Jefferson Luiz Brum Marques
Cesar Ramos Rodrigues
Daniela Ota Hisayasu Suzuki
José Marino Neto
Renato García Ojeda *Editors*

# IX Latin American Congress on Biomedical Engineering and XXVIII Brazilian Congress on Biomedical Engineering

Proceedings of CLAIB and CBEB 2022, October 24–28, 2022, Florianópolis, Brazil—Volume 2: Biomedical Signal Processing and Micro- and Nanotechnologies

Springer

# IFMBE Proceedings 99

The IFMBE Proceedings Book Series is an official publication of *the International Federation for Medical and Biological Engineering* (IFMBE). The series gathers the proceedings of various international conferences, which are either organized or endorsed by the Federation. Books published in this series report on cutting-edge findings and provide an informative survey on the most challenging topics and advances in the fields of medicine, biology, clinical engineering, and biophysics.

The series aims at disseminating high quality scientific information, encouraging both basic and applied research, and promoting world-wide collaboration between researchers and practitioners in the field of Medical and Biological Engineering.

Topics include, but are not limited to:

- Diagnostic Imaging, Image Processing, Biomedical Signal Processing
- Modeling and Simulation, Biomechanics
- Biomaterials, Cellular and Tissue Engineering
- Information and Communication in Medicine, Telemedicine and e-Health
- Instrumentation and Clinical Engineering
- Surgery, Minimal Invasive Interventions, Endoscopy and Image Guided Therapy
- Audiology, Ophthalmology, Emergency and Dental Medicine Applications
- Radiology, Radiation Oncology and Biological Effects of Radiation
- Drug Delivery and Pharmaceutical Engineering
- Neuroengineering, and Artificial Intelligence in Healthcare

IFMBE proceedings are indexed by SCOPUS, EI Compendex, Japanese Science and Technology Agency (JST), SCImago. They are also submitted for consideration by WoS.

Proposals can be submitted by contacting the Springer responsible editor shown on the series webpage (see "Contacts"), or by getting in touch with the series editor Ratko Magjarevic.

Jefferson Luiz Brum Marques ·
Cesar Ramos Rodrigues ·
Daniela Ota Hisayasu Suzuki ·
José Marino Neto ·
Renato García Ojeda

Editors

# IX Latin American Congress on Biomedical Engineering and XXVIII Brazilian Congress on Biomedical Engineering

Proceedings of CLAIB and CBEB 2022, October 24–28, 2022, Florianópolis, Brazil—Volume 2: Biomedical Signal Processing and Micro- and Nanotechnologies

*Editors*
Jefferson Luiz Brum Marques
Institute of Biomedical Engineering,
Department of Electrical and Electronic
Engineering
Federal University of Santa Catarina
Florianópolis, Brazil

Cesar Ramos Rodrigues
Institute of Biomedical Engineering,
Department of Electrical and Electronic
Engineering
Federal University of Santa Catarina
Florianópolis, Brazil

Daniela Ota Hisayasu Suzuki
Institute of Biomedical Engineering,
Department of Electrical and Electronic
Engineering
Federal University of Santa Catarina
Florianópolis, Brazil

José Marino Neto
Institute of Biomedical Engineering
Federal University of Santa Catarina
Florianópolis, Brazil

Renato García Ojeda
Institute of Biomedical Engineering,
Department of Electrical and Electronic
Engineering
Federal University of Santa Catarina
Florianópolis, Brazil

# Preface

The IX Latin American Congress on Biomedical Engineering and XXVIII Brazilian Congress on Biomedical Engineering (CLAIB&CBEB 2022) took place simultaneously on October 24–28, 2022, in Florianópolis-SC, Brazil, and were organised by the Institute of Biomedical Engineering of The Federal University of Santa Catarina (IEB-UFSC), the Regional Council of Biomedical Engineering for Latin America (CORAL) and the Brazilian Biomedical Engineering Society (SBEB). These events were held remotely for the most part, with a small set of conferences taking place in person on the premises of IEB-UFSC (Florianópolis, Brazil). They included 11 hands-on technical workshops for students, 26 keynote speakers and symposia, and 40 oral and poster presentation sessions attended by about a thousand participants, including undergraduate and graduate students, academic researchers, and public and private sector agents.

We are proud to present in this book a selection of papers presented at this event by researchers from all over the world, reporting recent and innovative findings and technological outcomes in the many areas of interest of biomedical engineering. These papers represent nearly 50% of those original contributions presented at the CLAIB&CBEB 2022. Their academic quality has been warranted by careful peer review coordinated by an expert scientific committee of leading Latin American senior researchers in biomedical engineering. The content is organised into four volumes and eleven chapters, covering the most relevant areas of scientific and technological developments within the broad spectrum of biomedical engineering interests. We are sure that the contributions presented in this book give a deep overview of the leading edge in your expertise and other areas.

On behalf of Scientific and Organising Committees, we thank authors, academic reviewers and sponsoring societies such as CORAL, SBEB, UFSC, FAPESC and IEB-UFSC for their contributions. Moreover, we encourage readers to enjoy this amazing piece of scientific literature as a breadth of knowledge in the biomedical engineering field.

# Organisation

CLAIB&CBEB 2022 was organised by the Regional Council of Biomedical Engineering for Latin America (CORAL) and the Brazilian Biomedical Engineering Society (SBEB) in cooperation with the International Federation for Medical and Biological Engineering (IFMBE).

## Committees

The Organising Committee of the IX Latin American Congress of Biomedical Engineering (CLAIB 2022) and the XXVIII Brazilian Congress of Biomedical Engineering (CBEB 2022) was composed of the following members:

**Organising Committee**

Renato Garcia Ojeda (President)
Jefferson Luiz Brum Marques (Vice-President)

**Scientific Committee**

Jefferson Luiz Brum Marques

**Programme Committee**

José Marino-Neto

**Special Events Committee**

Cesar Ramos Rodrigues

**Finance and Disclosure Committee**

Daniela Ota Hisaysu Suzuki

**Follow-up Committee**

Helio Schechtman—President of the Brazilian Society of Biomedical Engineering (SBEB)

Elliot Vernet—President Regional Council of Biomedical Engineering for Latin America (CORAL)

Shankar Krishnan—Presidente International Federation of Medical and Biological Engineering (IFMBE)

# Part Editors

### Biomedical Robotics, Assistive Technologies and Health Informatics

Teodiano Freire Bastos Filho (UFES-Brazil)
Luis Eduardo Rodriguez Cheu (CSEJG-Colombia)
Renato Garcia Ojeda (UFSC-Brazil)

### Biomedical Image and Signal Processing

Márcio Holsbach Costa (UFSC-Brazil)
Virginia Laura Ballarin (UNMdP-Argentina)
Jurandir Nadal (UFRJ-Brazil)

### Biomedical Optics and Systems and Technologies for Therapy and Diagnosis

Renato Amaro Zângaro (UAM-Brazil)
Regiane Albertini de Carvalho (UNIFESP-Brazil)

### Biomedical Devices and Instrumentation

Percy Nohama (PUC-PR-Brazil)
Guilherme Nunes Nogueira (PUC-PR-Brazil)

### Bioengineering, Modelling and Simulation, Bioinformatics and Computational Biology

Antonio Carlos Guimarães de Almeida (UFSJ-Brazil)
Antônio Márcio Rodrigues (UFSJ-Brazil)
Luis Enrique Bergues Cabrales (UO-Cuba)

**Biomaterials, Tissue Engineering and Artificial Organs**

Idágene Aparecida Cestari (USP-Brazil)
Sônia Maria Malmonge (UFABC-Brazil)

**Biosensors, Bioinstrumentation and Micro-nanotechnologies**

Alcimar Barbosa Soares (UFU-Brazil)

**Biomechanics, Neural Engineering and Rehabilitation**

Adriano de Oliveira Andrade (UFU-Brazil)

**Clinical Engineering and Health Technology Assessment**

Fabiola Margarita Martínez-Licona (UAM-Mexico)
Roberto Macoto Ichinose (UFRJ-Brazil)

**Health Technology Innovation and Development**

Antonio Adilton Oliveira Carneiro (USP-Brazil)

**Metrology and Quality of Healthcare Technologies**

Carlos Rubén Dell'Aquila (UNSJ-Argentina)
Rodrigo Pereira Barretto da Costa Félix (InMetro-Brazil)

**Special Topics**

Adson Ferreira da Rocha (UnB-Brazil)
Renato Garcia Ojeda (UFSC-Brazil)

## Support Committee

Alexandre Holzbach Junior
Bruna Fanchin
Clara Teresa de Souza Ramos
Daniella de Lourdes Luna Santana de Andrade
Daniel Ribeiro de Moraes
Esteferson Quadros
Felipe Rettore Andreis
Flavio Maurício Garcia Pezzolla
Guilberth Alves de Matos

Henrique Rezer Mosquér
Israel Santana
Jéssica Rodrigues da Silva
Jonas Martins Maciel
Jone Follmann
Juliano Martins
Lucas Bertinetti Lopes
Maicon Francisco
Marcelo Dérick Oliveira das Chagas
Maria Angelica Martins
Mariana Ribeiro Brandão
Mateus Andre Favretto
Matheus Gama Costa
Paulo Miguel Rodrigues da Rocha
Pedro Paulo Santos Gomes da Silva
Rafael Glatz
Rafael Schantone Silva
Raul Guedert
Sandra Cossul
Susana Bartnikowsky
Victor Hugo de Freitas Morales
Vinícius Rodrigues Zanon

## Scientific Reviewers

| | |
|---|---|
| Adilmar Coelho Dantas | Universidade Federal de Uberlândia, Brazil |
| Adriana Gabriela Scandurra | Universidad Nacional de Mar del Plata, Argentina |
| Adriana Kauati | State University of Western Paraná, Brazil |
| Adriane Parraga | Universidade Estadual do Rio Grande do Sul, Brazil |
| Adriano Pereira | Universidade Federal de Uberlândia, Brazil |
| Adriano Péricles Rodrigues | Universidade Federal de Goiás, Brazil |
| Agustina Bouchet | Universidad de Oviedo, Spain |
| Agustina Garces Correa | Universidad Nacional de San Juan, Argentina |
| Alberto López Delis | Universidad de Oriente, Cuba |
| Aldira Domínguez | Universidade de Brasília, Brazil |
| Alejandro Espinel Hernández | Universidad do Oriente, Cuba |
| Alexander Alexeis Suárez-León | Universidad de Oriente, Cuba |
| Alexandre Balbinot | Universidade Federal do Rio Grande do Sul, Brazil |
| Alexandre Luís Cardoso Bissoli | Instituto Nacional da Propriedade Industrial, Brazil |
| Alexandre Romariz | Universidade de Brasilia, Brazil |

| | |
|---|---|
| Alfredo de Oliveira Assis | Universidade Federal de Goiás, Brazil |
| Alvaro David Orjuela Cañon | Universidad del Rosario, Argentina |
| Ana B. Pimentel-Aguilar | Instituto Nacional de Enfermedades Respiratorias, Mexico |
| Ana Karoline Almeida da Silva | Universidade de Brasília, Brazil |
| André Adami | Universidade de Caxias do Sul, Brazil |
| André Dantas | Instituto Santos Dumont, Brazil |
| Andre Lazzaretti | Universidade Tecnológica Federal do Paraná, Brazil |
| André Luís Fonseca Furtado | Instituto Federal do Sudeste de Minas Gerais, Brazil |
| Andrés Alberto Ramírez Duque | Universidad El Bosque, Columbia |
| Andrés Felipe Ruíz Olaya | Universidad Antonio Nariño, Columbia |
| Angela Abreu Rosa de Sá | Universidade Federal de Uberlândia, Brazil |
| Angela Salinet | Universidade de São Paulo, Brazil |
| Antonio Mauricio Miranda de Sá | Universidade Federal de Rio de Janeiro, Brazil |
| Aparecido Carvalho | Centro Universitário Salesiano São Paulo, Brazil |
| Argenis Adrian Soutelo Jimenez | Universidad de Oriente, Columbia |
| Ariana Moura Cabral | Universidade Federal de Uberlândia, Brazil |
| Ariel Braidot | Universidad Nacional de Entre Rios, Argentina |
| Arnaldo Fim Neto | Universidade Federal do ABC, Brazil |
| Arquímedes Montoya Pedrón | Hospital General Docente Juan Bruno Zayas Alfonso, Cuba |
| Arthur Carvalho Pires | Universidade Federal de São João Del-Rei, Brazil |
| Aura Ximena González Cely | Universidad Pedagógica y Tecnológica, Columbia |
| Axayacatl Morales | Universidad Autónoma Metropolitana Iztapalapa, Mexico |
| Beatriz Fernandes | Pontificia Universidad Catolica de Parana, Brazil |
| Beatriz Janeth Galeano Upegui | Universidad Pontificia Bolivariana, Columbia |
| Belkys Amador | Universidad Austral de Chile, Chile |
| Bruno Bispo | Universidade Federal de Santa Catarina, Brazil |
| Caio C. E. De Abreu | Universidade do Estado de Mato Grosso, Brazil |
| Carla Scorza | Universidade Federal de São Paulo, Brazil |
| Carlos Alirio Lozano Ortiz | Universidade Federal de Rio de Janeiro, Brazil |
| Carlos Araujo | Instituto Federal do Paraná, Brazil |
| Carlos Castillo | Fundación Universitaria del Área Andina, Columbia |
| Carlos Danilo Miranda Regis | Instituto Federal da Paraiba, Brazil |
| Carlos Galvão Pinheiro Júnior | Universidade Federal de Goiás, Brazil |
| Carlos Maciel | Universidade de São Paulo, Brazil |
| Carlos Magno Medeiros Queiroz | Instituto Federal do Triângulo Mineiro, Brazil |
| Carlos Valadao | Instituto Federal do Espírito Santo, Brazil |

| | |
|---|---|
| Carolina Tabernig | Universidad Nacional de Entre Ríos, Argentina |
| Cesar Ferreira Amorim | Universidade Federal de Uberlândia, Brazil |
| Cesar Rodrigues | Universidade Federal de Santa Catarina, Brazil |
| César Teixeira | University of Coimbra, Portugal |
| Cicero Hildenberg Lima De Oliveira | Universidade Tecnológica Federal do Paraná, Brazil |
| Claudia Lescano | Universidad Nacional de San Juan, Argentina |
| Claudia Mirian de Godoy Marques | Universidade do Estado de Santa Catarina, Brazil |
| Cleber Zanchettin | Universidad Nacional de San Juan, Argentina |
| Cleison Daniel Silva | Universidade Federal do Pará, Brazil |
| Cristian Blanco | Universidad Antonio Nariño, Columbia |
| Cristian David Guerrero Mendez | Universidad Antonio Nariño, Columbia |
| Cristiano Jacques Miosso | Universidade de Brasília, Brazil |
| Cristiano Jacques Miosso | Universidade de Brasília, Brazil |
| Cristina Shimoda Uechi | MCTIC, Brazil |
| Dafne Mendes Soares | Universidade Federal de São João del-Rei, Brazil |
| Damien Depannemaecker | Universidade Federal de São Paulo, Brazil |
| Daniel Cavalieri | Instituto Federal do Espírito Santo, Brazil |
| Daniela del Carmen Gonzalez | Universidad Nacional de San Juan, Argentina |
| Daniela O. H. Suzuki | Universidade Federal de Santa Catarina, Brazil |
| Daniella De Lourdes L. S. Andrade | Universidade Federal de Santa Catarina, Brazil |
| Danilo Melges | Universidade Federal de Minas Gerais, Brazil |
| Danilo Silva | Universidade Federal de Santa Catarina, Brazil |
| David Sérgio Adães de Gouvêa | Universidade Federal de Juiz de Fora, Brazil |
| Débora de Fátima Camillo Ribeiro | Pontifícia Universidade Católica do Paraná, Brazil |
| Delmo Benedito Silva | Universidade Federal de São João Del-Rei, Brazil |
| Denis Delisle Rodriguez | Instituto Santos Dumont, Brazil |
| Diana Gutierrez | Universidad Manuela Beltran, Columbia |
| Diego Beltramone | Universidad Nacional de Córdoba, Argentina |
| Diego Sebastián Comas | National University of Singapore, Singapore |
| Éberte Freitas | Universidade Federal do Espírito Santo, Brazil |
| Edgard Morya | Instituto Santos Dumont, Brazil |
| Edras Pacola | Universidade Positivo, Brazil |
| Eduardo Antonio Fragoso Dias | Universidade Federal do Espírito Santo, Brazil |
| Eduardo M. Scheeren | Pontifícia Universidade Católica do Paraná, Brazil |
| Eduardo Naves | Universidade Federal de Uberlândia, Brazil |
| Elgison Da Luz Dos Santos | Centro Universitário Internacional Uninter, Brazil |
| Eliete Maria De Oliveira Caldeira | Universidade Federal do Espírito Santo, Brazil |
| Elisa Perez | Universidad Nacional de San Juan, Argentina |
| Elisangela Manffra | Pontifícia Universidade Católica do Paraná, Brazil |
| Elisângela Oliveira Carneiro | Universidade Estadual de Feira de Santana, Brazil |

| | |
|---|---|
| Emanuel Tello | Universidad Nacional de San Juan, Argentina |
| Enrique J. Marañón | Universidad de Oriente, Cuba |
| Enrique Mario Avila Perona | Universidad Nacional de San Juan, Argentina |
| Eric Laciar Leber | Universidad Nacional de San Juan, Argentina |
| Ernesto Suaste | CINVESTAV Sección Biolectrónica, Mexico |
| Esmirna Cascaret Fonseca | Centro Provincial de Medicina Deportiva, Cuba |
| Esteban Lanzarotti | Universidad de Buenos Aires, Argentina |
| Eugenio Orosco | Universidade Nacional de San Juan, Argentina |
| Euler Garcia | Universidade de Brasília, Brazil |
| Evandro Salles | Universidade Federal do Espírito Santo, Brazil |
| Fabiana Bertoni | Universidade Estadual de Feira de Santana, Brazil |
| Fabiana da Silveira Bianchi Perez | Faculdades Alfredo Nasser, Brazil |
| Fabio Viegas Caixeta | Universidade de Brasília, Brazil |
| Fabricio Lima Brasil | Instituto Santos Dumont, Brazil |
| Fabricio Neves Mendonca | Instituto federal de Minas Gerais, Brazil |
| Fabrício Noveletto | Universidade do Estado de Santa Catarina, Brazil |
| Fátima L. S. Nunes | Universidade de São Paulo, Brazil |
| Felipe Rettore Andreis | HST Aalborg University, Denmark |
| Fellipe Allevato Martins da Silva | Universidade Federal do Rio de Janeiro, Brazil |
| Fernando Henrique Magalhães | Universidade de São Paulo, Brazil |
| Fernando Jorge Muñoz Zapata | Universidad Nacional de San Juan, Argentina |
| Fernando Valdés-Pérez | Universidade do Oriente, Cuba |
| Flavio Buiochi | Universidade de Sao Paulo, Brazil |
| Floriano Salvaterra | Universidade Federal de Rio de Janeiro, Brazil |
| Francisco Assis de Oliveira Nascimento | Universidade de Brasília, Brazil |
| Franco Simini | Universidad de la Republica de Uruguay, Uruguay |
| Frieda Saicla Barros | universidade Tecnológica Federal do Paraná, Brazil |
| Gabriel Bruno M. Fernandes | Universidade Federal de Santa Catarina, Brazil |
| Gabriel Domingo Vilallonga | Universidad Nacional de Catamarca, Argentina |
| Gabriel Eduardo Cañadas Fragapane | Universidad Nacional de San Juan, Argentina |
| Gabriel Motta Ribeiro | Universidade Federal de Rio de Janeiro, Brazil |
| Gilcélio Silveira | Universidade Federal de São João del-Rei, Brazil |
| Glauco Cardozo | Instituto Federal de Santa Catarina, Brazil |
| Glécia Luz | Universidade de Brasília, Brazil |
| Gonzalo Quiroga | Universidad Nacional de San Juan, Argentina |
| Guilherme Augusto Gomes De Villa | Universidade Federal de Goiás, Brazil |
| Guillermo Abras | Universidad Nacional de Mar del Plata, Argentina |
| Gustavo Meschino | Universidad Nacional de Mar del Plata, Argentina |

| | |
|---|---|
| Gustavo Vivas | Empresa Brasileira de Serviços Hospitalares, Brazil |
| Harlei Leite | Universidade Federal do Espírito Santo, Brazil |
| Henrique Takachi Moriya | Universidade de Sao Paulo, Brazil |
| Hugo Líbero | Instituto Federal de Goiás, Brazil |
| Humberto Gamba | Universidade Tecnológica Federal do Paraná, Brazil |
| Humberto Romano | INTECNUS Foundation, Argentina |
| Isabela Alves Marques | Universidade Federal de Uberlândia, Brazil |
| Isabela Miziara | Universidade Federal do Pará, Brazil |
| Ismar Cestari | Instituto do Coração Universidade de São Paulo, Brazil |
| Jair Trapé Goulart | Universidade de Brasília, Brazil |
| Javier Camacho | Universidade EIA, Columbia |
| Jerusa Marchi | Universidade Federal de Santa Catarina, Brazil |
| Jessica Lara | Universidade Federal de São João Del-Rei, Brazil |
| João Henrique Kleinschmidt | Universidade Federal do ABC, Brazil |
| João Lameu | Universidae Federal do ABC, Brazil |
| Joao Marques | University of Saint Joseph, United States |
| João Otávio Bandeira Diniz | Instituto Federal do Maranhão, Brazil |
| João Panceri | Instituto Federal do Espírito Santo, Brazil |
| Joaquim Cezar Felipe | Universidade de São Paulo, Brazil |
| Joaquim Mendes | Universidade do Porto, Portugal |
| Joaquin Azpiroz Leehan | UAM Iztapalapa, Mexico |
| Jorge Takenaga | Universidad de Monterrey, Mexico |
| Jose Carlos Cunha | Pontificia Universidade Católica do Paraná, Brazil |
| José Di Paolo | Universidad Nacional de Entre Ríos, Argentina |
| José Felicio Da Silva | Universidade de Brasília, Brazil |
| José Foggiatto | Universidade Tecnológica Federal do Paraná, Brazil |
| José María Flores | Universidad Nacional de Entre Rios, Argentina |
| José Wilson Bassani | Universidade de Campinas, Brazil |
| Juan Carlos Iturrieta Gimeno | Universidad Nacional de San Juan, Argentina |
| Juan Carlos Perfetto | Universidad de Buenos Aires, Argentina |
| Juan Enrique Palomares Ruiz | Instituto Tecnológico Superior de Cajeme, Mexico |
| Juan Graffigna | Universidad Nacional de San Juan, Argentina |
| Juan I. Montijano | University of Zaragoza, Spain |
| Juan Ignacio Pastore | Universidad Nacional de Mar del Plata, Argentina |
| Juan Pablo Tello Portillo | Universidad del Norte, Columbia |
| Juan Pastore | Universidad Nacional de Mar del Plata, Argentina |
| Juan Vorobioff | Comisión Nacional de Energía Atómica CNEA, Argentina |

| | |
|---|---|
| Juliana Santa Ardisson | Universidade Federal do Espírito Santo, Brazil |
| Juliano Costa Machado | Instituto Federal Sul-Rio-Grandense, Brazil |
| Julio Alberto Rojas Vargas | Universidade do Oriente, Cuba |
| Julio Cesar Nievola | Pontifícia Universidade Católica do Paraná, Brazil |
| Karla Aparecida Ferreira | Universidade Federal de São João Del-Rei, Brazil |
| Katia Prus | Universidade Tecnológica Federal do Paraná, Brazil |
| Kelison Tadeu Ribeiro | Instituto Federal do Espírito Santo, Brazil |
| Lacordaire Kemel Pimenta Cury | Instituto Federal de Educação, Ciência e Tecnologia Goiano, Brazil |
| Leandro Bueno | Instituto Federal do Espírito Santo, Brazil |
| Lena Perez Font | Universidad de Oriente, Cuba |
| Leonardo Abdala Elias | Universidade Estadual de Campinas, Brazil |
| Leonardo Abrantes | Universidade Federal de Rio de Janeiro, Brazil |
| Leonardo Batista | Universidade Federal da Paraíba, Brazil |
| Leonardo Ramirez | Universidad Militar Nueva Granada, Columbia |
| Leticia Silva | Universidade Federal do Espírito Santo, Brazil |
| Lorena Orosco | Universidad Nacional de San Juan, Argentina |
| Lourenço Madruga Barbosa | Universidade Tecnológica Federal do Paraná, Brazil |
| Lucas Côgo Lampier | Universidade Federal do Espírito Santo, Brazil |
| Lucenildo Cerqueira | Universidade Federal de Rio de Janeiro, Brazil |
| Luciana Menezes Xavier de Souza | Universidade Federal de Santa Catarina, Brazil |
| Luciana Roberta Peixoto | Universidade de Brasília, Brazil |
| Luciane Aparecida Pascucci de Souza | Universidade Federal do TriânguloMineiro, Brazil |
| Luciano Menegaldo | Universidade Federal de Rio de Janeiro, Brazil |
| Luis Carlos Hernandez Barraza | National University of Singapore, Singapore |
| Luis Eduardo Maggi | Universidade Federal do Acre, Brazil |
| Luis Jiménez-Angeles | Universidade Nacional Autônoma do México, Mexico |
| Luis Miguel Zamudio | Universidade do Oriente, Cuba |
| Luis Pulenta | Universidad Nacional de San Juan, Argentina |
| Luis Roberto Barriere | Universidad Don Bosco, El Salvador |
| Luis Sifuentes | Universidad Cuauhtémoc Aguascalientes, Mexico |
| Luis Vilcahuaman | Pontificia Universidad Catolica del Peru, Peru |
| Maikel Noriega Alemán | Universidade do Oriente, Cuba |
| Maira Ranciaro | Pontifícia Universidade Católica do Paraná, Brazil |
| Malki-çedheq Benjamim C. Silva | Universidade Federal de Pernambuco, Brazil |
| Marcella Carneiro | Universidade de Brasília, Brazil |
| Marcelo Lencina | National Technological University, Argentina |
| Marcio Nogueira de Souza | Universidade Federal de Rio de Janeiro, Brazil |

| | |
|---|---|
| Marcio Rodrigues | Instituto Federal do Sudeste de Minas Gerais, Brazil |
| Marco Antonio Garcia | Universidade Federal de Juiz de Fora, Brazil |
| Marcos Hara | Instituto Federal do Paraná, Brazil |
| Marcus Vieira Fraga | Universidade de Goiás, Brazil |
| Maria Claudia F. Castro | Centro Universitário da FEI, Brazil |
| Mariana de Mello Gusso | Pontifícia Universidade Católica do Paraná, Brazil |
| Mariela Azul Gonzalez | Universidad Nacional de Mar del Plata, Argentina |
| Mário Fabrício Fleury Rosa | Universidade de Brasília, Brazil |
| Marly Guimaraes Fernandes Costa | Universidade Federal do Amazonas, Brazil |
| Martha Ortiz | Universidad Autonoma de Mexico-Iztapalapa, Mexico |
| Martha Ribeiro | Universidade de São Paulo, Brazil |
| Martha Zequera | Pontificia Universidad Javeriana, Columbia |
| Martín Guzzo | Universidad Nacional de San Juan, Argentina |
| Maryangel Jhoseline Medina | ComisiónNacional de Energía Atómica, Argentina |
| Mateus Andre Favretto | Universidade do Estado de Santa Catarina, Brazil |
| Mauren Abreu de Souza | Pontificia Universidade Católica do Paraná, Brazil |
| Mauricio Cagy | Universidade Federal do Rio de Janeiro, Brazil |
| Mauricio Kugler | Nagoya Institute of Technology, Japan |
| Mauricio Tavares | Contronic Tecnologia para Diagnósticos, Brazil |
| Mayra Aarao | Universidade Federal de São João del-Rei, Brazil |
| Mohamed Ahmed | Universidade Federal de São João Del-Rei, Brazil |
| Monica Teresita Miralles | Universidade de Buenos Aires, Argentina |
| Munir Gariba | Universidade de São Paulo, Brazil |
| Murilo Contó | Boston Scientific, Brazil |
| Nahuel M. Olaiz | Universidade de Buenos Aires, Argentina |
| Nicolas Jacobo Valencia Jimenez | Universidad Santiago de Cali, Columbia |
| Nilton Correia Da Silva | Universidade de Brasília, Brazil |
| Olavo Luppi Silva | Universidade Federal do ABC, Brazil |
| Pablo Caicedo | Universidade Autônoma da Colômbia, Columbia |
| Pablo Daniel Cruces | Universidad de Buenos Aires, Argentina |
| Pablo Diez | Universidad Nacional de San Juan, Argentina |
| Pablo Turjanski | Universidad de Buenos Aires, Argentina |
| Patrick Marques Ciarelli | Universidade Federal do Espírito Santo, Brazil |
| Paula Brandao Furlan | Universidade Federal de Rio de Janeiro, Brazil |
| Paula Hembecker | Pontifícia Universidade Católica do Paraná, Brazil |
| Paulo Ambrosio | Universidade Estadual de Santa Cruz, Brazil |
| Paulo Broniera Junior | Federação das Indústrias Estado de São Paulo, Brazil |
| Paulo Eduardo Narcizo de Souza | Universidade de Brasília, Brazil |

| | |
|---|---|
| Paulo Gubert | Universidade Católica de Pelotas, Brazil |
| Paulo Jose Abatti | Universidade Tecnológica Federal do Paraná, Brazil |
| Paulo Roberto Sanches | Hospital de Clínicas de Porto Alegre, Brazil |
| Paulo Rogerio Scalassara | Universidade Tecnológica Federal do Paraná, Brazil |
| Pedro Bertemes Filho | Universidade do Estado de Santa Catarina, Brazil |
| Pedro Gewehr | Universidade Tecnológica Federal do Paraná, Brazil |
| Pedro Moisés de Sousa | Universidade Federal de Uberlândia, Brazil |
| Pedro Xavier de Oliveira | Universidade Estadual de Campinas, Brazil |
| Rafael Sanchotene Silva | Universidade Federal de Santa Catarina, Brazil |
| Raimes Moraes | Universidade Federal de Santa Catarina, Brazil |
| Raul Guedert | Universidade Federal de Santa Catarina, Brazil |
| Renata Coelho Borges | Universidade Tecnológica Federal do Paraná, Brazil |
| Renato de Araujo | Universidade Federal de Pernambuco, Brazil |
| Renato Naville Watanabe | Universidade Federal do ABC, Brazil |
| Ricardo Borsoi | University of Lorraine, France |
| Ricardo Ruggeri | Autoridad Regulatoria Nuclear, Argentina |
| Ricardo Suyama | Universidade Federal do ABC, Brazil |
| Roberto Sagaró Zamora | Universidad de Oriente, Cuba |
| Rocio Belén Buenamaizón | Universidad Nacional de San Juan, Argentina |
| Rodolfo Dias Correia | Hospital das Clínicas de Ribeirão Preto, Brazil |
| Rodrigo Gomide | Instituto Federal Goiano, Brazil |
| Rodrigo Pinto Lemos | Universidade Federal de Goiás, Brazil |
| Rodrigo Santos | Universidade Federal de Juiz de Fora, Brazil |
| Rodrigo Varejão Andreão | Instituto Federal do Espírito Santo, Brazil |
| Roger Gomes Tavares de Mello | Universidade Federal do Rio de Janeiro, Brazil |
| Ronni Amorim | Universidade de Brasília, Brazil |
| Rosa Itzel Flores | Universidad Autonoma de Mexico - Iztapalapa, Mexico |
| Rosangela Jakubiak | Universidade Tecnológica Federal do Paraná, Brazil |
| Rossana Rivas Tarazona | Pontificia Universidad Catolica del Peru, Peru |
| Rubén Acevedo | Universidad Nacional de Entre Ríos, Argentina |
| Sandra Cossul | Universidade Federal de Santa Catarina, Brazil |
| Sandra Rocha Nava | Universidad La Salle, Columbia |
| Sara Rosa de Sousa Andrade | Faculdades Estácio de Sá, Brazil |
| Sérgio Francisco Pichorim | Universidade Tecnológica Federal do Paraná, Brazil |
| Sergio Furuie | Universidade de São Paulo, Brazil |
| Sergio Ioshii | Pontificia Universidade Católica do Paraná, Brazil |

## Sponsoring Institutions

CORAL: Consejo Regional de Ingeniería Biomédica para América Latina

www.coralbiomedica.org

SBEB: Sociedade Brasileira de Engenharia Biomédica

www.sbeb.org.br

IFMBE: International Federation of Medical and Biological Engineering

www.ifmbe.org

UFSC: Universidade Federal de Santa Catarina

www.ufsc.br

IEB-UFSC: Instituto de Engenharia Biomédica da UFSC

www.ieb.ufsc.br

FAPESC: Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina

www.fapesc.sc.gov.br

# Contents

# Biomedical Optics and Systems and Technologies for Therapy and Diagnosis

# Biomedical Image and Signal Processing

# Estimation of Heart Inclination Angle Using Posteroanterior Chest Radiograph and Comparison with Cardiac Axis Obtained from Synthesized Vectorcardiogram

Edison Silva Filho[1]([✉])〔ID〕, Lucas José da Costa[2]〔ID〕, Felipe Akio Matsuoka[2]〔ID〕,
Alembert Eistein Lino Alvarado[1]〔ID〕, Vinicius Ruiz Uemoto[2]〔ID〕,
Renato de Aguiar Hortegal[2]〔ID〕, Renata Valeri de Freitas[2]〔ID〕,
and Henrique Takachi Moriya[1]〔ID〕

[1] Biomedical Engineering Laboratory, University of São Paulo, São Paulo, Brazil
edison.silva2@usp.br
[2] Dante Pazzanese Institute of Cardiology, São Paulo, Brazil

**Abstract.** In medical practice, it is common to perform electrocardiography exams and by mathematical transformations to obtain the vectorcardiogram. The vectorcardiogram provides important information for medical diagnosis, such as the angle of inclination of the heart. This article aims to present a methodology for estimating the QRS vector-related angle of the heart using a posteroanterior chest radiograph image. We used an open source image processing software (Icy software version 2.3.0.0, Institut Pasteur, France, 2021) to perform a manual measurement of the target angle by analyzing relevant morphological structures from the x-ray images and using some functions to help the user to measure it. 18 radiographic images were selected to measure the angle of the heart by two independent individuals. The measured angles were compared using the mean absolute error (MAE). We then computed the QRS peak elevation angles of the vectorcardiogram (VCG) of the 57 patients collected at Dante Pazzanese Institute of Cardiology. In addition, an individual was randomly selected to measure a set of 57 radiographic images of these same patients. We performed the statistical treatments and the results suggested that the proposed manual method may be an alternative, viable and fast approach to estimating the anatomical heart axis for the purpose of aiding in medical diagnosis. However, further comparisons with more data and information are needed to determine its validity and possible method improvements.

**Keywords:** Heart angle · vectorcardiogram · chest x-ray image

## 1   Introduction

Commonly, at the Dante Pazzanese Institute of Cardiology, patients are submitted to routine tests such as electrocardiogram (ECG) and posteroanterior chest radiography. These exams are performed mainly due to their low cost, relative ease of execution and the delivery of information to support medical diagnosis.

The electrocardiogram (ECG) records in waveforms the electrical potentials on the surface of the body generated by the electrical activity of the heart. Any variation in cardiac impulse transmission can cause abnormal electrical potentials and change the waveforms on the electrocardiogram. These abnormal electrical potentials can be related to structural changes in the heart that affect its activity or simply electrical abnormalities. For this reason, an ECG is important in diagnosing serious heart muscle abnormalities [1,2].

Another way to analyze the electrical activity of the heart is through the vectorcardiogram (VCG). The VCG is the vector representation of the electrical activities of the heart considering three mutually orthogonal axes (vertical, transversal, and sagittal). In the VCG, the electrical activity of the heart is described by three loops that represent the P wave, the QRS complex, and the T wave. The loop of the QRS complex is oval and faces the same direction as the cardiac axis of the heart. In the normal heart, the direction of the mean QRS vector is approximately 59° [1,3,4].

A way to get the VCG is through the mathematical transformation of the electrocardiogram. There are several different transformation methods, such as the quasi-orthogonal transform, the Inverse Dower Transform (IDT), P Least Square Value (PLSV) and Q Least Square Value (QLSV) Transformations, Mason-Likar (ML) and the Kors regression transform, the latter with relative better results than the others. The Kors regression method is a technique for constructing the cardiogram vector from the ECG, using a linear combination of eight conventional ECG leads [3,4]. Figure 1 shows the Kors transformation matrix derived by regression technique in order to minimize the mean error between the measured VCG and the transformed VCG [5].

| Lead | I | II | V1 | V2 | V3 | V4 | V5 | V6 |
|---|---|---|---|---|---|---|---|---|
| X | 0.38 | −0.07 | −0.13 | 0.05 | −0.01 | 0.14 | 0.06 | 0.54 |
| Y | −0.07 | 0.93 | 0.06 | −0.02 | −0.05 | 0.06 | −0.17 | 0.13 |
| Z | 0.11 | −0.23 | −0.43 | −0.06 | −0.14 | −0.20 | −0.11 | 0.31 |

**Fig. 1.** Transformation coefficients of Kors regression method [5]

The joint use of the ECG and the VCG increases the precision in the diagnoses, but due to the difficulty in relation to the number of electrodes and their positioning, most of the time ECG is performed. However, VCG tests are more sensitive for detecting hypertrophy and ischemic heart disease. But the advent

of mathematical transformation mitigated the difficulty of joint analysis of the two methods [4,6].

In the posteroanterior chest X-ray, the heart is located close to the frame, decreasing the cardiac magnification that can cause a false impression of increased heart volume. The chest X-ray is important in helping to diagnose cardiac diseases. With it, it is possible to identify structures and morphological changes, as occurs, for example, in the left ventricular aneurysm by identifying the aneurysmal dilatation resulting from a myocardial infarction [7].

Through X-ray it is possible to estimate the inclination of the anatomical axis of the heart and compare it with the electrical axis. Many times there is a strong correlation, especially in the anteroposterior and longitudinal axes (refer to the article), but some diseases can affect the heart's inclination and having different methodologies at hand can help in the medical diagnosis [8].

This article aims to present a methodology for estimating the QRS vector-related angle of the heart using a posteroanterior chest radiograph image. The measurement of heart inclination angle by the method proposed here is relevant as an additional source of information that complements traditional ECG/VCG exams for medical decision. We also computed the QRS peak elevation angles of the same patients' VCG and compared both angles. We emphasize that we did not find studies in the literature for estimating the inclination of the angle of the heart with a methodology similar to that proposed by this article.

## 2    Materials and Methods

Data were collected as part of a study by the Dante Pazzanese Institute of Cardiology, and this project was approved according to the *Certificado de Apresentação de Apreciação Ética* 76085317.5.3001.5185.

### 2.1    Manual Measurement of the Heart Angle

A set of posteroanterior X-ray images of patients from Dante Pazzanese Institute of Cardiology was used to measure the approximated angle of the QRS vector of the heart. An open-source image processing software (Icy software version 2.3.0.0, Institute Pasteur, France, 2021) was employed to manually calculate the angle.

The methodology to estimate the angle of the heart took into account the anatomy of the heart and its location inside the rib cage. In this study, the anatomy of the heart was important to recognize structural regions inside chest x-ray images to allow the definition of two points of interest, and be able to trace the reference line to estimate the angle of inclination of the heart. According to Fig. 2, the estimated points of interest for plotting the reference line were the ones located in the apex of the heart, and amides the superior vena cava (SVC) and the right atrium (RA).

To draw the reference line, we used some features of the image processing software to facilitate the identification of regions of interest. First, the contour

of the heart was found using a block of contour detection [9]. Second, a vertical line along the spine was drawn by defining two points with sharing the same abscissa. Then, we utilized a tool to draw an inclined line that approximately passed through the two previously defined points.
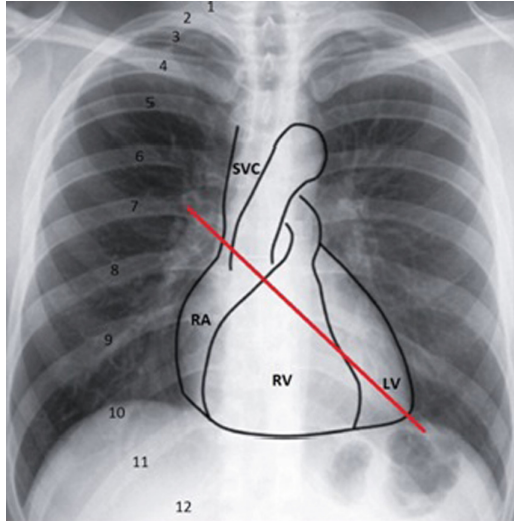


**Fig. 2.** Reference points for plotting the heart slope straight line. Points located at the apex of the heart and between the superior vena cava (SVC) and the right atrium (RA). [10]. Adapted by the author.

Finally, we measured the angle formed by the vertical and reference line with the Angle Helper tool, and calculated the angle of the heart by finding the supplement of the previously measured angle. This angle was considered as the manual measurement of the angle corresponding to the QRS vector of the heart by means of a posteroanterior chest x-ray image. Figure 3 shows the application of the method described in a chest x-ray image.

To validate the proposed method, two independent people performed measurements on a subset of 18 chest X-ray images. In both cases, images took an average time of 15 min to be analyzed.

## 2.2   Automatic Measurement

To estimate the heart angle in an automatic way, we implemented an algorithm in a mathematical programming environment (MATLAB, The MathWorks, USA) capable of processing electrocardiograms and returning quantitative parameters of global electrical heterogeneity (GEH), such as sum absolute QRST integral, spatial QRST angle, spatial ventricular gradient (SVG) magnitude, SVG elevation, SVG azimuth including the peak QRS elevation angle [11].
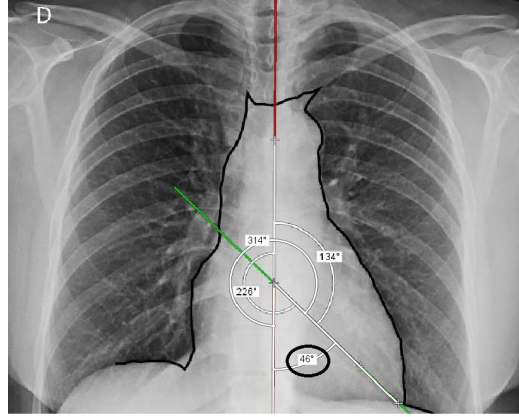
**Fig. 3.** Application of the Icy program on the chest x-ray image to estimate the angle of inclination of the heart.

We acquired a set of 57 standard 12-lead electrode ECGs using a digital electrocardiograph (C30+, TEB, Brazil). Then, the VCGs were synthesized from the 12-lead ECGs using the Kors matrix. The coefficients that we present in Fig. 1 allow the ECG to be transformed into a synthesized VCG with relevant information, such as the elevation angle of the QRS peak that we want to find. Figure 4 illustrates the use of the Kors regression matrix in ECG for the construction of the VCG.
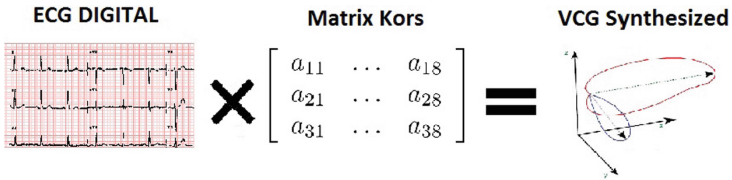


**Fig. 4.** Kors regression matrix in ECG for the construction of the VCG.

## 2.3    Method Comparison

After measuring and calculating the angle values for each subject using both methods, we calculated the mean absolute error (MAE) according to Eq. 1 for the 18 angles measured by the two different trained subjects. [12].

$$MAE = N^{-1} \sum_{i=1}^{n} |O_i - P_i| \tag{1}$$

where:

– $N$ is the number of samples
– $P_i$ and $O_i$ are measured angles by different persons or methods.

The mean and standard deviation were calculated for each method separately for later comparison of the results. Additionally, we calculated the coefficient of variation for the automatic method and for each trained subject.

## 3   Results

When applying Eq. 1 to the measurements of heart angles obtained by two trained individuals using the manual method in 18 chest x-ray images, a MAE of 2.4° was calculated. In Fig. 5, the relative frequency distribution of angle errors when compared two trained individuals are presented. We observe that most of the results have relatively low error. Only in the first column that has errors in the range from zero to two degrees are eight of the eighteen measurements obtained by the two trained individuals, which is equivalent to 44.4% of the data. On the other hand, the last column that has errors in the range of six to eight degrees are only two of the 18 measurements, which is equivalent to 11.1% of the data.
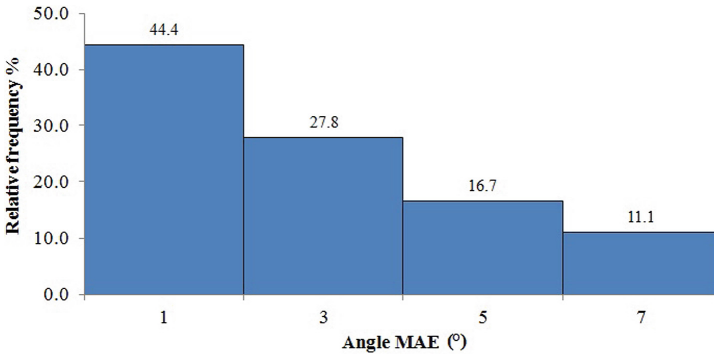


**Fig. 5.** Relative frequency distribution of measured angle errors between two trained subjects showing that most measurements have relatively low errors.

Table 1 presents a summary of the results of the calculations of means, expanded uncertainties and coefficients of variation for different subjects and methods

The first subject estimated, within the set of 18 chest radiographs, a mean angle of 50.3° with a standard deviation of 3.9°, coefficient of variation of 7.75% and an expanded uncertainty of 2.7°. However, the second subject with the same set of images obtained a mean angle of 47.9° with a standard deviation of 5.1°, coefficient of variation of 10.65% and an expanded uncertainty of 3.1°. Both results are presented with a confidence level of approximately 95%.

**Table 1.** Presentation of the calculated values for the automatic method and for the manual method by two different subjects.

|  | Result | Coefficient of Variation |
|---|---|---|
| Subject 1 | $(50.3 \pm 2.7)°$ | 7.75% |
| Subject 2 | $(47.9 \pm 3.1)°$ | 10.65% |
| Automatic method | $(57.2 \pm 7.5)°$ | 48.43% |

When considering the measurements obtained from the 136 chest x-ray images, an average angle of 52.1° with a standard deviation of 6.2° and an expanded uncertainty of 2.3° with approximately 95% confidence level was estimated.

In the automatic method on a set of 57 synthesized VCGs available, an average angle of 57.2° with a standard deviation of 27.7°, coefficient of variation of 48.43% and an expanded uncertainty of 7.5° was estimated. This same dataset we also analyzed and arrived at an average angle of 51.3° with a standard deviation of 5.2° and an expanded uncertainty of 2.4°. Both results are presented with a confidence level of approximately 95%.

## 4   Discussion

Vector analysis of electrocardiograms is a consolidated and widely used method to calculate, among other factors, the QRS vector angle. Based on our results, the proposed manual method proved to be a viable alternative to estimate the heart angle. The differentiation between the coefficients of variation of the two trained subjects may have its cause during the determination of the points of interest to trace the slope line.

Sometimes the anatomical axis is not intrinsically the same electrical axis that we find in VCGs. Some factors, such as existing diseases, can interfere with the results obtained and having the patient's anamnesis can help in the interpretation of the results.

The manual method with chest radiography appeared as an alternative option due to the low MAE between the measured angles of the two trained subjects. Further comparisons may help to determine the reproducibility of this method.

Currently, the manual method requires the individual to have knowledge of how the Icy software works and to recognize structural regions within chest radiography images. A trained individual takes approximately 15 min to analyze an image. However, in the future, the method of measuring the angle of inclination of the heart by x-ray imaging can be optimized using artificial intelligence and be subject to comparison with other studies.

In the literature we did not find a similar method for comparison. We were restricted to comparing the method between independent individuals and with the synthesized vectorcardiogram.

## 5    Conclusion

Although the proposed method to measure the angle of the heart by means of chest X-ray is performed completely manually, it has proved to be an interesting technique that delivers good results that, together with other exams, can help the medical team in decision making. However, despite the promising results of the methodology proposed in this article, it still lacks a broader evaluation with a larger set of images and qualitative and quantitative data from patients.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Hall, J.E.: Tratado de Fisiologia Médica. Elsevie, Rio de Janeiro (2011)
2. Antti, E., Aro Aapo L., Kenttä, T., et al.: 12-Lead electrocardiogram as a predictor of sudden cardiac death: from epidemiology to clinical practice. Scand. Cardiovasc. J. **50**, 253–259 (2016)
3. Ittalo, S., Cecilia, C., Renato, H., Carlos, R.: Validação do Vetorcardiograma de Kors pela Identificação de Infarto Agudo do Miocárdio in Anais de XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (Paraíba, Brazil) (2018)
4. Rene, J., Radek, M., Lukas, D.: Comparison of different electrocardiography with vectorcardiography transformations. Sensors **19**(14), 3072 (2019)
5. Jaroslav, V., Marek, P.: Review of processing pathological vectorcardiographic records for the detection of heart disease. Front. Physiol. **13**, 856590 (2022)
6. Sumche, M., Arie, C.M., Martin, J.S., Cees, A.S.: Vectorcardiographic diagnostic prognostic information derived from the 12-lead electrocardiogram: historical review and clinical perspective. J. Electrocardiol. **48**, 463–475 (2015)
7. Gerald, L., Simon, M., Berman, L.: The Chest X-Ray. Saunders Ltd. (2008)
8. Noble, O.F., John, R.B.: Anatomic and electrocardiographic position of the heart. Circulation **3**, 906–910 (1951)
9. Alexandre, D., Roman, T., Elisabeth, L., Nancy, G., Jean-Christophe, O.-M.: 3-D active meshes: fast discrete deformable models for cell tracking in 3-D time-lapse microscopy. IEEE Trans. Image Process. **20**, 1925–1937 (2010)
10. Reference points for plotting the heart slope straight line at. https://cardiopapers.com.br/conceitos-basicos-que-estruturascardiacas-sao-visualizadas-na-radiografia-de-torax
11. Waks, J.W., Sitlani, C.M., Soliman, E.Z., et al.: Global electric heterogeneity risk score for prediction of sudden cardiac death in the general population. Circulation **133**, 2222–2234 (2016)
12. David, L., Gregory, M.: Evaluating the use Of "Goodness-of-Fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. **35**, 233–241 (1999)

# EEG Signal Synchronization Patterns During Hand Laterality Judgment Task

Shirley Ferreira[1(✉)] , Rafaela Souza[1] , Antonio Silveira[2] ,
and Antonio Junior[2]

[1] Institute of Technology, Federal University of Pará, Belém, Brazil
shirley.ferreira@itec.ufpa.br
[2] Federal University of Pará, Electrical and Computer Engineering Laboratory,
Guamá, Belém, Brazil

**Abstract.** Motor imagery is a mental stimulation that triggers oscillatory events in sensorimotor rhythms often used in brain-computer interface applications, as well as a stimulus used to understand brain activity and improve the rehabilitation process of spinal cord injury people. Using the data extracted from an electronic game designed for a rehabilitation purpose, this paper describes the steps to extract the EEG synchronization patterns during the Hand Laterality Judgment Task that stimulates motor imagery processes, i.e., implicit motor imagery. As result, we obtained the curves of synchronization, in $dB$ values, that showed different levels of synchronization between alpha and beta bands and an activity mostly distributed over centro-parietal and parietal regions.

**Keywords:** Mental Rotation · Hand Laterality · ERD/ERS

## 1 Introduction

The decrease and increase in synchronization of the sensorimotor rhythms (SMR) are called event-related desynchronization (ERD) and event-related synchronization (ERS), respectively [1]. ERD/ERS phenomena are also known as an event-related spectral perturbation (ERSP) and it occurs during the motor imagery (MI) process, in which the subject imagines the movements of their body parts, commonly the hands, but also involving the foot and tongue, performed in repetitive trials. In the hand actual or motor imagery activity, the ERD patterns are, preponderantly, contralateral in the function of the hand imagined whilst the ERS are ipsilateral [2].

Electroencephalography (EEG) is an important tool to investigate brain activities, due to it is a noninvasive technique, relatively low-cost, and high temporal resolution [3]. Therefore, works using EEG data to estimate ERD/ERS patterns have been documented in the literature, focusing on classification problems, which is a fundamental step to the Brain-Computer Interface (BCI) systems application [4,5]. For MI, the time-frequency analysis has key-role in the detection and estimating of the ERD/ERS values [3], since the motor imagery

task induces events that modulate the ongoing alpha ($8-12$ Hz) and beta ($13-30$ Hz) activities [2], frequencies that compound the SMR. The modulation of the alpha and beta bands can be express, for instance, in relative change (unit in %), power ration (unit in $dB$), and subtraction (unit in $\mu V^2/Hz$) [3].

The methods for ERD/ERS estimation have importance to BCIs based on motor imagery also called MI-BCI. MI is extensively used in BCI systems due to discriminative stimulus proprieties, translated into EEG synchronization values, and also to involve a not expensive signal acquisition [6]. The experimental paradigm for motor imagery stimulated with a cue, in repetitive trials for each class, has been reported in research, with the goal to enhance MI-BCI systems accuracy [7,8]. When the motor imagery is stimulated in an indirect way, we have implicit motor imagery. For instance, the hand mental rotation engages motor imagery processes, because the subject imagines their hand rotating for a position. This protocol is often used when the goal is to identify hand laterality presented on a screen, that can be called Hand Laterality Judment Task (HLJT) [9]. Osuagwu *et al.* (2017) reported a classification performance of $83 \pm 3\%$ for implicit motor imagery of the left and right hand, whilst for explicitly stimulated motor imagery the performance was $81 \pm 8\%$ [10], showing the possibility to use implicit motor imagery for BCI application, until little described in the literature.

In the present study, a method to extract ERD/ERS values during HLJT was implemented. The research investigated how the task influence the oscillatory activity in SMR of the subjects performing a rehabilitation electronic game, described in the pioneer work [11], and called Alice in Land of the Hand, or only ALICE game.

## 2   Methodology

### 2.1   Participants

To perform ALICE game, twenty-three subjects participated (age: $25.65 \pm 3.88$ years), 10 men and 10 women, identified as $S_1, S_2, ..., S_{23}$. According to Edinburgh Lateral Dominance Inventory, 95% of the subjects were right-handed. One of the participants declared mixed preference. According to Mini-Mental State Examination (MMSE) all participants had no cognitive impairment. Furthermore, all subjects had no medical or neurological disorders and they hand-signed the Free and Informed Consent to participate of the experiment. The project was approved by the Ethics Committee of the University Hospital Onofre Lopes (HUOL/UFRN), released with CAAE number (Brazil Platform): 34478214.0.000 0.5292 and appreciation number: 821294. We prejudged the subject's performance and in this step, the data from the $S_5$, $S_{17}$, and $S_{20}$ subjects were eliminated. $S_5$ closed their eyes as a technique to response, increasing the amplitude of the alpha band. $S_{17}$ was eliminated due to be left-handed and, in the research context, the hemisphere dominance difference is not a desirable feature. $S_{20}$ failed in 104 trials of the 288 trials. Thus, the data from 20 subjects were used in the present work.

## 2.2  Experimental Paradigm

The game was development in collaboration between Brain Institute and the Department of Informatic and Applied Mathematics (DIMAP), both of the Federal University of Rio Grande do Norte (UFRN) [11], and the game was programmed in C language with the XNA Microsoft framework.

In the game world, a robot from the ALICE family disputes with an enemy robot on an electronic board game setting, and to take advantage over it, the participants should answer correctly the hand laterality that appears above the ALICE robot, inside circles. To perform the game, the subjects were seated on a comfortable chair, located 50 cm away from a monitor (Fig. 1). Then, they were asked to analyze and judge if the hand presented on the screen was left or right (HLJT) and then, press down the respective pedal (left or right) using their respective foot (left or right). The final score for each trial was based on reaction time (RT), i.e., the time from the stimulus shown on screen until the subject pressed down the pedal. The main goal was to be accurate with the shortest RT.
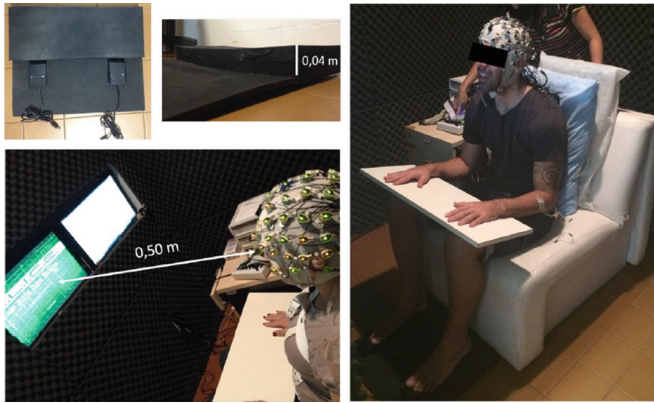


**Fig. 1.** Environment setting to perform ALICE game.

The hand presented have four features: laterality (left and right), orientation $(0°, 60°, 120°, 180°, 240°, 300°)$, view (back-view and palm-view), and posture (extension and flexion), in which the flexion posture was presented in three different conditions. These features are shown in Fig. 2 (A). So, there were 96 types of stimuli: left/right (2) × view (4) × orientation (6) × posture (2), but in this work, only trials with hand in back and palm-view were selected. The orientations $0°, 60°, 120°$ are classified as comfortable, due to being biomechanically easy to execute, and the orientations $180°, 240°, 300°$ are labeled as uncomfortable, being difficult to perform [12]. In total, 288 trials were performed by the participants, divided into three blocks. As seen in Fig. 2(B), the trial begins with a cross on the screen. 1500 ms later, the image of the hand is shown and stays on the monitor until the subject responds. Next, the robot gives the feedback

with a score based on RT. This feedback remains for 1000 ms and it is the final part of the trial.
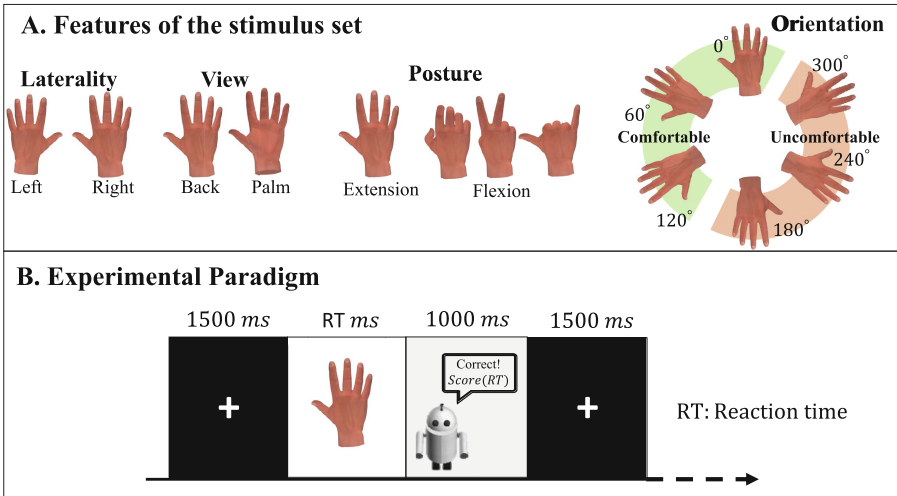


**Fig. 2.** A. Features of the stimulus. B. The experimental design of the game, involving the HLJT.

### 2.3   EEG Signal Acquisition and Preprocessing

The EEG signals were recorded at a sample rate of 1000 Hz with 64-channel ActiCap$^{TM}$ (Brain Products GmbH, Munich, Germany) and using the Vision Recorder (Version 1.20.0506, Brain Products GmbH, Munich, Germany). All electrodes were referenced to FCz, with impedance less than $10k\Omega$ and after EEG recording, the signals were re-referenced to the average of the left (TP9) and right (TP10) mastoids. Ocular artifacts were removed using the automated correction method of EOG based on regression analysis, using the BIOSIG toolbox [13], and three electrooculogram (EOG) electrodes were placed. The EEG data were digitally filtered with a 0.5–40 Hz band-pass FIR (Finite Impulse Response) filter and signal amplitude values exceeding $\pm100\mu$ V were automatically detected and rejected. The preprocessed EEG data were segmented into epochs defined in [−500 ms,1500 ms], i.e., 500 ms pre-stimulus and 1500 ms after the visual stimulus, that is the average of the RT for all subjects. The 500 ms pre-stimulus, of each trial, was used as baseline period, to compare the power changes after the visual stimulus.

### 2.4   ERD/ERS Estimation

To quantify the synchronization levels in alpha and beta bands, in decibel ($dB$), the following equation was used in each trial:

$$ERD/ERS = 10 \times \log_{10} \left( \frac{|F(f,t)|^2}{\left| \bar{F}(f,t)_{baseline} \right|^2)} \right) \qquad (1)$$

where $|F(f,t)|^2$ is the power extracted after the visual stimulus and immediately before the act of press down on the response pedal, and $\left| \bar{F}(f,t)_{baseline} \right|^2$ is the power during the rest interval (baseline period) [3]. The Welch's method [14] was used to estimate the periodogram, i.e., the power spectral density (PSD) [3].

To implement the method, the epoch $[-500\,ms, 1500\,ms]$ was segmented using a rectangular window with length of $N = 500$ samples and step of 250 samples. The segment of $N$ samples was divided in $L = 250$ parts, with $D = 125$ overlapping samples, following the Welch's method, and resulting in $x_k(l) = x(l + (K-1)D)$, where $l = 0, 1, ..., L-1$. So, we have $k = 1, 2, ..., K$ segments, i.e., $x_1(l), x_2(l), ..., x_K(l)$. For each $k = 1, 2, ..., K$ segment, the Fast Fourier Transform (FFT) $A_k(n)$ was extracted, with the signal windowed by Hanning function [15, 16]. Then, the modified periodogram given by:

$$I_k(f_n) = \frac{L}{U} |A_k(n)|^2 \qquad (2)$$

is estimated for each segment, where $f_n = n/L$ and $U$ is the window normalization factor, given by

$$U = \frac{1}{L} \sum_{l=0}^{L-1} w^2(l) \qquad (3)$$

where $w(l)$ is the Hanning window. As result, we have $K$ modified periodograms $I_k(f_n)$. Thus, a mean is calculated among these periodograms, resulting in

$$\widehat{P}_k(f_n) = \frac{1}{K} \sum_{k=1}^{K} I_k(f_n) \qquad (4)$$

i.e., the mean modified periodogram.

Trials with incorrect response, with hand movement detection during the RT [17, 18], without markers, with RT greater than 3500 ms or smaller than 500 ms [19], or EEG signal amplitude saturation ($\pm 100\mu V$) were automatically removed.

## 2.5   Region of Interest

The ERD/ERS patterns were analyzed in nine different regions: Full-montage, Reduced-montage, Simplified-areas, Frontal, Fronto-central, Central, Centro-pari et al, Parietal and Occipital (see Board 1 and Fig. 3). The goal was to investigate the activity of each brain region during the HLJT and the synchronization pattern.

Board 1: Regions of Interest and EEG channels selection.

| Region | EEG channels |
|---|---|
| Full-montage | 62 channels |
| Reduced-montage | 32 channels |
| Simplified-areas | F3, F4, FC3, FC4, C3, C4, CP3, CP4 P3, P4, PO3, and PO4 |
| Frontal | F1, F2, F3, F4, F5, and F6 |
| Fronto-central | FC1, FC2, FC3, FC4, FC5, and FC6 |
| Central | C1, C2, C3, C4, C5 and C6 |
| Centro-parietal | CP1, CP2, CP3, CP4, CP5, and CP6 |
| Parietal | P1, P2, P3, P4, P5, and P6 |
| Occipital | PO3, PO4, PO7, PO8, O1, and O2 |
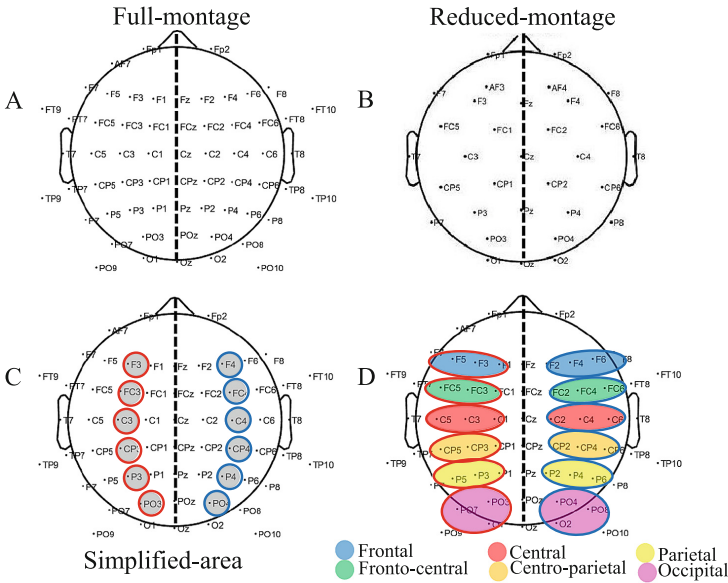


**Fig. 3.** Regions of Interest.

## 2.6  Statistical Analysis

Two statistic steps were performed, in order to confirm the main region(s) involved during the HLJT. In the first step, the main factors were HEMISPHERE (Left and Right), REGION (Frontal, Fronto-central, Central, Centro-parietal, Parietal, and Occipital), and FREQUENCY (alpha and beta), with the dependent variable being the mean ERSP. After the first step and following the results, the second analysis was performed including the hand features (LATERALITY, VIEW, ORIENTATION, POSTURE) and the average values of the ERSP, in

order to analyze the influence of the features on ERD/ERS values. To perform the statistic, the Generalized Estimating Equation (GEE) method was used [20, 21] and the significance level is $p = 0.05$.

## 3   Results

The ERD/ERS resulty curves, in $dB$, are shown in Fig. 4, for alpha and beta bands, respectively, including the features of the stimuli (orientation, view and posture), for each region of analysis (Board 1). The synchronization changes are more evident in alpha frequency than in beta frequency. An increase in alpha band is detected during the first 500 ms after the visual stimulus (see Fig. 5), following by an accentuated desynchronization, more evident in the alpha band and in parietal and occipital lobes, that include the central, parietal, and occipital regions, but also in frontal region. The ERD/ERS distribution during the HLJT, differently in explicit motor imagery, has a distributuion more uniform over the cortex, mainly over parietal and occipital lobes. Also, we can see in Fig. 5 that the desynchronization is present in both bands.



**Fig. 4.** Average of the alpha and beta synchronization patterns over trials and subject means, for each region of analysis and features of the stimulus.

**Fig. 5.** ERD/ERS topography characteristic during the **first** 500 *ms* **after the stimulus** of each hand that appears on the screen, in the alpha and beta bands; and the topography during the whole reaction time, for each hand laterality and frequency range.

The first statistical analysis showed significant main effect of REGION ($\chi^2\,(5) = 562.497$; $p < 0.001$), with the Centro-parietal and Parietal regions obtaining the largest mean ERD: $-1.80\,dB$ and $1.78\,dB$, respectively. These regions did not show any significant differences between them. The interaction HEMISPHERE × REGION and REGION × FREQUENCY were significant, but without differences of hemisphere and frequency to the same region. So, the factors HEMISPHERE and FREQUENCY were considered for the step.

In the second statistical analysis, there were no significant main effects on the HEMISPHERE, FREQUENCY, LATERALITY, VIEW, POSTURE, and ORIENTATION factors. But, there was significant main effect of HEMISPHERE × LATERALITY ($\chi^2\,(3) = 7.501$; $p = 0.05$) interaction, in which the right-hand judgment generated a higher ERD in the left hemisphere (mean $-2.02\,dB$) than in the right hemisphere (mean $-1.72$ dB).

## 4    Discussion

In the present work, we implemented a method to extract ERD/ERS patterns during the Hand Laterality Judgment Task - HLJT, that engages motor imagery processes [9]. The EEG data from 20 subjects was used. The results showed largest changes in the alpha band than in the beta band, as also reported by Chen *et al.* (2013). The increase in alpha rhythm during the first 500 ms of the reaction time is due to the visual stimulation, that enhances the activity over the posterior cortex. Also, the hand mental rotation is able to generate an activation over the post-central gyros (M1 region), superior and inferior parietal lobes, and in the primary visual cortex (occipital lobe) [22]. The synchronization of the frontal cortex is related to the attention so requested during the game performance [23]. Furthermore, during the implicit motor imagery, the distribution of the

ERD over the cortex is lower lateralized than in explicit motor imagery [2] and besides, is more uniformly distributed over the parietal and occipital regions. The Hand Laterality Judgment Task it is not an easy activity to perform, due to be a multifaceted task, due to involves skills as visual codification, mental rotate capability, judge and also the attention [24].

## 5    Conclusion

This paper implemented a method to extract the ERD/ERS patterns during the Hand Laterality Judgment Task, used by a rehabilitation gaming system. This task engages motor imagery processes in an implicit way since the stimulus is part of the body. The method proposed includes Welch's method using the Hanning window with a sliding approach. The results showed the synchronization levels in SMR, predominantly in the alpha band, and uniform distribution over the left and right hemispheres, in the centro-parietal and parietal brain regions. The method used to extract ERD/ERS and the experimental design can be used for BCI system.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Pfurtscheller, G., Silva, F.H.L.: Event-related EEG/MEG synchronization and desynchronization: basic principles. J. Clin. Neurophysiol. **110**, 1842–1857 (1999). https://doi.org/10.1016/S1388-2457(99)00141-8
2. Pfurtscheller, G., Neuper, C.: Motor imagery activates primary sensorimotor area in humans. Neurosci. Lett. **239**, 65–68 (1997). https://doi.org/10.1016/S0304-3940(97)00889-6
3. Hu, L., Zhang, Z. (eds.): EEG Signal Processing and Feature Extraction. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-9113-2
4. Schlögl, A., Lee, F., Bischof, H., Pfurtscheller, G.: Characterization of four-class motor imagery EEG data for the BCI-competition 2005. J. Neural Eng. **2**(4), L14–L22 (2005). https://doi.org/10.1088/1741-2560/2/4/L02
5. Freitas, D.R.R., Inocêncio, A.V.M., Lins, L.T., Santos, E.A.B., Benedetti, M.A.: A real-time embedded system design for ERD/ERS measurement on EEG-based brain-computer interfaces. In: Costa-Felix, R., Machado, J.C., Alvarenga, A.V. (eds.) XXVI Brazilian Congress on Biomedical Engineering. IP, vol. 70/2, pp. 25–33. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-2517-5_4
6. Padfield, N., Zabalza, J., Zhao, H., Masero, V., Ren, J.: EEG-based brain-computer interfaces using motor-imagery: techniques and challenges. Sensors **19**, 1423 (2019). https://doi.org/10.3390/s19061423

7. Hou, Y., Zhou, L., Jia, S., Lun, X.: A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN. J. Neural Eng. **17**, 016048 (2020). https://doi.org/10.1088/1741-2552/ab4af6

8. Mahamune, R., Laskar, S.H.: Classification of the four-class motor imagery signals using continuous wavelet transform filter bank-based two-dimensional images. Int. J. Imaging Syst. Technol. **31**, 2237–2248 (2021). https://doi.org/10.1002/ima.22593

9. Parsons, L.M.: Imagined spatial transformations of one's hands and feet. Cogn. Psychol. **19**, 178–241 (1987). https://doi.org/10.1016/0010-0285(87)90011-9

10. Osuagwu, B.A., Zych, M., Vuckovic, A.: Is implicit motor imagery a reliable strategy for a brain–computer interface?. IEEE Trans. Neural Syst. Rehabilitation Eng. **25**, 2239–2248 (2017). https://doi.org/10.1109/TNSRE.2017.2712707

11. Souza, R.F.: Influência da postura da mão na identificação da lateralidade: uma análise eletrofisiolágica (2016). Master's Thesis, Federal University of Rio Grande do Norte, Rio Grande do Norte, Natal (2016)

12. ter Horst, A.C., Van Lier, R., Steenbergen, B.: Mental rotation strategies reflected in event-related (de) synchronization of alpha and mu power. Psychophysiology **50**, 858–863 (2013). https://doi.org/10.1111/psyp.12076

13. Vidaurre, C., Sander, T.H., Schlögl, A.: BioSig: the free and open source software library for biomedical signal processing. Comput. Intell. Neurosci. **2011**, 935364 (2011). https://doi.org/10.1155/2011/935364

14. Welch, P.: The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust. **15**, 70–73 (1967). https://doi.org/10.1109/TAU.1967.1161901

15. Akin, M., Kiymik, M.K.: Application of periodogram and AR spectral analysis to EEG signals. J. Med. Syst. **24**, 247–256 (2000). https://doi.org/10.1023/A:1005553931564

16. Stoica, P., Moses, R.L.: Spectral Analysis of Signals. Pearson Prentice Hall, New Jersey. vol. 452 (2005)

17. Staude, G., Wolf, W.: Objective motor response onset detection in surface myoelectric signals. Med. Eng. Phys. **21**, 449–467 (1999). https://doi.org/10.1016/S1350-4533(99)00067-3

18. Staude, G., Flachenecker, C., Daumer, M., Wolf, W.: Onset detection in surface electromyographic signals: a systematic comparison of methods. EURASIP J. Adv. Signal Process. **2001**, 67–81 (2001). https://doi.org/10.1155/S1110865701000191

19. Ionta, S., Perruchoud, D., Draganski, B., Blanke, O.: Body context and posture affect mental imagery of hands. PLoS ONE **7**, e34382 (2012). https://doi.org/10.1371/journal.pone.0034382

20. Zeger, S.L., Liang, K.Y., Albert, P.S.: Models for longitudinal data: a generalized estimating equation approach. Biometrics **44**, 1049–1060 (1988). https://doi.org/10.2307/2531734

21. Hardin, J.W., Hilbe, J.M.: Generalized Estimating Equations. Chapman and Hall/CRC, New York (2002)

22. Kosslyn, S.M., DiGirolamo, G.J., Thompson, W.L., Alpert, N.M.: Mental rotation of objects versus hands: neural mechanisms revealed by positron emission tomography. Psychophysiology **35**, 151–161 (1998). https://doi.org/10.1111/1469-8986.3520151

23. Jueptner, M., Stephan, K.M., Frith, C.D., Brooks, D.J., Frackowiak, R.S.J., Passingham, R.E.: Anatomy of motor learning. I. Frontal cortex and attention to action. J. Neurophysiol. **77**, 1313–1324 (1997). https://doi.org/10.1152/jn.1997.77.3.1313
24. Tagaris, G.A., Kim, S.G., Strupp, J.P., Andersen, P., Uğurbil, K., Georgopoulos, A.: Mental rotation studied by functional magnetic resonance imaging at high field (4 Tesla): performance and cortical activation. J. Cogn. Neurosci. **9**, 419–432 (1997). https://doi.org/10.1162/jocn.1997.9.4.419

# Ultrasound Scan Converter Implemented on Xilinx Zynq-7000 All Programmable Systems-on-chip

Matheus Jose da Silva Ruzyk[1](✉), Amauri Amorin Assef[1,2],
Joaquim Miguel Maia[1,3], and Eduardo Tavares Costa[4]

[1] Graduate Program in Electrical and Computer Engineering (CPGEI), Federal
University of Technology - Paraná (UTFPR),, Curitiba, Brazil
`ruzyk@alunos.utfpr.edu.br`
[2] Academic Depart. of Electrical Engineering (DAELT), UTFPR, Curitiba, Brazil
[3] Biomedical Engineering Program (PPGEB) and Academic Department of
Electronics Engineering (DAELN), UTFPR, Curitiba, Brazil
[4] DEEB-FEEC & CEB, University of Campinas (UNICAMP), Campinas, Brazil

**Abstract.** Ultrasound imaging is widely used in medical diagnostics. One of its critical processing algorithms is scan conversion (SC). Although there is extensive research on ultrasound imaging, novel schemes for SC using System on Chips (SoCs) and Field Programmable Gate Arrays (FPGAs) are expected to improve image quality. This article aims to assess the SC image processing technique on a Xilinx Zed-Board development kit. This board has a Zynq-7000 All Programmable SoCs, which was used to execute the proposed SC algorithm with real ultrasound data from a multipurpose phantom. The results of the experimental data investigations are quantitatively and qualitatively compared with the results of a Matlab simulation. The image output is displayed on a monitor screen through the VGA port of the board. A significant agreement between the simulated and practical images was observed, with CR and CNR errors of 2.63% and 3.38%, respectively, and with minor artifacts caused by precision and interpolation differences. In conclusion, the proposed SC scheme was satisfactorily implemented on the ZedBoard, and future studies may explore other processing steps on this platform.

**Keywords:** Ultrasound · scan-conversion · FPGA · SoCs · image processing

## 1 Introduction

The scan conversion (SC) is one of the most important image processing steps to form ultrasound B-mode images, and it is necessary for compatibility with monitors when using convex array transducers. The technique is responsible for converting the coordinates of the acquired image from polar two-dimensional

(2D) datasets, in a sector of reference, to Cartesian coordinates of the target display. This conversion can be done through image processing algorithms running on computers or other processing platforms. The context of revisiting this and other well-known image processing techniques is to explore their performance and suitability to new hardware-based open platforms, that are idealized to allow implementation of new imaging algorithms and comply with attributes of easy reconfiguration, compact size, and realtime requirements to facilitate research purposes [5]. An example of a commercial convex array transducer used in biomedical applications is presented in Fig. 1 [6]. Its geometry allows a larger area to be scanned due to the elements placed on the convex surface.
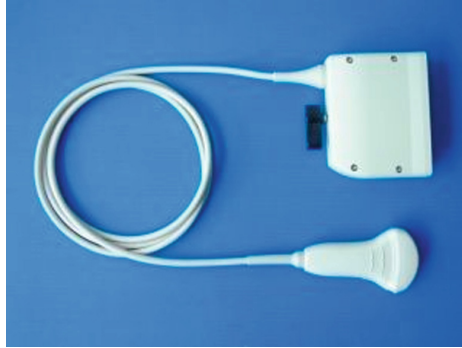


**Fig. 1.** Commercial convex array transducer AT3C52B (Broadsound Corp.,Taiwan) [6].

The coordinate conversion process is illustrated in Fig. 2, which shows a rectangular matrix composed of scanlines and samples transformed into a conical-shaped image. The scanline index is associated with an aperture angle for each transducer array element, and data samples are associated with the distance traveled by the echo. The process of SC resamples the polar-sample image into Cartesian format representing the path the waves have traveled, forming an undistorted image.

Implementation of SC uses a mathematical conversion of coordinates and interpolation to prevent discontinuities in the image on the cartesian points not covered by the original data. The principle of conversion is based on Equations (1) and (2), where $r$ and theta ($\theta$) are the polar coordinates. The first is the radial direction distance of the transducer, and the other is the angle between the center of the radial axis. The parameters $x$ and $y$ are the Cartesian coordinates [12,14].

$$r = \sqrt{x^2 + y^2}, \tag{1}$$

$$\theta = \tan^{-1}\frac{x}{y}, \tag{2}$$
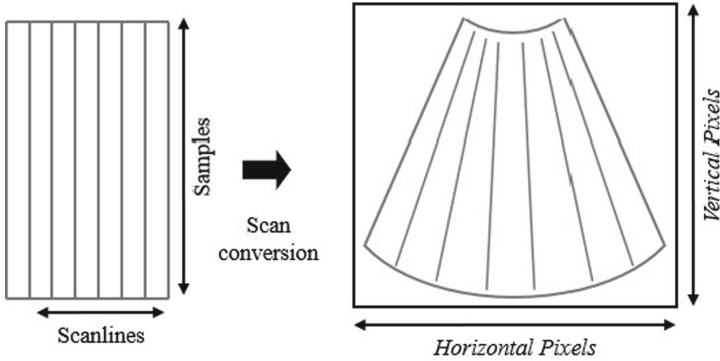
**Fig. 2.** Illustration of the SC processing for polar to Cartesian coordinates.

Various types of interpolation or image processing can fill the missing data on the cartesian coordinate system, and the most suitable will depend on the selected processing unit constraints [7,10]. Several authors have explored implementations of ultrasound image and signal processing techniques in various platforms [2,9,17]. For the cases where requirements of versatility and flexibility are necessary to test new algorithms, Field Programmable Gate Array (FPGA) or System on Chips (SoCs) are reliable choices [5,10,11]. One alternative is to use a ZedBoard Xilinx platform. The ZedBoard is a low-cost development board designed with an Artix-7 FPGA and a dual-core ARM Cortex-A9 processor that has been used to perform other ultrasound processing steps, such as beam-forming [3,18].

In this paper, we present the initial results of an SC algorithm implemented on a Xilinx ZedBoard. The design was implemented using Xilinx Intellectual Property (IP) blocks and C to be synthesized and programmed in the Vivado integrated platform and evaluated by comparing the output image with a Matlab simulation using cost function metrics.

## 2   Methodology

The validation study was performed by acquiring data from a tissue-mimicking phantom (model 84-317, Fluke Biomedical) and a 3.2 MHz 128-element convex array transducer (AT3C52B, Broadsound Corporation, Taiwan), as in [1,15]. Two B-mode ultrasound images were generated and evaluated using the same data set. The first is the reference image, generated by a Matlab script developed by our research group. The second image was implemented on the ZedBoard, and the processing ran in its Cortex-A9 processor.

### 2.1   MATLAB Simulation

The Matlab native functions are used to perform the SC of the input data. At first, two vectors corresponding to polar coordinates are calculated using the

transducer physical parameters, which are the convex angle between scanlines, and the distance in the axial direction estimated according to the medium characteristics. These vectors are used to create a tridimensional profile with the function *meshgrid*. Next, the function *pol2cart* is used to convert these vectors to cartesian coordinates. Finally, the function *surf* is used to plot a tridimensional surface that associates the calculated profile with the brightness values of the empirical polar input matrix. The result shows the scan-converted image when projected in a two-dimensional perspective.

## 2.2   Implementation on ZedBoard

ZedBoard is used to process and display an image and send its data to a computer for further analysis. The SoC is implemented using IP hardware blocks of the software Vivado 2020.2. In addition, the software is written in C with the program Vitis 2020.2 to perform the SC and control the peripherals. Data processing in ZedBoard begins with a lookup table (LUT) calculation to store cartesian coordinates that already express a conic form and yield the final positioning at the image matrix. Next, the values of echo are plotted in the image matrix addressed by the LUT coordinates, and at the same time, pixel values are determined using the nearest neighbor interpolation. These steps are processed in a microcontroller inside the SoC, and results are available through peripherals. The overall hardware architecture of the proposed system is shown in Fig. 3 and presents the SoC and the peripherals used.

The SoC is divided into the programmable logic (PL) and the processing system (PS). The PS controls the UART, GPIO, and memories and is responsible for processing and transferring the image from the DDR memory to the PL, with the aid of the direct memory access controller (DMA) and its version for video purposes, VDMA. It moves data from the memory to the processor and from this to the VDMA controller using the AXI4 bus. Both receive commands through the AXI4-Lite interface and are configured by the programmer using libraries given by Xilinx. The PL generates the signals needed to operate the 4-bit VGA interface available in ZedBoard. It comprises a video timing controller that operates at 25.2MHz to generate the required synchronization signals for a 480p resolution image and a stream to video IP used to adequate the data stream received from the PS. This stream, present in the AXI4 bus, is transformed into a vector ready to be sent to the VGA digital-to-analog converter (DAC).

The PS programs run in the Cortex-A9 bare metal, with a clock frequency of 669 MHz, and are programmed in C. It performs the SC and the nearest-neighbor interpolation with the beamformed data, saved in its DDR memory as a *string of unsigned characters*, displays the results as a 480p VGA image, and sends the processed data to a computer through the UART interface at a baud rate of 115200 bps. This resulting image has a resolution of $500 \times 320$ pixels, expressed in 160.000 bytes.
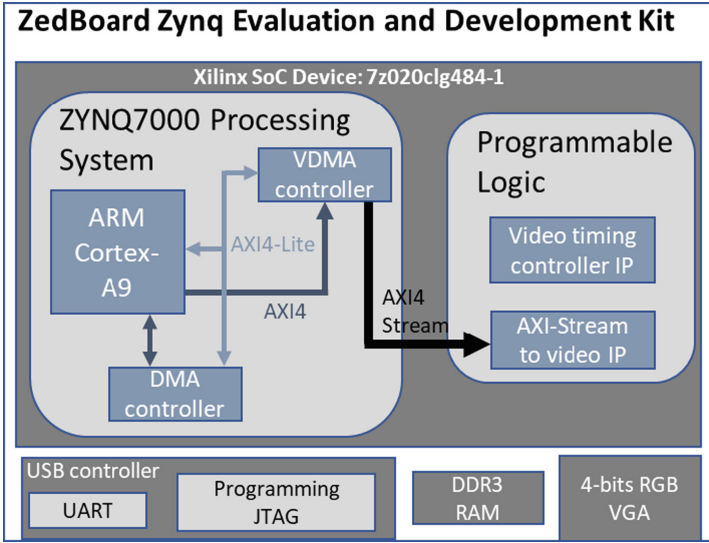
**Fig. 3.** System architecture implemented on ZedBoard.

### 2.3   Cost Functions to Evaluate Resulting Images

This work uses two cost functions: the contrast-to-noise ratio (CNR) and contrast resolution (CR). They are defined by Equations (3) and (4), respectively, where $\mu_t$ and $\mu_b$ represent the mean intensity regions of the image in dB grayscale, first at a point of interest and the other at the background. The other variable, denoted by $\sigma_b$, is the background intensity standard deviation

$$CR = 1 - \frac{\mu_t}{\mu_b}, \tag{3}$$

$$CNR = \frac{|\mu_t - \mu_b|}{\sigma_b}, \tag{4}$$

## 3   Results

Radiofrequency (RF) data acquired by the ULTRA-ORS platform, which means Ultrasound Open Research System, and beamformed in Simulink, both explained in detail in previous works [1,15], are used to evaluate the proposed hardware-based SC. The data acquisition parameters are summarized in Table 1.

The overall experimental setup is described in the diagram of Fig. 4, which denotes that the acquisition and beamforming steps were done previously, providing a matrix of 121 channels by 2046 samples to be scan-converted both on ZedBoard and Matlab. The first uses a data precision of eight bits, and the other double for data processing. The VGA is limited to a 4-bit precision for each pixel in this board.

**Table 1.** Parameters used to acquire de RF data.

| Item | Notation | Value | Units |
|---|---|---|---|
| Speed of sound | $c$ | 1540 | m/s |
| Sampling frequency | $fs$ | 40 | MHz |
| Number of samples per channel | $N$ | 2046 | - |
| Number of transducer elements | $Ne$ | 128 | - |
| Aperture | $n$ | 8 | - |
| Kerf | $kerf$ | 115 | $\mu$m |
| Element width | $D$ | 0,41 | mm |
| Radius of curvature | $R_{convex}$ | 40 | mm |

After beamforming, the RF data supplied to ZedBoard can be first displayed but still not in its final form. The data presented in polar coordinates is shown in Fig. 5, with a dynamic range of -40 dB. Next, the data is processed both in a ZedBoard setup and in a computer running the Matlab script, with results presented in the following sub-sections.

## 3.1   ZedBoard Setup

The scan-converted image is satisfactorily exhibited on a VGA display connected to the ZedBoard, as presented in Fig. 6. Although a brightness span degradation happens due to the mentioned VGA low resolution of the platform, the main details of the image are visible, and it complies with the expected conical shape. In addition, the 8-bit resolution data processed is sent successfully by USB in the same size packages to the host computer. Small data packages prevented data losses due to buffer congestions on the USB-emulated serial port.

Table 2 presents the utilization of the resources of the FPGA to implement the VGA controller and peripheral controls.

**Table 2.** Resource utilization of implemented scan converter on Zynq-7000 SoC.

| Resource | LUT | LUT RAM | FF | BRAM | MMCM | DSPs |
|---|---|---|---|---|---|---|
| Utilization | 1711 | 133 | 2506 | 3.5 | 1 | 0 |
| % | 3.22 | 0.76 | 2.36 | 2.5 | 25 | 0 |

## 3.2   Comparison of Results Between Matlab and ZedBoard

The scan-converted results of the Matlab processing and ZedBoard are presented in Fig. 7a and Fig. 7b, respectively. The two resulting images are similar in shape
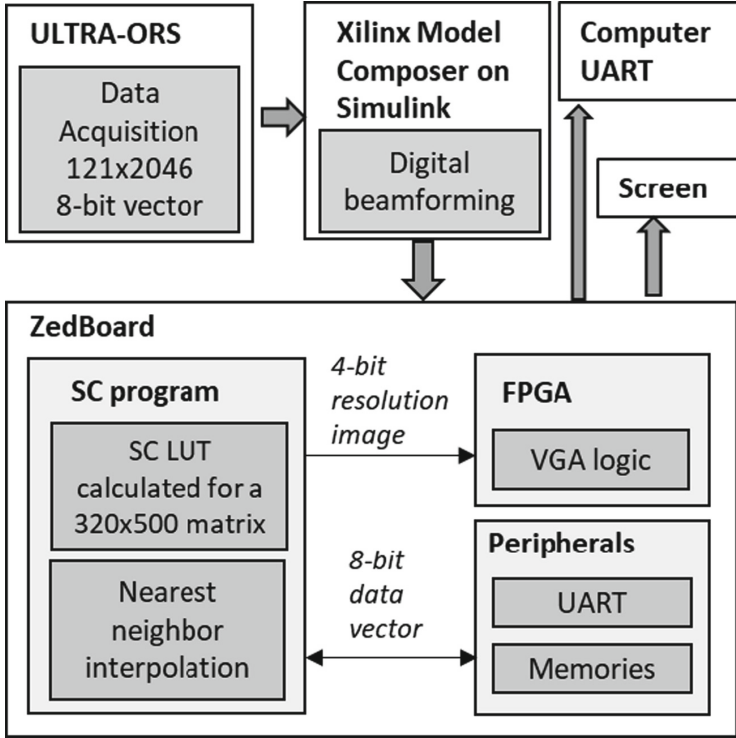
**Fig. 4.** Experimental setup showing the data transfer between modules.

and over the most prominent characteristics. However, there is an expected divergence due to the precision differences between the two platforms, making Matlab the more efficient and, therefore, the golden standard. In addition, another factor that influences the results is that Matlab uses a linear interpolation form through the function *surf*. In contrast, in ZedBoard, a simpler but less effective method, the nearest neighbor interpolation, is used, resulting in an increased granularity in higher depths.

In addition to the qualitative analysis, Table 3 compares the two resulting images using CR and CNR metrics in the regions defined in Fig. 7. The obtained simulation and experimental results are in excellent agreement, with errors smaller than 4% for both CR and CNR metrics.

**Table 3.** Image evaluation metrics.

| Metric | MATLAB | ZedBoard | Error (%) |
|--------|--------|----------|-----------|
| CR     | 0.0924 | 0.0900   | 2.6295    |
| CNR    | 1.9777 | 1.9108   | 3.3812    |
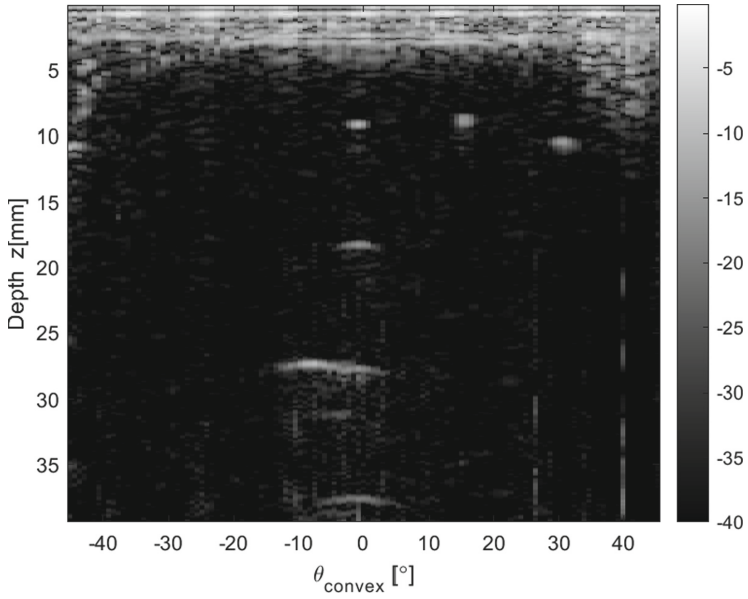
**Fig. 5.** Beamformed data in polar coordinates.
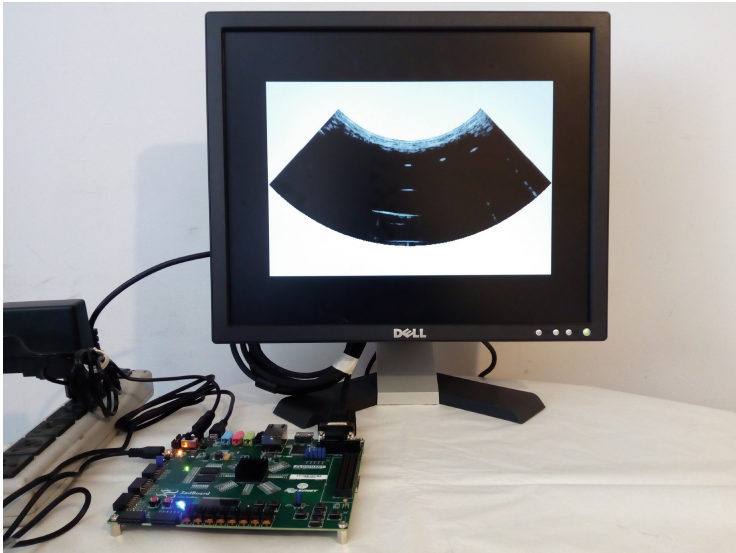


**Fig. 6.** Scan conversion result displayed by the ZedBoard through a VGA display. The data results are sent through a USB interface to the host computer.
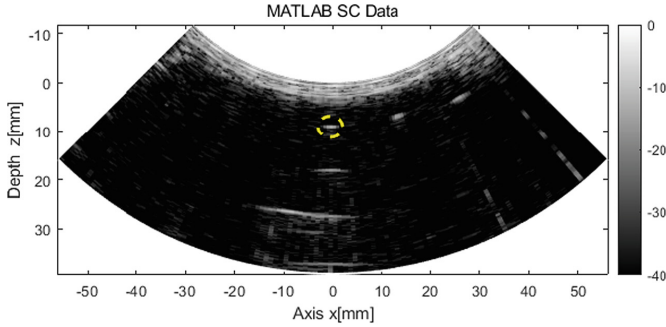
(a) Matlab simulation resulting image.
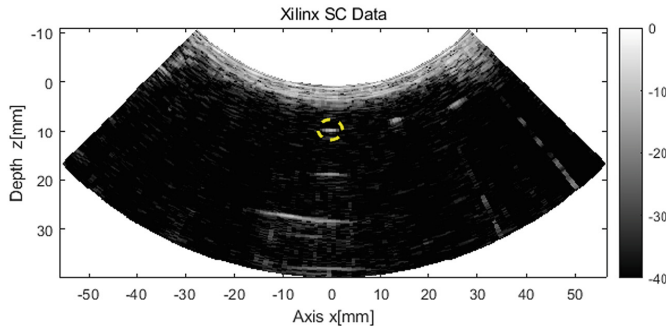


(b) ZedBoard resulting image.

**Fig. 7.** Scan conversion results. (**a**) Matlab simulation and (**b**) ZedBoard hardware-based implementation. Both images have yellow dashed circle regions assigned for metrics calculation.

## 4   Discussion

The scan conversion is successfully implemented in the ZedBoard, using its processor to execute the algorithm and command both peripherals and programmable logic. It is part of the effort to test image processing techniques in a modular way to explore new back-end processing platforms [5,13]. The achieved results show that the platform can produce images with a quality close to a high-end processing environment, and, in addition, this is proven by quantitative analysis. Other works explore ZedBoard as a platform for general control of front-end ultrasound modules as well as a back-end processing module and test SC approaches. These works have processed the SC using the LUT approach for coordinate transformation and bilinear interpolation to fill scan line gaps [3]. In another work that aimed at portability in a point of care design, the authors [4] did the SC using another approach, through the CORDIC algorithm, that is part of Xilinx's IP resources in Vivado, with $2 \times 2$ linear interpolation, achieving the proposed qualitative result. In comparison, the implementation done in this article has used the ZYNQ processing system to perform all the needed calculations

of the SC and added steps such as the image display directly from ZedBoard to the screen through VGA and also showed an additional quantitative comparison, even though a simpler interpolation technique was used here.

In future research, results can be compared with other implementations, such as the SC in hardware, as proposed by [10], and include other processing steps of digital beamforming. Furthermore, the use of the ZYNQ7000 allows a comparison between softcore processors, the in-built ARM Cortex A9 and FPGA implementations, which can give insights on how to best allocate resources when using this as a US central processing platform [8]. Finally, timing comparisons of algorithms performance in ZYNQ and other platforms can also be done with SC and other techniques to evaluate the timing requirements for US real-time imaging [16].

## 5   Conclusion

Scan conversion is fundamental for displaying ultrasound images acquired with convex transducers in a correct geometry. In this work, two implementations of SC were compared, one in Matlab and the other in the ZedBoard. The results show that the image generated in the platform is similar to the golden standard of Matlab, with CR and CNR errors of 2.63% and 3.38%, respectively, although expected minor deviations occur due to precision differences between the systems and the use of different interpolation approaches. Future works may explore other image display methods, such as HDMI, and execute the SC calculations on FPGA instead of using the SoC embedded processor to pursue real-time imaging.

**Compliance with ethical standards**

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Assef, A.A., Maia, J.M., Costa, E.T.: Initial experiments of a 128-channel FPGA and PC-based ultrasound imaging system for teaching and research activities. In: Proceedings 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), vol. 2016-Oct, pp. 5172–5175 (2016). https://doi.org/10.1109/EMBC.2016.7591892

2. Batbayar, E., Tumenjargal, E., Song, C., Ham, W.: Implementation of an ultrasound platform for proposed photoacoustic image reconstruction algorithm. In: 2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 293–298. IEEE (2018). https://doi.org/10.1109/BIBE.2018.00064

3. Bharath, R., Kumar, P., Reddy, D.S., Rajalakshmi, P.: Compact and programmable ultrasound front-end processing module for research activities. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), vol. 2018-July, pp. 921–924 (2018). https://doi.org/10.1109/EMBC.2018.8512343

4. Bharath, R., Reddy, D.S., Kumar, P., Rajalakshmi, P.: Novel architecture for wireless transducer based ultrasound imaging system. In: 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), vol. 8, pp. 432–436. IEEE (2016). https://doi.org/10.1109/IECBES.2016.7843487

5. Boni, E., Yu, A.C., Freear, S., Jensen, J.A., Tortoli, P.: Ultrasound open platforms for next-generation imaging technique development. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **65**(7), 1078–1092 (2018). https://doi.org/10.1109/TUFFC.2018.2844560

6. Broadsound, C.: AT3C52B (C5-2 for Verasonics) (2022). https://www.broadsound.com.tw/replacement-probe-139-product366

7. Chang, J.H., Yen, J.T., Shung, K.K.: High-speed digital scan converter for high-frequency ultrasound sector scanners. Ultrasonics **48**(5), 444–452 (2008). https://doi.org/10.1016/j.ultras.2008.03.001

8. Gilliland, S., Govindan, P., Gonnot, T., Saniie, J.: Performance evaluation of FPGA based embedded ARM processor for ultrasonic imaging. In: 2013 IEEE International Ultrasonics Symposium (IUS), pp. 519–522. IEEE (2013). https://doi.org/10.1109/ULTSYM.2013.0135

9. Hassan, M.A., Kadah, Y.M.: Digital signal processing methodologies for conventional digital medical ultrasound imaging system. Am. J. Biomed. Eng. **3**(1), 14–30 (2013). https://doi.org/10.5923/j.ajbe.20130301.03

10. Kassem, A., Hamad, M., Sawan, M.: An efficient SoC dedicated to ultrasonic digital imaging systems. In: 2006 6th International Workshop on System on Chip for Real Time Applications IWSOC, pp. 165–168 (2006). https://doi.org/10.1109/IWSOC.2006.348229

11. Kassem, A., Sawan, M., Boukadoum, M.: A New Digital Scan Conversion Architecture For Ultrasonic Imaging Systems. J. Circuits, Syst. Comput. **14**(02), 367–382 (2005). https://doi.org/10.1142/S0218126605002374

12. Li, X.: Ultrasound Scan Conversion on TI's C64x+ DSPs. Application Report (March), pp. 1–15 (2009). https://www.ti.com/lit/an/sprab32/sprab32.pdf

13. Martin-Arguedas, C.J.: A SoC-based fully configurable point-of-care ultrasound system for research purposes. In: 2019 IEEE International Ultrasonics Symposium (IUS), vol. 2019-Oct, pp. 815–817. IEEE (2019). https://doi.org/10.1109/ULTSYM.2019.8926191

14. Robinson, D.E., Knight, P.C.: Interpolation scan conversion in pulse-echo ultrasound. Ultrason. Imaging **4**(4), 297–310 (1982). https://doi.org/10.1177/016173468200400401

15. Ruzyk, M.J.S., Assef, A.A., Maia, J.M., Oliveira, J., Souza, M.A., Costa, E.T.: Modelling and simulation of an ultrasound reception beamforming using Xilinx model composer. In: Proceedings 17th Symposium on Medical Information Processing and Analysis, vol. 12088. SPIE (2021). https://doi.org/10.1117/12.2606292

16. Vasudevan, V., Wang, B., Govindan, P., Saniie, J.: Design and evaluation of reconfigurable ultrasonic testing system. In: 2015 IEEE International Conference on Electro/Information Technology (EIT), vol. 2015-June, pp. 310–313. IEEE (2015). https://doi.org/10.1109/EIT.2015.7293359

17. Viswanath, K., Gunasundari, R.: Analysis of Kidney Stone Detection by Reaction Diffusion Level Set Segmentation Using Xilinx System generator on FPGA. VLSI Des. **2015**, 10 (2015). https://doi.org/10.1155/2015/581961

18. Wang, B., Saniie, J.: Ultrasonic Signal Acquisition and Processing platform based on Zynq SoC. In: 2016 IEEE International Conference on Electro Information Technology (EIT), vol. 2016-Aug, pp. 448–451. IEEE (2016). https://doi.org/10.1109/EIT.2016.7535282

# Motif Synchronization and Space-Time Recurrences for Biometry from Electroencephalography Data: A Proof-of-Concept

Manuela V. A. Davanço[1] , Marina C. de Paulo[1] , Paula G. Rodrigues[2] ,
Diogo C. Soriano[2,3] , and Gabriela Castellano[1,3(✉)]

[1] Gleb Wataghin Institute of Physics (IFGW), Universidade Estadual de Campinas
(UNICAMP), Campinas, SP, Brazil
gabriela@ifi.unicamp.br
[2] CECS, Federal University of ABC, São Bernardo, SP, Brazil
[3] Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, SP, Brazil

**Abstract.** The electroencephalography (EEG) technique has the capability of identifying individual traits. Previous work has already used functional connectivity (FC) features obtained from EEG data for biometric purposes. In this work, we explored two FC methods not yet used in this context: motifs synchronization and space-time recurrences. Fifty subjects with two resting-state EEG acquisitions (one with eyes open and another with eyes closed) were included in the study. FC matrices for 1 s and 5 s epochs were computed for each acquisition. Subject's identification was sought by comparing the FC matrices from both acquisitions using the Pearson correlation coefficient. The motifs method achieved 48% accuracy for both epoch sizes, and the space-time recurrences achieved 36% and 38% accuracies for 1 s and 1 s epochs respectively. Although the accuracies were low, they were well above the 2% chance level. Also, unlike other similar studies, the comparison was made between signals acquired in different conditions. In general, the obtained low accuracies illustrate the challenging problem of performing biometry from EEG and the need for further adjustments in the feature extraction and classification stages.

**Keywords:** Biometry · EEG · motifs synchronization · recurrence quantification · functional connectivity

## 1 Introduction

The Electroencephalography (EEG) technique has been used for several clinical applications, particularly for diagnosis of some neurological diseases such as epilepsy [1] and sleep disorders [2]. Former studies have also shown that this technique provides information about differences between individuals related to anatomical and functional

---

M. V. A. Davanço and M. C. de Paulo—These authors contributed equally to this work.

brain traits [3, 4]. More recently, the idea of using EEG signals to distinguish between different individuals aiming to implement a biometrical system [5, 6] has been explored in greater depth in several works [7–9], usually considering resting state acquisitions [10–12]. Indeed, the resting state paradigm has the benefit of enabling the use of data ac-quired from any individual, including those with restricted mobility, as well as diminishing the problem of movement artifacts, since the person must be still during the EEG evaluation.

Among the studies that aim to obtain biometric information from EEG signals are those that have used characteristics extracted from specific electrodes, and those that have explored the relationships between EEG signals obtained by different electrodes, a method known as brain connectivity. Connectivity-based approaches assume that many brain functions are executed under a specific engagement of different brain regions [13] or networks. Thus, the understanding of how these interactions take place may play a key role in providing additional information regarding the individual.

Functional connectivity (FC) is a data-driven, exploratory method, which seeks to find similarities between the dynamics of different brain regions. This method establishes relationships between the regions, which can be analyzed through graph theory [14]. For that, the graph nodes and a measure of similarity to provide the links between the nodes must be defined. For EEG data, the nodes are usually chosen to be the electrodes. In the context of using FC obtained from EEG to identify individuals, some similarity measures have already been explored, such as spectral coherence [9, 15, 16], the Spearman correlation applied to the Hilbert transform of the time series [17], the phase-locking value [17, 18], the imaginary part of the phase-locking value [19], the phase lag index or phase lag index [15], and mutual information applied to ordinal patterns [20], among others.

This work aims to analyze resting-state EEG under the perspective of graph-based measures to identify individuals. Two different FC similarity measures are used here: spatial-temporal recurrences [21] and motifs synchronization [22] based on ordinal patterns [23, 24]. It is important to stress that, to the best of our knowledge, neither type of similarity measure has yet been used in this context.

This article is divided as follows: Sect. 2 presents the EEG database used, the preprocessing steps and the two similarity methods used to evaluate the FC; Sect. 3 presents the identification results; and Sect. 4 presents a discussion about these and a brief conclusion of the work.

## 2 Materials and Methods

Figure 1 shows a flowchart summarizing the signal processing pipeline adopted in this work, including the chosen database, the preprocessing steps, the feature extraction approach, and the methodological analysis. All these steps are described in more detail in the following sections.

### 2.1 Database and Preprocessing

The EEG data used was from the Physionet database [25, 26], in which data from 109 subjects were recorded using a 64-channel EEG BCI2000 system, with electrodes
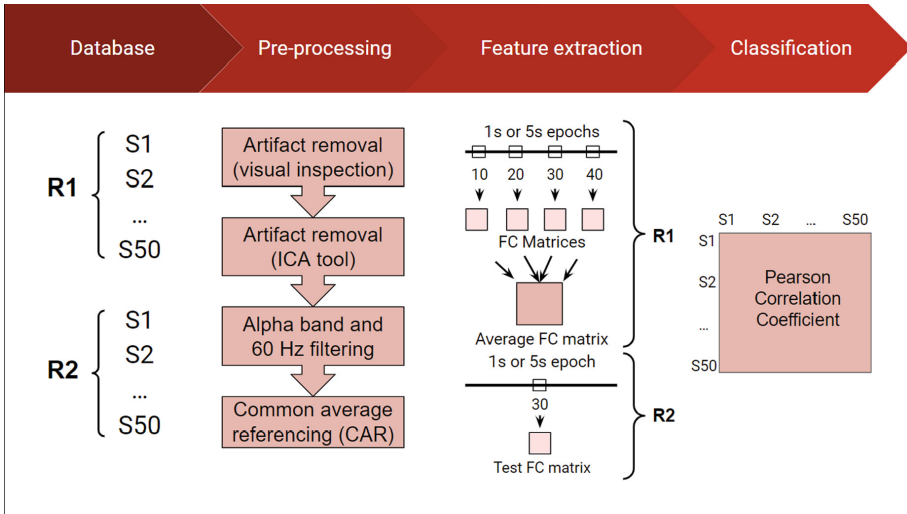
**Fig. 1.** Signal processing pipeline. The database consisted of 50 subjects with two EEG acquisitions each (R1 and R2). These were preprocessed and 1 s or 5 s epochs were extracted from the time series (four epochs from R1 and one epoch from R2). FC matrices were calculated from these epochs, using the motifs synchronization and space-time recurrences methods. A mean FC matrix from R1 was used as reference and the FC matrix from R2 was used as test. These were compared among all subjects using Pearson's correlation coefficient.

positioned following the 10–10 system (Fig. 2). The subjects performed 14 experimental runs. The first two runs were acquired in resting state, during one minute each, with eyes open (R1 – condition) and closed (R2 - condition), respectively. These runs were used in the analysis performed here.

After downloading the data in EDF format, the preprocessing was performed using EEGLAB [27], and consisted of four steps: first, removal of artifacts by simple inspection; second, decomposition of the data using Independent Component Analysis (ICA) and removal of undesired components; third, removal of alpha and power grid frequencies; fourth, Common Average Referencing (CAR) of the data [28].

In order to remove more blatant artifacts, the "*Inspect/Reject data by eye*" tool was used, in which sections of the signals could be marked for removal. Sections that had greater (at least five-fold) amplitude than the rest of the signal were removed.

Then, the signal was decomposed into independent components (ICA), using the tool "*Decompose Data by ICA*". This tool displays the obtained independent components through scalp map projections of the EEG activity. Components related to muscle movements, eye blinks and other eye movements can be easily recognized, and were thus removed.

Next, the alpha band was removed, using a stop band filter (with the "*Basic FIR Filter*" tool of EEGlab, considering an interval of 7 Hz to 13 Hz, and selecting the option "*Notch filter the data instead of pass band*"). This was done because we wanted to compare signals obtained from eyes closed and eyes open paradigms, and this band is known to be strikingly different between opened and closed eyes signals. Finally,

the signal was bandpass filtered (again with the same tool, but without selecting the option "*Notch filter the data instead of pass band*") between 4 and 50 Hz, to eliminate low-frequency artifacts and high-frequency noise.
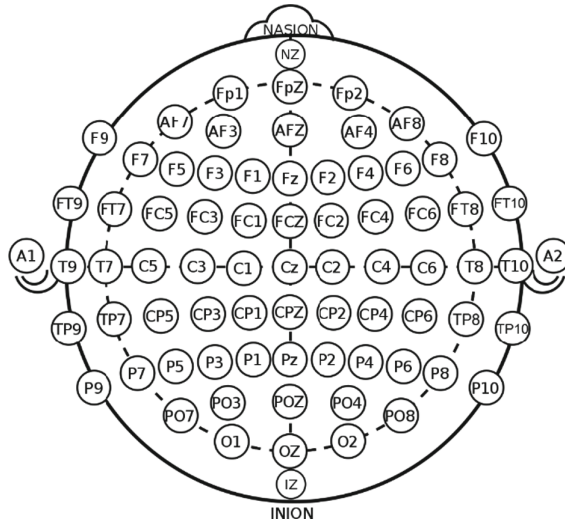


**Fig. 2.** 10–10 electrode positioning system. Obtained from https://upload.wikimedia.org/wikipedia/commons/3/38/International_10-20_system_for_EEG-MCN.png. Author: Brylie Christopher Oxley.

The final preprocessing step implied in a spatial filter for re-referencing the signals using CAR [28]. This method consists in calculating the mean of the signals over electrodes and then subtracting this value from each electrode signal.

## 2.2 Functional Connectivity Matrices

All the database was preprocessed using the four steps aforementioned, however, only data from 50 subjects were used in this work. These subjects were selected considering the duration of the acquisitions after preprocessing. Sub-jects with acquisitions with less than 45 s were discarded.

From the R1 acquisition, four epochs were extracted, starting at seconds 10, 20, 30 and 40. From the R2 acquisition, only one epoch was extracted, starting at second 30. Lengths of 1 s and 5 s were tested for these epochs. Then, FC matrices were computed for both R1 and R2 epochs for all subjects, to be used as features in the identification problem. A template matching approach was used, in which the R1 matrices were further averaged to give one reference FC matrix per subject, while the R2 matrix was used as a test sample.

Two different similarity methods were used to compute the FC matrices: motifs synchronization [22] and space-time recurrences [21]. Both methods were implemented in MATLAB (2018, Natick, Massachusetts: The MathWorks Inc). These methods are detailed in the following.

**Motifs Synchronization.** A motif series is basically a series of behavior patterns in the EEG signal. In this work, motifs with three points were used, as in Fig. 3. Thus, a temporal series of an EEG electrode can be "translated" into a motif series, according to the types of motifs in the signal.
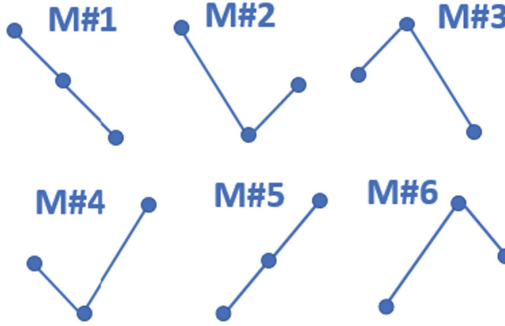


**Fig. 3.** Three-point motifs used in this work.

The motif series of two electrodes can be used to evaluate the similarity between signals considering different lag values. In this work, a lag $t = 0$ was used. Mathematically, the similarity between the motif series of electrodes $i$ and $j$ can be calculated using the coefficient $c_{ij}$, defined as follows [22]:

$$c_{ij} = \sum_{k=1}^{L_M} J_k \tag{1}$$

where $L_M$ is the motif series length, $J_k = 1$ if the motif at position $k$ is the same in both series, and $J_k = 0$ otherwise.

Then, the degree of synchronization between electrodes $i$ and $j$ is calculated:

$$Q_{ij} = \frac{c_{ij}}{L_M} \tag{2}$$

which varies from 0 to 1.

With that, an $N \times N$ connectivity matrix is obtained, where $N$ is the number of electrodes used for the acquisition (in this work, $N = 64$), and each element of the matrix is the degree of synchronization between the row electrode and the column electrode.

**Space-Time Recurrences.** Space-time recurrences is a method used to identify whether a system has returned to a previous configuration during a given time period [29].

The space-time recurrence between two time series $x_i$ and $x_j$ is defined as:

$$STR_{i,j}(\varepsilon, n) = \Theta\big[\varepsilon - \big|x_i(n) - x_j(n)\big|\big] \tag{3}$$

The structure $STR$ is called the space time recurrence matrix: a tridimensional data structure of $N \times N \times N_s$, with $N$ being the number of channels (or electrodes; in this case,

$N = 64$); and $N_s$ the total number of samples in the chosen time frame (e.g. $N_s = 160$ for 1 s frames or $N_s = 800$ for 5 s frames, since the sampling rate was 160 Hz). $\Theta$ is the Heaviside function, therefore: $\Theta(x) = 0$ if $x < 0$ and $\Theta(x) = 1$ if $x \geq 0$. Finally, $\varepsilon$ is an arbitrary distance threshold. In the present work, we chose $\varepsilon = 50\%$ of the maximum distance ($|x_i(n) - x_j(n)|$) between electrode time series.

From the *STR* it is possible to calculate the connectivity matrix, which consists in normalizing the sum of the values of each electrode pair in the $STR_{i,j}$ structure:

$$A_{i,j} = \frac{1}{N} \sum_{n=1}^{N} STR_{i,j}(\varepsilon, n) \tag{4}$$

Thus, it is possible to reduce the dimension of the problem, since $A_{i,j}$ is a two dimensional $N \times N$ matrix. Each element of the matrix describes the similarity between two temporal series of EEG.

### 2.3   Comparison by Pearson Correlation Coefficient

To evaluate the similarity among the signals, and thus identify a given subject, the Pearson correlation coefficient was calculated between the mean R1 (eyes open) matrix and the R2 (eyes closed) matrix of all subjects.

If the highest correlation value was for R1 and R2 of the same individual, it was possible to identify the person, because it indicated greater similarity between different acquisitions of the same person. If not, it was not possible to identify the person.

Finally, the methods were compared in terms of their hit rate, or accuracy (i.e., percentage of correctly classified individuals).

## 3   Results

Table 1 shows the accuracy values obtained for subject identification, for each FC method and epoch length.

**Table 1.** Subject identification accuracies for all combinations of functional connectivity methods and epoch lengths.

| Epoch length | Motifs synchronization | Space-time recurrences |
| --- | --- | --- |
| 1 s | 48% | 36% |
| 5 s | 48% | 38% |

Using motif synchronization, for both epoch lengths (1 s and 5 s), 24 individuals were correctly identified among 50 analyzed, which corresponds to 48% of accuracy in both cases. Interestingly, the epoch length did not seem to influence the performance of this method.

The space-time recurrences method was able to correctly identify 18 out of 50 individuals for FC matrices computed using 1 s data, which corresponds to 36% accuracy, and for 5 s matrices it could identify 19 individuals among 50, which corresponds to a 38% accuracy. Therefore, the results using 5 s epochs to compute the FC matrices were slightly better.

## 4 Discussion

Regarding a comparison between methods, the motifs method achieved better accuracy than the space-time recurrences method, for all epoch lengths. This indicates that the motifs method was more capable than the recurrences method of extracting relevant information from the EEG signals regarding individual traits. The motifs method has been shown to be more efficient than other usual EEG FC methods, such as mean squared coherence and imaginary coherence, for extracting relevant information regarding interictal epileptiform discharges [30].

Nevertheless, the accuracies obtained with both methods used here, for all epoch lengths, were too low for practical purposes. Indeed, in [9], La Rocca and colleagues achieved up to 100% recognition rates with this same database, using features obtained both from power spectral density (PSD) and from coherence-based connectivity. They looked at individual electrode PSD features and individual channel (electrode pair) coherence features, and then combined the features from a given region (e.g., central, parietal or frontal) and fed them to a classifier based on the Mahalanobis distance. However, it is important to note that they only compared epochs within a given acquisition (eyes open or eyes closed); they did not attempt to use one acquisition to predict the other, as done here.

This work has several limitations. The number of subjects was low for the type of application (biometry). Notwithstanding, it is important to note that when the number of subjects is increased, the rate of accuracy decreases, since more comparisons are being made and the chance that there will be a correlation coefficient smaller than that of the right person increases. We previously tested the method with a sample of 11 subjects and the accuracies were indeed much better (64% for both methods).

The number of acquisitions per subject was also low, and additionally, the two acquisitions used did not follow exactly the same conditions, since despite both being in resting-state, one was acquired with eyes open and the other with eyes closed. Closing one's eyes is known to increase the amplitude of alpha band oscillations in EEG signals. In a first analysis (not reported here) we attempted to use these different signals without subtracting the alpha band, but the results were worse than the ones reported here.

Also, the first preprocessing step (artifact removal by visual inspection) is somewhat subjective and may not have been exactly the same for all signals. Additionally, the STR requires adjusting the recurrence threshold for optimum FC evaluation [22] and a further detailed analysis considering a specific dataset for hyperparameter tuning outlines a natural perspective.

Nevertheless, it is important to highlight that the method presented may be taken further by exploring different options in each step of the methodology. Preprocessing could benefit from an automatic artifact removal algorithm such as SOUND [31], which

would take away the subjectivity of removing signal stretches and ICA components by simple inspection. Also, other types of referencing methods, such as REST [32], could be tried instead of CAR. In the feature extraction step, graph parameters computed from the FC matrices could be explored. The motifs method could be improved by looking at delays other than zero, as in [22], while STR can be improved by means of threshold adaptations. Finally, in the classification step, a very simple classification method was used, namely, the Pearson correlation coefficient, but comparatively more sophisticated classification approaches could be investigated, such as Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) or even deep neural networks.

In conclusion, both methods of FC calculation, motif synchronization and space-time recurrences, produced results that remained below what would be considered an accurate pattern of subject identification. That said, these results were highly above the chance level (which, for 50 subjects, would have been 2%), showing that the methods have potential for this application. Also, our results were obtained attempting to match two signals acquired in different moments, while other works in the literature using similar approaches have compared only signal epochs within the same acquisition (and condition). Finally, this was a pilot study, which aimed to explore the use of two FC measures that, to the best of our knowledge, had not yet been applied to biometry studies based on EEG data.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. Koutroumanidis, M., et al.: The role of EEG in the diagnosis and classification of the epilepsy syndromes: a tool for clinical practice by the ILAE Neurophysiology Task Force (Part 1). Epileptic Disord. **19**(3), 233–298 (2017)
2. Campbell, I.G.: EEG recording and analysis for sleep research. Curr. Protoc. Neurosci. **49**(1), 10–2 (2009)
3. Berkhout, J., Walter, D.O.: Temporal stability and individual differences in the human EEG: an analysis of variance of spectral values. IEEE Trans. Biomed. Eng. BME **15**(3), 165–168 (1968)
4. Van Dis, H., Corner, M., Dapper, R., Hanewald, G., Kok, H.: Individual differences in the human electroencephalogram during quiet wakefulness. Electroencephalogr. Clin. Neurophysiol. **47**(1), 87–94 (1979)
5. Chan, H.L., Kuo, P.C., Cheng, C.Y., Chen, Y.S.: Challenges and future perspectives on electroencephalogram-based biometrics in person recognition. Front. Neuroinf. **12**, 66 (2018)
6. Fraschini, M., Meli, M., Demuru, M., Didaci, L., Barberini, L.: EEG fingerprints under naturalistic viewing using a portable device. Sensors **20**(22), 6565 (2020)
7. Campisi, P., La Rocca, D., Scarano, G.: EEG for automatic person recognition. Computer **45**(7), 87–89 (2012)
8. Marcel, S., Millan, J.R.: Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. IEEE Trans. Pattern Anal. Mach. Intell. **29**(4), 743–752 (2007)

9. La Rocca, D., et al.: Human brain distinctiveness based on EEG spectral coherence connectivity. IEEE Trans. Biomed. Eng. **61**(9), 2406–2412 (2014)

10. Mantini, D., Perrucci, M.G., Del Gratta, C., Romani, G.L., Corbetta, M.: Electrophysiological signatures of resting state networks in the human brain. Proc. Natl. Acad. Sci. **104**(32), 13170–13175 (2007)

11. Campisi, P., La Rocca, D.: Brain waves for automatic biometric-based user recognition. IEEE Trans. Inf. Forensics Secur. **9**(5), 782–800 (2014)

12. Moctezuma, L.A., Molinas, M.: Towards a minimal EEG channel array for a biometric system using resting-state and a genetic algorithm for channel selection. Sci. Rep. **10**(1), 1–14 (2020)

13. Pessoa, L.: Understanding brain networks and brain organization. Phys. Life Rev. **11**(3), 400–435 (2014)

14. de Vico Fallani, F., Richiardi, J., Chavez, M., Achard, S.: Graph analysis of functional brain networks: practical issues in translational neuroscience. Phil. Trans. Royal Soc. B: Biol. Sci. **369**(1653), 20130521 (2014)

15. Garau, M., Fraschini, M., Didaci, L., Marcialis, G.L.: Experimental results on multi-modal fusion of EEG-based personal verification algorithms. In: 2016 International Conference on Biometrics (ICB), pp. 1–6 (2016)

16. Boutorabi, S., Sheikhani, A.: Evaluation of electroencephalogram signals of the professional pianists during iconic memory and working memory tests using spectral coherence. J. Med. Signals Sensors **8**(2), 87 (2018)

17. Cox, R., Schapiro, A.C., Stickgold, R.: Variability and stability of large-scale cortical oscillation patterns. Netw. Neurosci. **2**(4), 481–512 (2018)

18. Pereda, E., García-Torres, M., Melián-Batista, B., Mañas, S., Méndez, L., González, J.J.: The blessing of dimensionality: feature selection outperforms functional connectivity-based feature transformation to classify ADHD subjects from EEG patterns of phase synchronisation. PLoS ONE **13**(8), e0201660 (2018)

19. Dimitriadis, S.I., Salis, C., Tarnanas, I., Linden, D.E.: Topological filtering of dynamic functional brain networks unfolds informative chronnectomics: a novel data-driven thresholding scheme based on orthogonal minimal spanning trees (OMSTs). Front. Neuroinf. **11**, 28 (2017)

20. Kottlarz, I., et al.: Extracting robust biomarkers from multichannel EEG time series using nonlinear dimensionality reduction applied to ordinal pattern statistics and spectral quantities. Front. Physiol. **11**, 614565 (2021)

21. Rodrigues, P.G., Filho, C.A.S., Attux, R., Castellano, G., Soriano, D.C.: Space-time recurrences for functional connectivity evaluation and feature extraction in motor imagery brain-computer interfaces. Med. Biol. Eng. Comput. **57**(8), 1709–1725 (2019)

22. Rosário, R.S., Cardoso, P.T., Muñoz, M.A., Montoya, P., Miranda, J.G.V.: Motif-synchronization: a new method for analysis of dynamic brain networks with EEG. Phys. A Stat. Mech. its Appl. **439**, 7–19 (2015)

23. Olofsen, E., Sleigh, J.W., Dahan, A.: Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. Br. J. Anaesth. **101**(6), 810–821 (2008)

24. Quintero-Quiroz, C., Montesano, L., Pons, A.J., Torrent, M.C., García-Ojalvo, J., Masoller, C.: Differentiating resting brain states using ordinal symbolic analysis. Chaos Interdiscip. J. Nonlinear Sci. **28**(10), 106307 (2018)

25. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R.: BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans. Biomed. Eng. **51**(6), 1034–1043 (2004)

26. Goldberger, A.L., et al.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000)

27. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods **134**(1), 9–21 (2004)

28. Ludwig, K.A., Miriani, R.M., Langhals, N.B., Joseph, M.D., Anderson, D.J., Kipke, D.R.: Using a common average reference to improve cortical neuron recordings from microelectrode arrays. J. Neurophysiol. **101**(3), 1679–1689 (2009)
29. Eckmann, J.-P., Kamphorst, S.O., Ruelle, D.: Recurrence plots of dynamical systems. Europhys. Lett. **4**(9), 973–977 (1987)
30. Costa, L.R.D., Campos, B.M.D., Alvim, M.K., Castellano, G.: EEG signal connectivity for characterizing interictal activity in patients with mesial temporal lobe epilepsy. Front. Neurol. **12**, 673559 (2021)
31. Mutanen, T.P., Metsomaa, J., Liljander, S., Ilmoniemi, R.J.: Automatic and robust noise suppression in EEG and MEG: the SOUND algorithm. Neuroimage **166**, 135–151 (2018)
32. Dong, L., et al.: MATLAB toolboxes for reference electrode standardization technique (REST) of scalp EEG. Front. Neurosci. **11**, 601 (2017)

# Forty-Class SSVEP-Based Brain-Computer Interface to Inter-subject Using Complex Spectrum Features

Christian Flores[1,2(✉)] , Romis Attux[3] , and Sarah N. Carvalho[1,2,3]

[1] School of Electrical and Computer Engineering, University of Campinas, Campinas, SP 13083-852, Brazil
[2] Department of Electrical Engineering, Universidad de Ingenieria y Tecnologia - UTEC, Lima, Peru
[3] Department of Electrical Engineering, Federal University of Ouro Preto, João Monlevade, MG 35931-008, Brazil
c265561@dac.unicamp.br

**Abstract.** The Steady-State Visually Evoked Potential (SSVEP) is one of the most popular paradigms for Brain-Computer Interface (BCI) applications. In this study, we address two challenges in designing SSVEP-based BCI. Firstly, our BCI system must be able to discriminate among the 40 available visual stimuli. In addition to the complexity brought by the high number of classes, visual stimuli flicker at close frequencies, only 0.2 Hz apart in the range of 8 to 15.8 Hz. The second challenge we addressed was the attempt to eliminate individualized system tuning. Our SSVEP-based BCI was designed using only data from subjects other than the user, that is, with cross-subject training. In the treatment of these two challenges, we extracted features with frequency and phase information for each of the 40 visual stimuli and applied them to a Linear Discriminant Analysis. The database has data from 35 subjects, so we trained with 34 subjects and tested with the remaining ones. We applied three different time windows of 1, 2 and 3 s to segment brain data and analyze the effect on classification accuracy. Our results reached an average classification, considering 40 classes, of 28.14%, 56.85% and 71.45% for a time window of 1, 2 and 3 s, respectively.

**Keywords:** Brain-Computer Interfaces · Steady-State Visually Evoked Potentials · spectrum features · Linear Discriminant Analysis

## 1 Introduction

A Brain-Computer Interface (BCI) is a technology that translates human brain activity into commands to control machines [1]. The brain signals can be recorded using different brain scan techniques, such as electroencephalography (EEG) and magnetic resonance imaging (MRI). Then, these brain signals are the input to a

BCI, which translates them using a classifier developed for a specific application. The EEG signals are usually employed in many applications due to their non-invasive character and low cost compared to other techniques [2].

People with motor or neurodegenerative disabilities, such as paraplegia, stroke, or amyotrophic lateral sclerosis, are potential users of BCI technology since it can provide alternative means to interact with their surroundings, improving their lifestyle [3].

EEG signals can capture information about different mental activities performed by a subject, such as Motor Imagery (MI), Steady-State Evoked Potential (SSVEP) and Event-Related Potential (ERP) [4,5]. The SSVEP is considered the most suitable paradigm of BCI, since it manages to guarantee a high Information Transfer Rate (ITR), without needing conditioning and preparation of the subjects [6]. The SSVEP is a brain potential that occurs in response to the observation of a constant oscillating visual stimulus, even if the subject is not paying full conscious attention to it [7]. These brain signals can be used to build a BCI to command different devices by showing patterns flickering in specific frequencies on a computer monitor, for example. The EEG signal can capture brain activity oscillations of SSVEP, which appear in the same frequencies and their harmonics engendered by visual stimulation [8].

The authors in [9] developed a framework for detecting SSVEP signals in an inter-subject classification experiment of 35 subjects using a Convolutional Neural Network (CNN) and two visual stimuli. Their results reached 82.2% mean test classification accuracy.

Furthermore, a Convolutional Neural Network on two types of features: Magnitude spectrum features (M-CNN) and Complex spectrum features (C-CNN) approaches were proposed in [10] to classify SSVEP. Two training scenarios were also compared: User-Independent (UI) training and User-Dependent (UD) training. The proposed C-CNN method is suitable for SSVEP and improves the performance of both UD and UI training scenarios. The classification results reported a maximum accuracy of 98.16% ± 12.3%, and 99.25% ± 5% for UI-C-CNN and UD-C-CNN, respectively.

On the other hand, Canonical Correlation Analysis (CCA) and its different variations, such as Filter Bank Canonical Correlation Analysis (FBCCA), were used to analyze SSVEP to discover the input target. In [11], the authors proposed the FBCCA to enhance the CCA performance, incorporating fundamental and harmonics components in the target identification. Their results stated that FBCCA outperformed the CCA method. In [12], the authors consider a cross-subject classification of 35 subjects, selecting two and five subjects with the highest recognition accuracy as a template reference for a transfer learning process. Their classification accuracy reported roughly 51% for a window length of 1 s.

Regarding robotics applied to SSVEP-based BCI, the authors [13] proposed a robotic arm for performing pick and place tasks, in combination with computer vision based on object recognition. The system was tested with ten healthy subjects obtaining a maximum classification accuracy of 97.75%.

In this paper, we designed an SSVEP-based BCI and addressed two major challenges still, open in the literature:

– A machine learning algorithm for BCI is frequently designed for one subject, in the context of what is called intra-subject classification. One disadvantage of this design is that we need to recalibrate the machine learning algorithm for a new subject. One solution is to develop BCI systems capable of suiting any subject, in what is termed inter-subject classification. Our SSVEP-based BCI was trained with data from 34 subjects and evaluated with data from the remaining subject.

– The BCIs are frequently designed to command electronic devices with few options. However, building multiclass BCI is more challenging, because the machine learning algorithm needs to operate with more choices and error possibilities. In this study, we designed a multiclass SSVEP-based BCI able to distinguish the 40 visual stimuli. This is a challenging problem since the frequencies of visual stimuli are spaced from only 0.2 Hz in the range of 8 to 15.8 Hz. In our solution, we used as features the spectral magnitude of the 40 evoked frequencies and phase data of visual stimulation, this information served as input for a Linear Discriminant Analysis (LDA).

The remainder of this paper is organized as follows: Sect. 2 describes the dataset and the techniques employed during pre-processing, feature extraction and classification. In Sect. 3, the results are analyzed and compared. Conclusions and future works are presented in Sect. 4

## 2 Materials and Methods

In this section, we present the database and the techniques employed in pre-processing, feature extraction and classification to build an SSVEP-based BCI.

### 2.1 Dataset

The database used in this work contains EEG signals acquired under the SSVEP paradigm. A virtual keyboard displayed 40 letters on a monitor, each of which flickered with a specific frequency and phase. The stimulation frequencies were distant from each other at 0.2 Hz, and they were in the range of 8.0 to 15.8 Hz. Adjacent frequencies had a phase difference of $0.5\pi$ radians. The brain data were acquired from 35 healthy subjects, of which 8 are experienced in using BCI, and 27 subjects did not have any prior experience in using BCI. Each subject observed each stimulus in six trials. The data matrix consists of 240 samples (40 targets x 6 trials) per subject. In each trial, signals were collected using 64 bipolar dry-electrodes for 6 s, with a sample rate of 250 Hz, resulting in 1500 points (6 s x 250 Hz). Therefore, the data length of six seconds includes 0.5 s before stimulus onset, 5 s for visual stimulation and 0.5 s after stimulus offset. More details about this dataset can be found in [14].
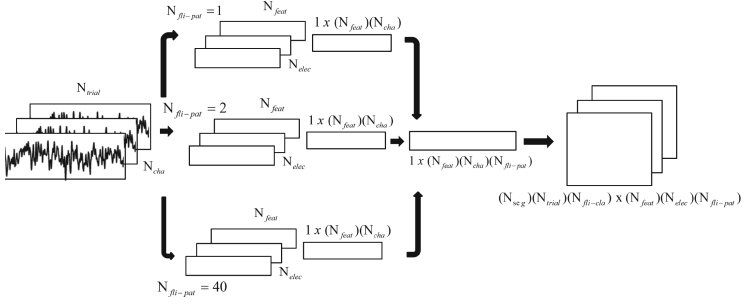
**Fig. 1.** Framework to depict the construction matrix features $X$ for each subject. (i) Columns construction: calculated over each segment $N_{seg}$, so features of magnitude and phase $N_{feat}$ are computed. This process is repeated for each electrode $N_{elec}$ and for all flicker patterns $N_{fli-pat}$. (ii) Rows construction: columns construction is repeated for all segments $N_{seg}$ calculates over all trials $N_{trial}$ and for each flickers classes $N_{fli-cla}$.

## 2.2 Signal Pre-processing

Although the database provides EEG data from 64 electrodes, this study used only the signals from the electrodes placed at POz, PO7, PO8, PO3, PO4, CB1, O1 and Oz. These signals were chosen because they are in the occipital lobe or close to the visual cortex region. These data were already filtered by a notch filter at 50 Hz in the data recording, to remove common powerline noise. Thus, the preprocessing of the data consisted of:

1. The first 0.5 s and the last 0.5 s of each sample were eliminated, to analyze only the 5 s interval where visual stimulation actually occurred.
2. The spatial filter common average reference (CAR) was applied according to Eq. 5.

$$x_i^{CAR}(n) = x_i(n) - \frac{1}{L} \sum_{i=1}^{L} x_j(n) \qquad (1)$$

3. The EEG signals were filtered by a seven-order elliptic filter passband, with cutoff frequencies at 2 and 30 Hz.
4. Each 5 s data sample was windowed at 1 s (250 points), 2 s (500 points) and 3 s (750 points), to allow the analysis in the three temporal scenarios.

## 2.3 Feature Extraction

The features were extracted from pre-processed EEG data. Firstly, the segments of 250, 500, or 750 points were filled with zeros up to 4098 points. We get a frequency resolution of 0.0610, so this procedure aims to guarantee precision in the calculation of the frequencies of the stimuli since the evoked frequencies have a spacing of only 0.2 Hz. Then, the Fast Fourier Transform (FFT) algorithm was applied and three magnitude and three angle points were extracted, the point

corresponding to the stimulus frequency or flicker pattern, the point before and after it.

All samples from the database were submitted to this feature extraction procedure, keeping correspondence with the evoked stimulus label. In this way, a feature matrix $X$ was generated for each subject, containing the multiplication of segments (i.e., 1, 2, or 3 s), trials (6), flicker classes (40) which set up the rows and the multiplication of complex features (6), electrodes channels (8) and flicker patterns (40) which set up the 1920 columns i.e., $X \in \mathbb{C}^{(N_{seg}.N_{trial}.N_{fli-cla}) \times (N_{feat}.N_{elec}.N_{fli-pat})}$, where:

$N_{seg}$: number of segments generated with windows of 1, 2 or 3 s per trial yielding 5, 3 and 1 segments, respectively.

$N_{trial}$: number of trials.

$N_{fli-cla}$: number of flickers classes.

$N_{feat}$: number of extracted features of magnitude and phase.

$N_{elec}$: number of electrodes.

$N_{fli-pat}$: number of flickers patterns.

Figure 1 describes a framework to build a matrix $X$.

## 2.4 Classification

Multiclass LDA was designed using data from 34 subjects and validated with data from the remaining subject, in a scheme of leave-one-subject-out-cross-validation. The classifier should be able to identify which of the 40 stimuli (classes) the subject was focusing his/her gaze on each instant.

A multiclass LDA classifier is based on the analysis of scattering matrices within-class scattering matrix and between-class scatter matrix. Given a set of samples $x_1,x_2,..,x_n$ and their classes targets $y_1,y_2,..,y_n$. The definition of a within-class scatter matrix is:

$$S_w = \sum_{i=1}^{n} (x_j - \mu_{y_j})(x_i - \mu_{y_j})^T \tag{2}$$

And, the definition of the between-class scatter matrix is:

$$S_b = \sum_{i=1}^{m} \eta_k (\mu_k - \mu)(\mu_k - \mu)^T \tag{3}$$

where, $\mu_k$ is the sample mean of the $k$-th class. Therefore, $m$ is the number of classes, $\mu$ is the overall sample mean, and $\eta_k$ is the number of samples in the $k$-th class.

Moreover, multiclass LDA is formulated as an optimization problem to find a set of linear combinations with coefficients $\mathbf{w}$ maximizing the ratio of the between-class scattering to the within-class scattering, as

$$\hat{\mathbf{w}} = \arg\max_{\theta} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}} \tag{4}$$

The generalized eigenvalue problem computed the solution:

$$S_b\mathbf{w} = \lambda S_w\mathbf{w} \tag{5}$$
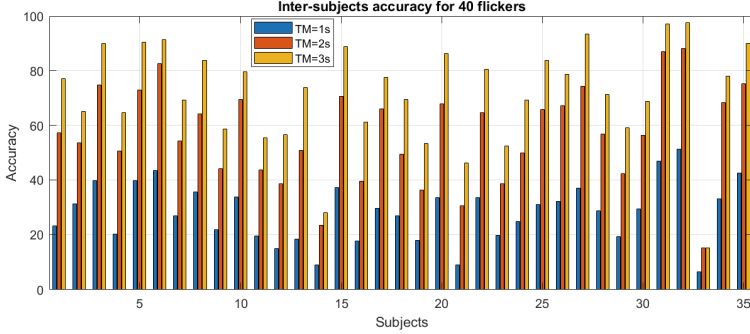


**Fig. 2.** Classification accuracy for each subject using leave-one-subject-out-cross-validation and for each time window of 1, 2 and 3 s.

## 3    Results and Discussion

We analyzed the classification accuracy using Linear Discriminant Analysis and the leave-one-subject-out cross-validation was performed for inter-subject classification. Hence, the samples of one subject were used for the testing stage and the samples of the rest of the subjects were used for the training stage.

Figure 2 shows the classification accuracy for each subject. The great variability of accuracy between subjects is noticed, and it is maintained for the three-time windows adopted. Subject 32 achieved the best accuracy with 51.33%, 88.13% and 97.50% for time windows of 1, 2 and 3 s, respectively. On the other hand, Subject 33 reported the worst performance, with a poor accuracy of only 6.58%, 15.00% and 15.00% for time windows of 1, 2 and 3 s, respectively. This variability is associated with the intrinsic characteristics of each subject [15]. There is a lot of variability around the mean value, for the three windows considered.

Figure 3 allows analyzing this widespread classification accuracy for each time window through a boxplot. LDA classifier reported mean classification accuracies of 28.14%, 56.85%, and 71.45% for time windows of 1, 2 and 3 s, respectively.

All subjects for Inter-Subject classification reported better classification accuracy for time windows of 3 s. Therefore, classification accuracy performance is enhanced by increasing the time window length. This is expected since larger windows have better spectral resolution and a higher signal-to-noise ratio. However, given the challenge of differentiating 40 visual stimuli with frequencies so
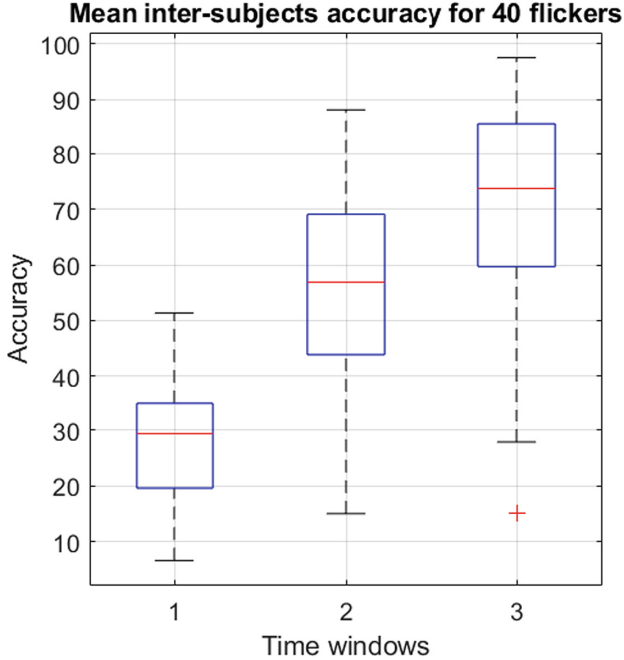
**Fig. 3.** Box plot for all subjects and for each time window of 1, 2 and 3 s.

close (0.2 Hz), the results of the SSVEP-based BCI system, considering a cross-subject model training, is interesting, being much superior to randomness (1/40), even in the most challenging scenario considered (with windowing of 1 s).

In the literature, we found some related studies that address the multiclass issue in SSVEP-based BCI, such as [11,13,15,16]. However, these papers consider an intra-subject classification. The study [10] deals with the cross-subject approach using a CNN, but they employ a maximum of 12 visual stimuli. While in [12], the authors reported accuracy of roughly 51% for a window length of 1 s using some subjects as a template reference. The selection of subjects a priori to train the system must be taken with caution concerning the representativeness of potential users of the conceived SSVEP-based BCI.

Thus, it is interesting to consider that our proposed challenges can be considered satisfied. In future works, we intend to address the transfer learning approach. Also, we will focus on more robust signal filtering techniques, feature extraction, and classifiers with deep learning, in order to improve the performance of the BCI system and mitigate the wide variability.

## 4    Conclusions

In this study, we employed the magnitude and phase of spectrum frequency of 40 flickers to classify SSVEP corresponding to a range from 8 Hz to 15.8 Hz, with an

interval of 0.2 Hz. Moreover, we design an inter-subject classification experiment for 35 subjects to address the BCI challenge of re-calibrating the system for any new subjects. Our SSVEP-based BCI presented an average performance of 28.14%, 56.85%, and 71.45%, for windows of 1, 2 and 3 s, respectively. This result is very promising, considering the two difficulties considered, 40 nearby visual stimuli and cross-subject training. It is interesting to observe that 8 subjects achieved a hit rate above 85% in the differentiation of the 40 stimuli, with a time window of 3 s. The high variability of classification accuracy for each subject, in part, due to the neurophysiological particularities of the subject himself, and in part, due to the developed BCI system. This last installment can be enhanced in future works using other machine learning techniques.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. Tiwari, N., Edla, D.R., Dodia, S., Bablani, A.: Brain computer interface: a comprehensive survey. Biologically Inspired Cogn. Architectures **26**, 118–129 (2018). https://doi.org/10.1016/j.bica.2018.10.005, https://www.sciencedirect.com/science/article/pii/S2212683X18301142
2. Sorger, B., Goebel, R.: Chapter 21 - real-time fMRI for brain-computer interfacing. In: Ramsey, N.F., del R Millán, J. (eds.) Brain-Computer Interfaces, Handbook of Clinical Neurology, vol. 168, Elsevier, pp. 289–302 (2020). https://doi.org/10.1016/B978-0-444-63934-9.00021-4, https://www.sciencedirect.com/science/article/pii/B9780444639349000214
3. Zhuang, M., Wu, Q., Wan, F., Hu, Y.: State-of-the-art non-invasive brain-computer interface for neural rehabilitation: a review. J. Neurorestoratology **8**(1), 12–25 (2020). https://doi.org/10.26599/JNR.2020.9040001, https://www.sciencedirect.com/science/article/pii/S2324242622000080
4. Nam, C., Nijholt, A., Lotte, F.: Brain-Computer Interfaces Handbook: Technological and Theoretical Advances (2018)
5. de Carvalho, S.N.: Space-time filter for SSVEP brain-computer interface based on the minimum variance distortionless response. Med. Biol. Eng. Comput. **59**(450), 1133–1150 (2021)
6. Kwak, N.S., Müller, K.R., Lee, S.W.: A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. PLoS ONE **12**(2), 1–20 (2017). https://doi.org/10.1371/journal.pone.0172578
7. Nik Aznan, N.K., Bonner, S., Connolly, J., Al Moubayed, N., Breckon, T.: On the classification of SSVEP-based dry-EEG signals via convolutional neural networks. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp 3726–3731 (2018). https://doi.org/10.1109/SMC.2018.00631
8. Norizadeh Cherloo, M., Kashefi Amiri, H., Daliri, M.R.: Spatio-spectral CCA (SS-CCA): a novel approach for frequency recognition in SSVEP-based BCI. J.

Neurosci. Methods **371**, 109499 (2022). https://doi.org/10.1016/j.jneumeth.2022.109499, https://www.sciencedirect.com/science/article/pii/S0165027022000267

9. Bassi, P.R., Rampazzo, W., Attux, R.: Transfer learning and specaugment applied to SSVEP based BCI classification. Biomed. Signal Process. Control **67**, 102542 (2021). https://doi.org/10.1016/j.bspc.2021.102542, https://www.sciencedirect.com/science/article/pii/S1746809421001397

10. Ravi, A., Beni, N.H., Manuel, J., Jiang, N.: Comparing user-dependent and user-independent training of CNN for SSVEP BCI. J. Neural Eng. **17**(2), 026028 (2020). https://doi.org/10.1088/1741-2552/ab6a67

11. Chen, X., Wang, Y., Gao, S., Jung, T.P., Gao, X.: Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. J. Neural Eng. **12**(4), 046008 (2015)

12. Yan, W., Wu, Y., Du, C., Xu, G.: Cross-subject spatial filter transfer method for SSVEP-EEG feature recognition. J. Neural Eng. **19**(3), 036008 (2022)

13. Chen, X., Zhao, B., Wang, Y., Gao, X.: Combination of high-frequency SSVEP-based BCI and computer vision for controlling a robotic arm. J. Neural Eng. **16**(2), 026012 (2019)

14. Wang, Y., Chen, X., Gao, X., Gao, S.: A benchmark dataset for SSVEP-based brain-computer interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. **25**(10), 1746–1752 (2017). https://doi.org/10.1109/TNSRE.2016.2627556

15. Oikonomou, V.P., Nikolopoulos, S., Kompatsiaris, I.: A Bayesian multiple kernel learning algorithm for SSVEP BCI detection. IEEE J. Biomed. Health Inform. **23**(5), 1990–2001 (2019). https://doi.org/10.1109/JBHI.2018.2878048

16. Li, Y., Xiang, J., Kesavadas, T.: Convolutional correlation analysis for enhancing the performance of SSVEP-based brain-computer interface. IEEE Trans. Neural Syst. Rehabil. Eng. **28**(12), 2681–2690 (2020). https://doi.org/10.1109/TNSRE.2020.3038718

# Effect of Acoustic Intensity on Multichannel Cortical Auditory Evoked Potential Elicited by Spanish Words in Adults with Normal Hearing

Norma Castañeda-Villa$^{(\boxtimes)}$ (iD), Pilar Granados-Trejo (iD), and Juan M. Cornejo-Cruz (iD)

Universidad Autónoma Metropolitana, Iztapalapa, 09310 CDMX, México
`ncv@xanum.uam.mx`

**Abstract.** This research studied the effect of the acoustic intensity level of four Spanish words on the multichannel Auditory Cortical Evoked Potential (CAEP); the amplitude of the characteristic peaks of this potential (P1, N1, and P2) will depend on the intensity of the stimulus used to elicit it. Eight young adults (four females and four males), within an age range of 18 and 29, with normal hearing, participated in this study. Four disyllabic words (/bota/, /papa/, /lobo/, and /sabor/) and three acoustic intensity levels (20, 40, and 60 dBHL) were used to evoke the CAEP. The amplitudes of the N1 and P2 peaks and N1-P2 inter-peak were analyzed as a function of the acoustic intensity of the stimuli for each word. The significance differences between the mentioned amplitudes and the different intensities were determined using an ANOVA analysis. Linear Regression was used to determine which recording electrodes better reflected the relationship between the amplitudes of the peaks and the stimuli' acoustic intensity.

**Keywords:** Cortical Auditory Evoked Potential · Spanish · disyllabic words · Acoustic intensity

## 1 Introduction

Different studies have used Cortical Auditory Evoked Potentials (CAEP) to evaluate the changes in stimuli perception due to synchronized neural activity underlying auditory processing. If the CAEP can be recorded for determinate acoustics stimuli characteristics, it is assumed the stimuli are audible for the subject [1].

Garinis and Cone (2007) were pioneers in assessing the effect of stimulus levels on CAEP. The authors used the syllables /ta/, /da/, and /sa/ to elicit the response in 15 normal-hearing adults; the acoustic intensity levels used were 0, 20, and 40 dBSL over the lowest level at which syllables were discriminated with more than 95% accuracy. Finally, the authors determined that the amplitude of the characteristics P1, N1, and P2 response peaks increased as the level increased, while the latency of those peaks decreased as the stimulus increased [2].

Another important group studying the effects of acoustic intensity levels on the speech-evoked CAEP is the one led by Purdy et al. (2013). This group selected low and high frequency speech sounds (/m/ and /t/ respectively) and three acoustic intensity levels (30, 50, and 70 dBSPL or 40, 60, and 80 dBSPL) to evaluate a group of 16 infants. They found different behaviors in the amplitude and latency of the P1 peak using the two speech sounds at different acoustic intensity levels. The authors mention that their results in infants are different from those reported in adults, which could be attributed to developmental effects on acoustic intensity coding by the auditory cortex. Finally, they recommend that the effects of stimulus characteristics (temporal, spectral, and level) on CAEPs should be considered when using this response for hearing aid or cochlear implant evaluation in infants [3].

Other authors have proposed to study the Latency-Acoustic intensity function of the Auditory Late Latency Response in normal-hearing subjects using speech stimuli; the syllable /ta/ with an acoustic intensity of 30, 50, 70, and 90 dBSPL. They found that the acoustic intensity influences P1 and N1 latencies nonlinearly while their amplitude decreases with a reduction in acoustic intensity, although there is no significance [4].

More recently, Barry et al. (2022) examined the CAEP peaks and Principal Components as a function of tone acoustic intensity: 1000 Hz at 50 (soft) and 80 (loud) dB SPL. They report an increase in the amplitude of some 70% from soft to loud components. Additionally, the authors include a table with an extensive review of studies that reported relationships between CAEP peak amplitudes and stimulus acoustic intensity [5]. One year before, Soleimani et al. (2021) published a review article that includes different researches that have demonstrated the utility of CAEP as an objective method in the management of deaf children [6]. Additionally, CAEPs have been proposed to estimate hearing thresholds in children and adults with normal hearing or users of hearing aids or cochlear implants [7, 8].

A common element among those researchers is that they use tones or consonant-vowel contrast stimuli to elicit the auditory response. The novelty of our study was to investigate the effect of the acoustic intensity level in the CAEP but using disyllabic Spanish words.

## 2   Material and Methods

### 2.1   Subjects

Our study included eight normal-hearing voluntary young adults between 18 and 29 years old ($23.67 \pm 3.57$); four females and four males. All were neurologically normal, with lower than 25 dBHL hearing thresholds at 250 Hz to 8 kHz. The participants remained seated in a recliner in an anechoic chamber during a test session of approximately one hour. They were asked to stay on alert, trying to identify which words they were hearing.

### 2.2   Stimuli

We used four disyllabic Spanish words as stimuli (/bota/, /papa/, /lobo/, and /sabor/). Previous research determined the temporal characteristics and topographical distribution

of P1-N1-P2 peaks elicited by these words at 75 dBHL acoustic intensity and the difference between the words' amplitude [9]. We use three different intensities 20, 40, and 60 dBHL. The words were presented randomly via TDH-49 headphones and repeated each approximately 100 times.

### 2.3 EEG Recording

The EEG was recorded on 20 electrodes (FP1–2, F3–4, C3- 4, P3–4, O1–2, F7–8, T3–4, T5–6, Cz, Pz, Fz, and AFz) using NeuroScan™ equipment, the impedance of the electrodes was kept below 5 kHz. FPz was the ground electrode, and the linked mastoids were the reference. EEG acquisition sampling rate of 1000 Hz, with a gain of 12,500 for accuracy of 0.007 μV/LSB. Low pass filter of 200 Hz and High pass filter of 1.0 Hz, 12 dB/oct. Notch filter On. Artifact rejection of ±70 μV. Groups of four randomized words, probability of 0.25 each, were presented to the subject up to 150 times for each word and intensity test. The amplitude of every.SND file was equalized to not exceed, on average, the 0 dB mark on the VU meter of the audiometer at the time of the test. The study complied with all applicable research and ethical standards and laws followed by the Declaration of Helsinki principles. The recording consisted of 150 epochs in a time window of 100 ms before the stimulus and 900 ms after the stimulation. Offline processing of the recordings was carried out before their analysis, which consisted of 1) Digital filtering between 1 and 30 Hz, 2) original sampling frequency reduction to 500 Hz, 3) epochs extraction in an analysis window of 700 ms (100 to 600 ms), 4) elimination of epochs whose amplitude exceeded ±50 μV, 5) remotion of the baseline across epochs, and 6) elimination of EEG artifacts (blinks, eye movements, and line noise) by Independent Component Analysis. Finally, we averaged 90 epochs with 350 points for each word and intensity.

### 2.4 Statistical Analysis

After carrying out the shapiro-wilk test (P > 0.05) we determined the no normality of our data; an anova test was used to determine the significant differences between the amplitudes of the CAEP peaks and the acoustic intensity for each word. After determining whether there were significance differences between the amplitudes and acoustic intensities, a post hoc range test was used to determine which responses presented the differences. The peak and inter-peak amplitudes of the four words and the three stimuli acoustic intensity were compared using a t-test. Finally, a linear regression was used to determine which electrodes better fit the relationship between the CAEP peaks' amplitudes and the stimuli' acoustic intensity.

## 3 Results

Figure 1 shows the grand average for each word (columns) and each acoustic intensity level (rows) of the CAEP for all the subjects (electrode Cz). It was impossible to recognize the P1 peak in any of these recordings. P2 was the most positive peak within the latency range of 250 and 450 ms at an acoustic intensity level of 60 dBHL, and N1 was the most

negative before P2; with these marks, the peaks at 40 y 20 dBHL were located. The N1 and P2 amplitudes are sensitive to acoustic intensity changes, which we observe in all the words. On the other hand, a clear relationship between the latencies of this peak and the acoustic intensity level is not observed.
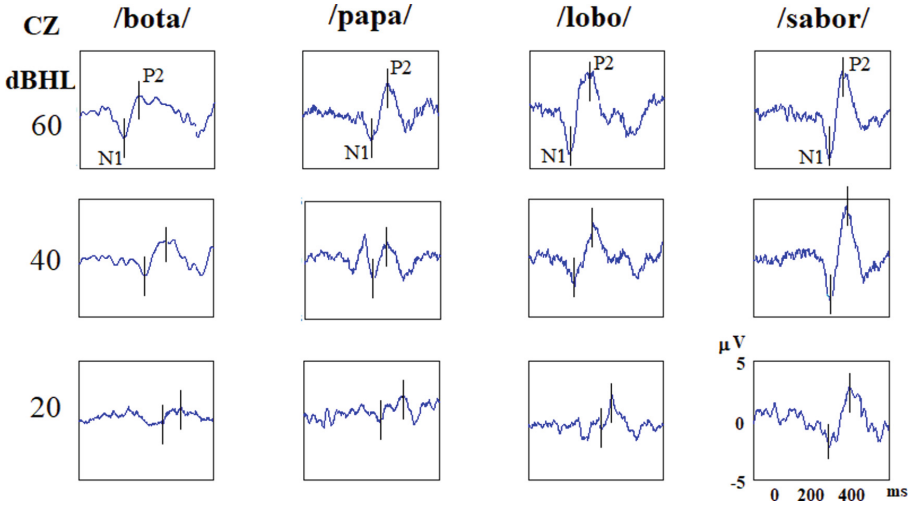


**Fig. 1.** The grand average (n = 8) for each word (column) and each acoustic intensity level (row) taken across all subjects. Marks indicate N1 and P2 peaks at Cz

### 3.1 Statistical Analysis

To determine which electrodes reflexed the acoustic intensity changes, we plot the amplitude of the mentioned peak to 20 and 60 dBHL. Figure 2 shows an example of the dispersion plot corresponding to P2 amplitudes (means across the selected electrodes). The regression line fits Cz, C3–4, AFz, Fz, FP1–2, and F3; better than F4, F7–8, Pz, and P4 with a greater deviant.

Average CAEP (n = 8) N1, P2, and N1-P2 amplitudes for 20, 40, and 60 dBHL intensities were analyzed for each one of the following electrodes: FP1–2, F3–4, F7–8, C3–4, AFz, Cz, Fz, Pz. And P4; we select these sites because corresponded to the topographical distribution of the peaks observed in our previous research [9]; see Fig. 3.

The ANOVA test showed that for the word /bota/, there are significance differences between the amplitudes of the peaks N1, P2, and the interpeak N1-P2 and the intensities (p = $5.78 \times 10^{-6}$, p = $2.44 \times 10^{-7}$ and p = $5.75 \times 10^{-7}$, respectively). Post hoc comparison indicated significance differences between N1 amplitude at 20 and 40 dBHL (p = $4.11 \times 10^{-4}$) and 20 and 60 dBHL (p = $7.76 \times 10^{-7}$), for P2 only between 20 and 60 dBHL (p = $2.97 \times 10^{-6}$) and for N1-P2 in 20 and 40 dBHL (p = $1.22 \times 10^{-4}$) and 20 and 60 dBHL (p = $9.87 \times 10^{-7}$).

In /papa/, ANOVA showed significance differences in N1 (p = $1.99 \times 10^{-8}$) and N1-P2 (p = $7.04 \times 10^{-5}$). There are no significance differences between the amplitude
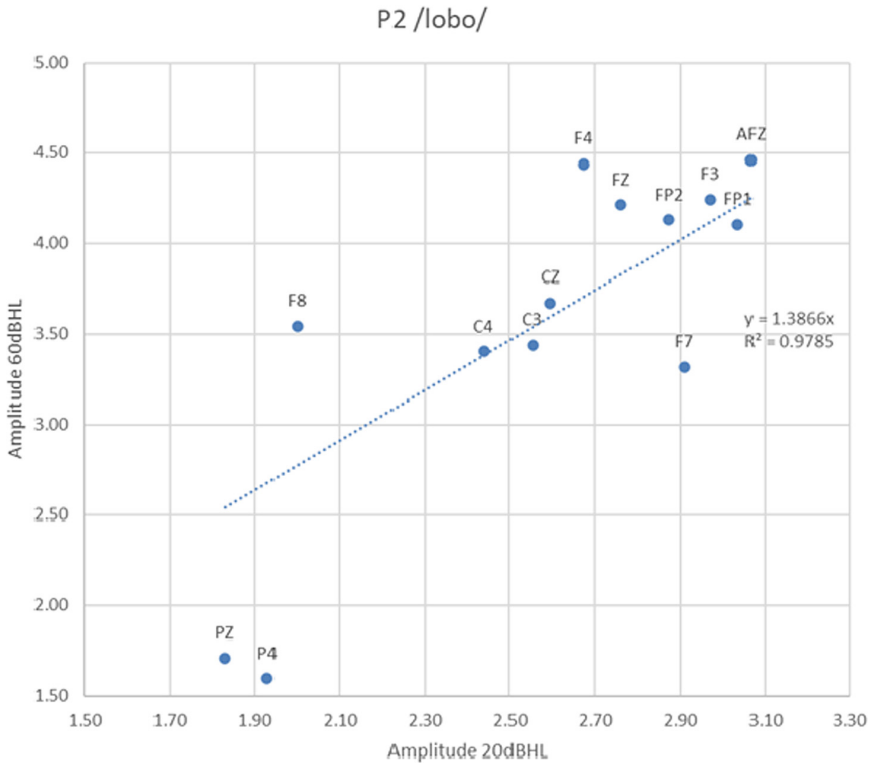
**Fig. 2.** A Scatter plot shows the relation between P2 amplitude at 20 and 60 dBHL; a linear regression line was fit to determine the electrodes closer to this behavior.

of P2 ($p > 0.005$) and intensities. The t-test indicated that in both N1 and N1-P2, there are significance differences between 20 and 40 dBHL ($p = 2.71 \times 10^{-6}$ and $p = 5.86 \times 10^{-5}$, respectively). For the word /lobo/ ANOVA determined significance differences between N1, P2 and N1-P2 and the intensities ($p = 5.73 \times 10^{-7}$, $p = 14.00 \times 10^{-4}$ and $p = 1.31 \times 10^{-5}$, respectively), the post hoc comparison indicated significance differences only between the acoustic intensity of 20 and 60 dBHL (N1 $p = 21 \times 10^{-4}$, P2 $p = 1.75 \times 10^{-5}$ and N1-P2 $p = 6.72 \times 10^{-6}$). Finally, for the word /sabor/ significance differences were found for the amplitudes N1 ($p = 3.47 \times 10^{-7}$), P2 ($p = 219 \times 10^{-4}$) and N1-P2 ($p = 4.87 \times 10^{-5}$). The subsequent comparison showed significance differences between the amplitudes of N1 at 20 and 40 dBHL ($p = 1.23 \times 10^{-5}$) and 20 and 60 dBHL ($p = 6.01 \text{x} 10^{-8}$), for P2 only between 20 and 60 dBHL ($p = 20.00 \times 10^{-4}$) and for N1-P2 between 20 and 40 dBHL ($p = 3.69 \times 10^{-4}$) and 20 and 60 dBHL ($p = 9.45 \times 10^{-7}$).
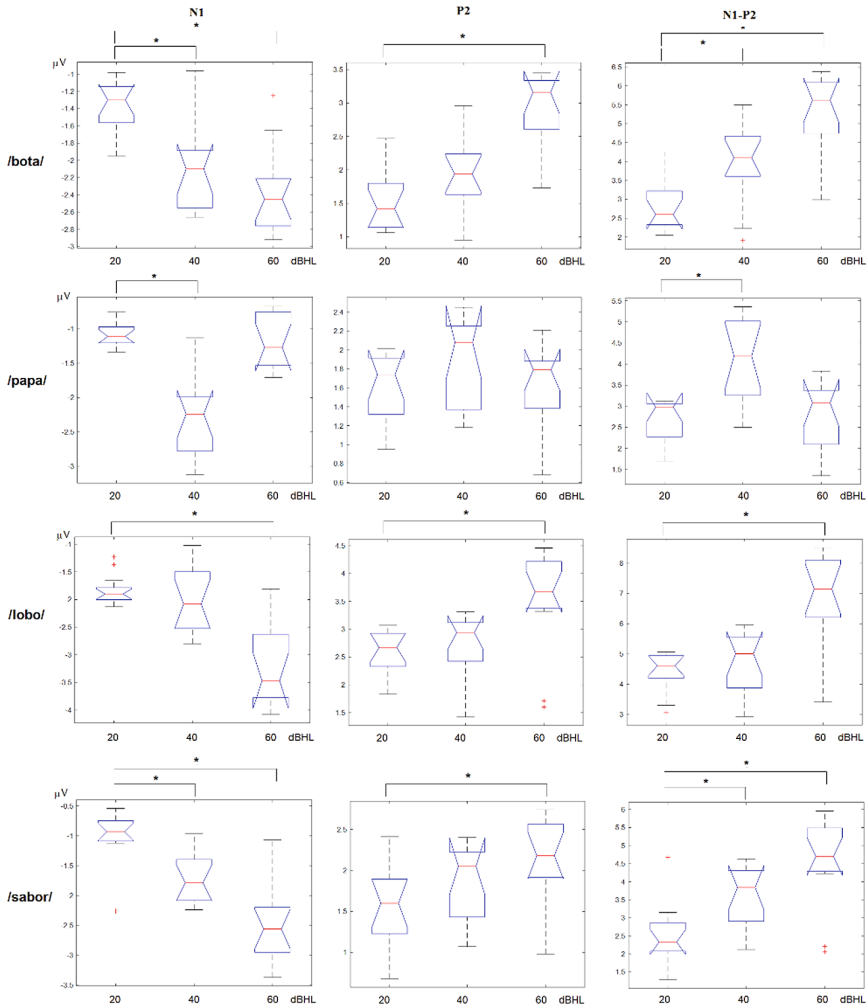
**Fig. 3.** Boxplots of the amplitudes for N1, P2, and N1-P2 for three intensities; values correspond to the average data for 13 electrodes position. The asterisk (*) represents the level of significance ($p < 0.05$).

## 4 Discussion

The amplitudes of N1. P2 and N1-P2 in words/bota/, /lobo/, and /sabor/ increased as the acoustic intensity of the stimulus increased (for the 13 electrodes analyzed), the expected behavior in this range of stimuli intensities. The above is under the findings of other authors who suggest that these are mandatory markers of sensory processing, leading to a stimulus-classification stage of processing [1]. The above did not happen in the case of the words /papa/. Since CAEP is a response to the characteristics of the stimulus, this could be explained by the fact that in this word, there are two equal syllables, less contrast than in the other three words [10].

N1 seems the most sensitive peak to acoustic intensity changes, with significance differences between 20 and 40 dBHL and between 20 and 60 dBHL. On the other hand, P2 only shows significance differences in amplitude when there is an acoustic intensity change from 20 to 60 dBHL. The above could be helpful if you want to use the procedure followed in this work to determine the detection threshold with N1 (liminal test) and the discrimination threshold with P2 (supraliminal test) [8].

Finally, the electrodes that showed the expected behavior of the increase in amplitude with increasing acoustic intensity were Cz, C3–4, AFz, Fz, FP1–2, and F3.

## 5   Conclusion

In this study, we analyze the effect of the stimuli acoustic intensity on the characteristic peaks of the CAEP. The objective was to determine the changes in amplitude of the P1, N1, P2, and N1-P2 peaks to three acoustic intensity levels (20. 40 and 60 dBHL). We use four disyllabic Spanish words (/bota/, /papa/, /lobo/ y /sabor/) as stimuli. P1 was not observed in the recordings using these acoustic intensity levels. Although we observed a relation between the N1 and P2 amplitudes with the acoustic intensity changes, there were no significance differences for all the words and intensities. Two words showed significance differences in amplitude peaks and acoustic intensity levels, /bota/ and /sabor/. Although it is necessary to repeat this protocol in other populations, for example, in normal-hearing children and hearing aid or cochlear implant users, our protocol could be used to determine the perception and discrimination word threshold in Spanish objectively. Finally, the electrodes that showed the expected behavior of the increase in amplitude with increasing acoustic intensity were Cz, C3–4, AFz, Fz, FP1–2, and F3. Knowing this could reduce the time of electrode placement and the test, which is very suitable in registries with children. This study provides the bases to develop procedures for the objective fit of a Hearing Aid or a Cochlear Implant using Spanish words.

## References

1. Mathew, A.K., Purdy, S.C., Welch, D., Pontoppidan, N.H., Rønne, F.M.: Electrophysiological and behavioural processing of complex acoustic cues. Clin. Neurophysiol. **127**(1), 779–789 (2016). https://doi.org/10.1016/j.clinph.2015.04.002
2. Garinis, A., Cone-Wesson, B.: Effects of stimulus level on cortical auditory event-related potentials evoked by speech. J. Am. Acad. Audiol. **18**(2), 107–116 (2007). Accessed 20 Nov 2014. http://www.ingentaconnect.com/content/aaa/jaaa/2007/00000018/00000002/art00003
3. Purdy, S.C., Sharma, M., Munro, K.J., Morgan, C.L.A.: Stimulus level effects on speech-evoked obligatory cortical auditory evoked potentials in infants with normal hearing. Clin. Neurophysiol. **124**(3), 474–480 (2013). https://doi.org/10.1016/j.clinph.2012.09.011
4. Prakash, H., Abraham, A., Rajashekar, B., Yerraguntla, K.: The effect of intensity on the speech evoked auditory late latency response in normal hearing individuals. J. Int. Adv. Otol. **12**(1), 67–71 (2016). https://doi.org/10.5152/iao.2016.1776
5. Barry, R.J., De Blasio, F.M., Rushby, J.A., MacDonald, B., Fogarty, J.S., Cave, A.E.: Stimulus intensity effects and sequential processing in the passive auditory ERP. Int. J. Psychophysiol. **176**, 149–163 (2022). https://doi.org/10.1016/j.ijpsycho.2022.03.005

6. Soleimani, M., Rouhbakhsh, N., Rahbar, N.: Towards early intervention of hearing instruments using cortical auditory evoked potentials (CAEPs): A systematic review. Int. J. Pediatr. Otorhinolaryngol. **144**, 110698 (2021). https://doi.org/10.1016/j.ijporl.2021.110698

7. Lightfoot, G.: Summary of the N1–P2 cortical auditory evoked potential to estimate the auditory threshold in adults. Semin. Hear. **37**(1), 1–8 (2016). https://doi.org/10.1055/s-0035-1570334

8. Cardon, G., Sharma, A.: Cortical neurophysiologic correlates of auditory threshold in adults and children with normal hearing and auditory neuropathy spectrum disorder. Am. J. Audiol. **30**(1), 28–42 (2021). https://doi.org/10.1044/2020_AJA-20-00062

9. Granados-Trejo, P., Castañeda-Villa, N., Cornejo-Cruz, J.M.: Características temporales y distribución topográfica de los potenciales evocados auditivos corticales evocados por palabras en español en una población adulta. Rev. Mex. Neurocienc. **17**(2), 42–49 (2015)

10. Mueller, J.L., ten Cate, C., Toro, J.M.: A Comparative perspective on the role of acoustic cues in detecting language structure. Top. Cogn. Sci. **12**(3), 859–874 (2020). https://doi.org/10.1111/tops.12373

# On Hyperparameters Tuning for Deep Learning Segmentation Models Applied to COVID-19 Computerized Tomography Images

Karina Sequia, Fernanda Nascimento, and Anderson Gabriel Santiago(✉)

Center for Engineering, Modelling and Applied Social Sciences,
Federal University of ABC, São Bernardo do Campo, Brazil
`gabriel.santiago@ufabc.edu.br`

**Abstract.** This paper presents a study of hyperparameters tuning for two different image deep learning semantic segmentation architectures applied to COVID-19 Computerized Tomography images, UNet and UNet++, using as encoder layer the feature extractors of five different Convolutional Neural Networks - ResNeXt, Xception, MobileNet, VGG13 with Batch Normalization and DenseNet121. The hyperparameters considered are the learning rate, number of epochs for training, batch size and the optimizer algorithm. In order to find the optimal set, instead of merely combining each possible set o hyperparameters which results in a high computational cost, it employs a Bayesian search algorithm with early stopping, leading to faster analysis and provides not only the hyperparameters optimal values but also the impact of each one of them over the resulting mean of Intersection over Union - metric chosen to evaluate the models performance - as well as the segmentation results. All images are available as a public anonymized datasets.

**Keywords:** COVID-19 · Hyperparameter Optimization · Image Segmentation · Deep Learning · Computerized Tomography

## 1 Introduction

Coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2 virus, has surpassed 200 million confirmed cases and caused more than 4 million deaths worldwide as of August 2021, according to the Organization World Health. Because this virus mainly causes respiratory problems, Computerized Tomography (CT) of the chest is one of the most applied imaging methods for the detection and follow-up of the treatment of patients with COVID-19. With a high sensitivity for diagnosing the disease, CT chest scans allows the verification of certain patterns that reveal lesions associated with SARS-CoV-2, such as the appearance of ground-glass opacity, reticulation and consolidation, according to variations in texture, size and position in the images.

The demand for applications of Artificial Intelligence (AI) in medical field is increasing due to the large number of imaging tests produced, as a result of the development of advanced techniques to obtain this type of analysis [18]. Thus, in order to have a better interpretation of these images in a shorter period of time, Deep Learning (DL), a subarea of AI, is the most commonly used technique for this application, capable of effectively performing data classification, detection and semantic segmentation of objects given an image dataset with large amounts of data [19].

Segmentation of medical images is already of vital importance in many biomedical applications, such as computer-assisted diagnosis, treatment and monitoring of many diseases. Typical processes of medical image segmentation include brain, tumor, cardiac, pulmonary segmentation, among others [11]. By identifying the pixels that define organs and lesions in exams such as CT and MRI, segmentation becomes a tool that allows obtaining fundamental information about the shapes and volumes of anatomical and pathological structures, allowing physicians to analyze only the significant parts of biomedical imaging [6].

UNet [13] is a DL architecture specialized in object segmentation and uses a two-step procedure: (i) image encoding, which extracts the image features and (ii) image decoder, which creates a mask for the segmented regions. The encoder itself can be any Convolutional Network architecture such as VGG or ResNet, and can be thought as one of the hyperparameters used in the model training. Figure 1 presents the basic structure of a UNet model.
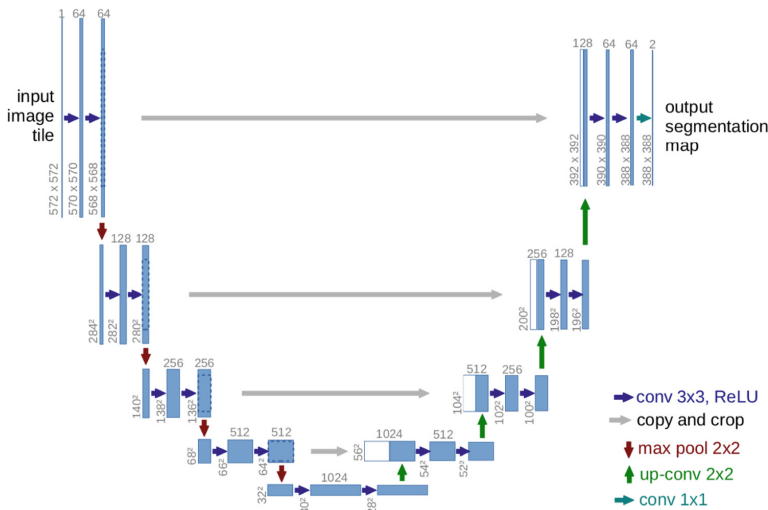


**Fig. 1.** UNet basic structure.

Aslan et al. [2] present a study using Chest X-Ray images images (CRX) using CNN's models for feature extraction and Machine Learning algorithms such as

Support Vector Machines (SVM) and Decision Trees (DT) in substitution of the conventional fully connected layers along Bayesian search for hyperparameters optimization pointing out that there are few studies that provide such information.

Since finding the optimal set of hyperparameters represents a crucial step in improving the performance of a DL model, this paper presents a study of hyperparameters tuning - learning rate, number of epochs, batch size and optimizer - considering UNet and its variation, UNet++ [20], and five different CNN's architectures as encoders: ResNeXt [17], Xception [4], MobileNet [7], VGG13 with Batch Normalization [16] and DenseNet121 [8] applied to COVID-19 image segmentation. A Bayesian search is used in order to determine the optimal hyperparameter sets that are presented for each one of the ten different combination of segmentation architecture and encoder. The results also presents the hyperparameters that posses the highest influence in the metric analysed (the mean Intersection over Union for the test dataset), meaning that their variation impacts over 60% of the metric analysed.

## 2   Methodology

### 2.1   COVID-19 Image Database and Data Augmentation

All images used are anonymized and available as public dataset[1] and consists of 349 Computerized Tomography (CT) of COVID-19 positive images splited in 70% - 30% training-test ratio, since the goal of this paper is not provide a full segmentation model, but a study on how the most common deep learning hyperparameters can affect the training and testing procedures of a segmentation model. The segmentation was performed by an specialist using ImageJ[2] and observing the ground glass opacity (GGO) as an indicative of COVID-19 lesions.

In order to increase the generality of the network and decrease overfitting during the training procedure, Data Augmentation was used via Albumentation [3] Python API and the image transformations considered are listed below:

- Image Rotation with maximum rotation angle of $\theta_{max} = 40^o$ and 50% probability of occurrence;
- Horizontal flipping with 50% probability of occurrence;
- Vertical Flipping with 50% probability of occurrence;
- Random size crop with maximum height and width of $100 \times 100$ pixels with 50% probability of occurrence;

It is important to mention that before any of those transformations are applied, the training dataset goes under a minimum-maximum normalization, resulting in pixel value range limited to (0, 1), followed by standard scaling[3]

---

[1] https://www.kaggle.com/datasets/luisblanche/covidct?select=CT_COVID.

[2] https://imagej.nih.gov/ij/.

[3] The mean and standard deviation values were evaluated according to the training dataset.

with mean and standard deviation of $\mu = 0.6255$ and $\sigma = 0.3183$ and image resizing to $128 \times 128$ pixels. For the testing procedure no Data Augmentation was applied, only pixel normalization, standard scaling and image resizing are applied.

## 2.2   Deep Learning Segmentation Models

The segmentation architectures chosen for the proposed work are UNet and UNet++, both consists of two-step model: an encoder, responsible for extracting features from the image dataset and a decoder which takes the image features from the first step and aims to reconstruct a segmentation map for the desired image classes. UNet++ posses a more complex decoder than UNet and both of them can use other CNN's such as VGG13 and Xception as encoders. The loss function used to update the model parameters is the "Binary Cross Entropy Loss" in Eq. 1, where $N_c$ is the number of classes, $y_i$ is the predicted value for the i-eth class and $\bar{y}_i$ its true label.

$$H(y, \bar{y}) = -\frac{1}{N_c} \sum_{i=1}^{N_c} [y_i \log(\bar{y}_i) + (1 - y_i) \log(1 - \bar{y}_i)] \qquad (1)$$

All architectures were implemented using Python 3.9 Anaconda distribution and PyTorch[4] Application Programming Interface (API). Concerning the hardware, an Intel i7 Core, 16 Gb of RAM and a Tesla K80 Nvidia Graphics Processing Unit (GPU) with 24 Gb of RAM running Linux Ubuntu 22.04 LTS was used.

The CNN architectures used as encoders for UNet and UNet++ were chosen based on algorithm structure and the number of trainable parameters are presented in Table 1. It is important to emphasize that only the feature extraction layer of each CNN presented were used, i.e., the classification layer was completely discarded and no transfer learning applied.

**Table 1.** CNN architectures used as encoders with corresponding number of trainable parameters.

| Encoder | Trainable parameters |
|---|---|
| DenseNet 121 | $6 \times 10^6$ |
| Xception | $22 \times 10^6$ |
| VGG13 BN | $9 \times 10^6$ |
| MobileNet V3 Small Minimal 100 | $0.43 \times 10^6$ |
| ResNeXt 50 | $22 \times 10^6$ |

The metric used to evaluate the performance of each model is the "mean Intersection over Union" ($IoU$) [12], given by Eq. 2, where $N$ is the number of

---
[4] https://pytorch.org/.

images being segmented, $M_{pred}$ the mask predicted by the segmentation model and $M_{ref}$ the mask of reference.

$$IoU = \frac{1}{N} \sum_{i=1}^{N} \frac{M_{pred}^i \cap M_{ref}^i}{M_{pred}^i \cup M_{ref}^i} \tag{2}$$

### 2.3 Hyperparameters Optmization

In order to perform the hyperparameter optmization, several algorithms are available such as grid search from Scikit-Learn [15] library, which performs a time expensive brute force search over the entire hyperparameter space, or random grid search, also from Scikit-Learn, which may never lead to an optimal solution. Akiba et al. [1] presents a model based on Baeysian search in order to provide the optimal hyperparameters values and this strategy is employed in this work and the set of hyperparameters considered along their respective ranges are presented in Table 2:

**Table 2.** Hyperparameter range.

| Hyperparameter | Range |
|---|---|
| Epochs | [100, 250] |
| Batch Size | [10, 50] |
| Learning Rate | [$10^{-6}$, $10^{-1}$] |
| Optimizer | [Adam [9], SGD [14]] |

The API devoloped by Akiba et. al. also allows that unpromising tries be suspended (pruned) before their conclusion by monitoring the evaluation of a chosen metric over a pre-determined number of epochs. If the metric does not improve, the execution is halted and the computational resources are freed to another set of hyperparameters.

For the purpose of model analysis, Eq. 2 was evaluated over the test dataset and all combinations of segmentation architectures and encoders were evaluated over 200 tries.

## 3 Results

### 3.1 Optimal Hyperparameters

After evaluating each combination of segmentation architecture (SA) and encoders, Table 3 presents the highest $IoU$ for each architecture.

The optimal hyperparameters are presented in Tables 4 and 5 for UNet and UNet++ respectively.

Tables 6 and 7 presents the two most relevant hyperparameters, i.e., changing their values impacts over 60% of the $IoU$ for each pair architecture/encoder.

**Table 3.** Segmentation architecture, encoder and highest $IoU$.

| S.A. | Encoder | $IoU_{max}$ |
|---|---|---|
| UNet | DenseNet 121 | 0.8223 |
| | Xception | 0.8285 |
| | VGG13 BN | 0.8367 |
| | MobileNet V3 | 0.8132 |
| | ResNeXt 50 | 0.8282 |
| UNet++ | DenseNet 121 | 0.8275 |
| | Xception | 0.8377 |
| | VGG13 BN | 0.8309 |
| | MobileNet V3 | 0.8201 |
| | ResNeXt 50 | 0.8309 |

**Table 4.** UNet optimal parameters.

| Encoder | LR | Epochs | BS | Optimizer |
|---|---|---|---|---|
| DenseNet 121 | $2.9 \times 10^{-4}$ | 217 | 12 | Adam |
| Xception | $4.7 \times 10^{-4}$ | 115 | 12 | Adam |
| VGG13 BN | $5.0 \times 10^{-4}$ | 190 | 11 | Adam |
| MobileNet V3 | $3.7 \times 10^{-3}$ | 186 | 20 | Adam |
| ResNeXt 50 | $1.8 \times 10^{-4}$ | 123 | 20 | Adam |

## 3.2  Segmentation Results

Each pair architecture/encoder was trained using the optimal parameters presented in Tables 4 and 5. The final loss function value ($H^{\star}(y, \bar{y})$) and training time are presented in Tables 8 and 9 along the resulting $IoU$ of each model applied to a image sample extracted from the test dataset. For illustration purposes, Fig. 2a and 2b present the image sample and respective segmentation mask used to evaluate the $IoU$ value for Tables 8 and 9 while Figs. 3a to 3b the predicted segmentation mask for UNet/VGG13 and UNet++/MobileNet respectively with their corresponding difference from the ground truth in Figs. 3c and 3d.

**Table 5.** UNet++ optimal parameters.

| Encoder | LR | Epochs | BS | Optimizer |
|---|---|---|---|---|
| DenseNet 121 | $9.7 \times 10^{-4}$ | 241 | 21 | Adam |
| Xception | $4.6 \times 10^{-4}$ | 217 | 12 | Adam |
| VGG13 BN | $3.0 \times 10^{-4}$ | 108 | 19 | Adam |
| MobileNet V3 | $1.3 \times 10^{-3}$ | 164 | 12 | Adam |
| ResNeXt 50 | $9.0 \times 10^{-4}$ | 116 | 37 | Adam |

**Table 6.** Hyperparameter relevance over $IoU$: UNet.

| Encoder | Hyperparameter 1 | Hyperparameter 2 |
|---|---|---|
| DenseNet 121 | LR (70%) | Optimizer (24%) |
| Xception | Optimizer (63%) | Batch size (17%) |
| VGG13 BN | Batch size (54%) | Optimizer (21%) |
| MobileNet V3 | LR (56%) | Batch size (25%) |
| ResNeXt 50 | LR (51%) | Batch size (33%) |

**Table 7.** Hyperparameter relevance over $IoU$: UNet++.

| Encoder | Hyperparameter 1 | Hyperparameter 2 |
|---|---|---|
| DenseNet 121 | Epochs (36%) | Optimizer (35%) |
| Xception | Optimizer (57%) | LR (34%) |
| VGG13 BN | Optimizer (37%) | LR/ Batch size (24%) |
| MobileNet V3 | Optimizer (73%) | LR (15%) |
| ResNeXt 50 | Optimizer (38%) | Batch size (31%) |

**Table 8.** $IoU$, $H^\star(y, \bar{y})$ and training time for UNet.

| SA | $IoU$ | $H^\star(y, \bar{y})$ | Training time (min) |
|---|---|---|---|
| DenseNet 121 | 0.6412 | 0.2320 | 14.73 |
| Xception | 0.8055 | 0.0192 | 20.64 |
| VGG13 BN | 0.8233 | 0.0345 | 18.33 |
| MobileNet V3 | 0.7827 | 0.0313 | 13.25 |
| ResNeXt 50 | 0.8123 | 0.0345 | 19.52 |

**Table 9.** $IoU$, $H^\star(y, \bar{y})$ and training time for UNet++.

| SA | $IoU$ | $H^\star(y, \bar{y})$ | Training time (min) |
|---|---|---|---|
| DenseNet 121 | 0.7248 | 0.0275 | 25.87 |
| Xception | 0.7781 | 0.0215 | 22.29 |
| VGG13 BN | 0.8035 | 0.0346 | 14.57 |
| MobileNet V3 | 0.8296 | 0.0270 | 13.76 |
| ResNeXt 50 | 0.7919 | 0.0284 | 15.78 |

# 4   Discussion

The results presented in this paper show that, considering the $IoU$, there is
no relevant variation considering the segmentation architecture and the encoder
used given the optimal hyperparameters set. This conclusion can also be observed
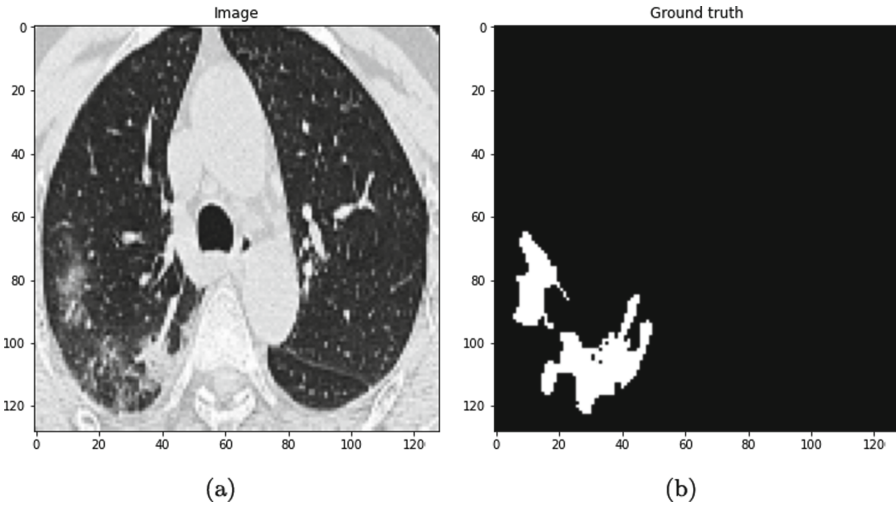
**Fig. 2.** Image sample extracted from the test dataset: a) CT image; b) Corresponding segmentation mask.

in Tables 8 and 9. Although, it is important to mention that the training time is considerably different from each architecture, with the shortest time of 13.25 minutes for UNet/MobileNet. This can be explained not only by the higher learning rate value among all other models, which directly impacts over the learning speed of the model, but also due to the fact that CNN posses less trainable parameters compared to others (Table 1). If considered the highest $IoU$ in Tables 8 and 9, the pair UNet++/MobileNet also presented the second shortest training time with 13.76 min.

Another important observation that can be made by observing Tables 4 and 5 is that the choice of the proper optimizer, learning rate and batch size are crucial in achieving the optimal values for the $IoU$, being the optimizer the most relevant hyperparamter in 80% of the models tested while the learning rate and batch size in 60%.

Although Monshi et al. [10] and Hamida et al. [5] present studies concerning hyperparameters optimization for COVID-19 detection using standard CNN's models and other ML algorithms, no study using segmentation DL models and CT was found to provide a comparison of results, which emphasizes the original aspect of the present work.
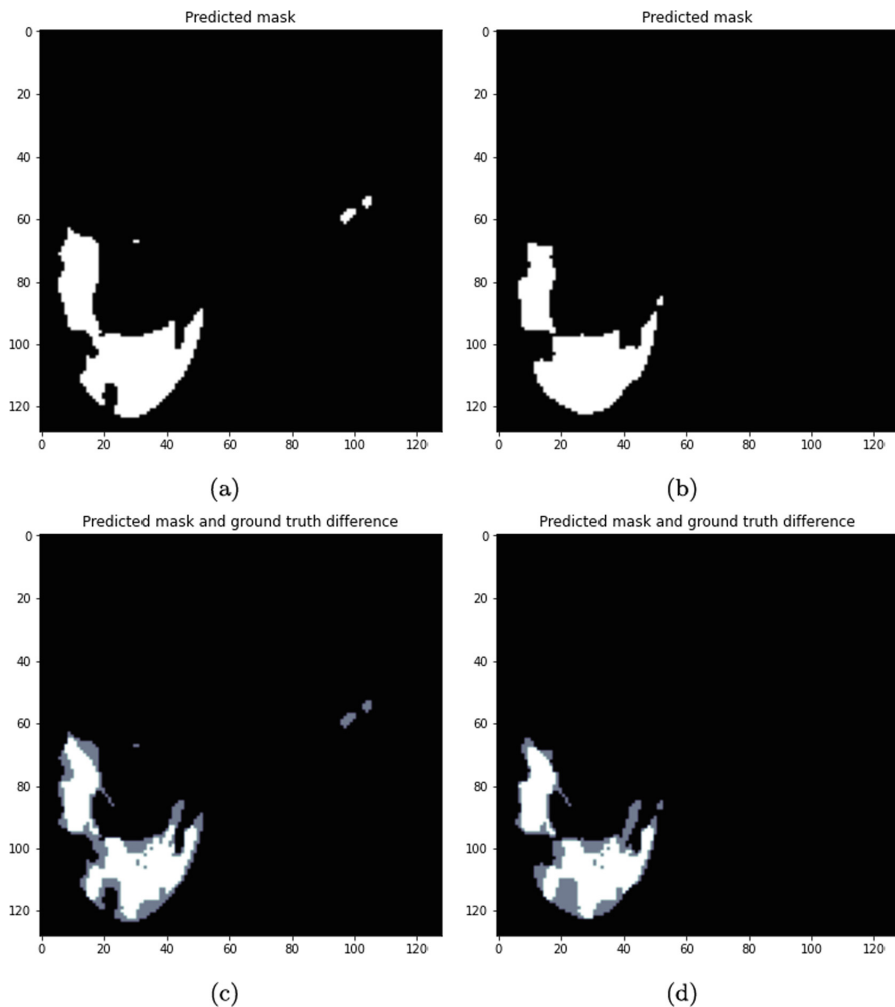
**Fig. 3.** Results of the evaluation of two segmentation models: a) UNet/VGG13; b) UNet++/MobileNet; c) Difference between the ground truth and UNet/VGG13 segmentation; f) Difference between the ground truth and UNet++/MobileNet segmentation.

## 5   Conclusion

This paper presented a study of hyperparameters tuning for segmentation architectures considering several CNN's as encoders. It was observed that given the optimal hyperparameter set, all combinations segmentation architecture/encoder did not present any relevant difference considering the $IoU$, being a decisive factor the training time and the encoder number of trainable parameters. The three most relevant hyperparameters are the optimizer, learning rate

and batch size, while the number of epochs did not presented any significant change in most of the experiments. Is valid to notice that COVID-19 lung damage is still an area of lots of questions withe the need of more data and this study intended to help fill this gap.

**Acknowledgements.** The authors would like to thank the Federal University of ABC.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)
2. Aslan, M.F., Sabanci, K., Durdu, A., Unlersen, M.F.: Covid-19 diagnosis using state-of-the-art cnn architecture features and bayesian optimization. Comput. Biol. Med. **142**, 105244 (2022). ISSN 0010-4825. https://doi.org/10.1016/j.compbiomed.2022.105244. https://www.sciencedirect.com/science/article/pii/S0010482522000361
3. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: fast and flexible image augmentations. Information **11**(2) (2020). ISSN 2078-2489. https://doi.org/10.3390/info11020125. https://www.mdpi.com/2078-2489/11/2/125
4. Chollet, F.ç.: Xception: deep learning with depthwise separable convolutions. CoRR abs/1610.02357 (2016). http://arxiv.org/abs/1610.02357
5. Hamida, S., et al.: Optimization of machine learning algorithms hyper-parameters for improving the prediction of patients infected with covid-19. In: 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), pp. 1–6 (2020). https://doi.org/10.1109/ICECOCS50124.2020.9314373
6. Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. J. Digit Imaging **32** (2019)
7. Howard, A.G., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications. CoRR abs/1704.04861 (2017). http://arxiv.org/abs/1704.04861
8. Huang, G., Liu, Z., Weinberger, K.Q.: Densely connected convolutional networks. CoRR abs/1608.06993 (2016). http://arxiv.org/abs/1608.06993
9. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2014). https://arxiv.org/abs/1412.6980
10. Monshi, M.M.A., Poon, J., Chung, V., Monshi, F.M.: Covidxraynet: optimizing data augmentation and cnn hyperparameters for improved covid-19 detection from cxr. Comput. Biol. Med. **133**, 104375 (2021). ISSN 0010-4825. https://doi.org/10.1016/j.compbiomed.2021.104375. https://www.sciencedirect.com/science/article/pii/S0010482521001694
11. Peng, J., Wang, Y.: Medical image segmentation with limited supervision: a review of deep network models (2021)
12. Rezatofighi, H., Tsoi, N., Gwak, J.Y., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: a metric and a loss for bounding box regression (2019). https://arxiv.org/abs/1902.09630

13. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015). http://arxiv.org/abs/1505.04597

14. Ruder, S.: An overview of gradient descent optimization algorithms. CoRR abs/1609.04747 (2016). http://arxiv.org/abs/1609.04747

15. SciKit Learn. https://scikit-learn.org/stable/modules/classes.html#module-sklearn.model

16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)

17. Xie, S., Girshick, R.B., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. CoRR abs/1611.05431 (2016). http://arxiv.org/abs/1611.05431

18. Zeng, N., Zuo, S., Zheng, G., Ou, Y., Tong, T.: Editorial: Artificial intelligence for medical image analysis of neuroimaging data. Front. Neurosci. **14** (2020)

19. Zhao, Q., Zheng, P., Xu, S.T.: Object detection with deep learning: a review. Trans. Neural Netw. Learn. Syst. **30** (2019)

20. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: a nested u-net architecture for medical image segmentation. CoRR abs/1807.10165 (2018). http://arxiv.org/abs/1807.10165

# Microvolt T-Wave and Ventricular Repolarization Duration Alternans

Thaís Winkert[1]([✉]) [iD], Jurandir Nadal[1] [iD], and Paulo Roberto Benchimol-Barbosa[2] [iD]

[1] Biomedical Engineering Program, COPPE Institute, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
thaiswinkert@peb.ufrj.br
[2] Hospital Universitário Pedro Ernesto, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil

**Abstract.** The ventricular tachyarrhythmia is an important cause of sudden cardiac death, impacting economic, social, and health systems. Microvolt T-wave alternans (MTWA) represent transmural gradient alternation in phase-3 ventricular action potential duration, leading to T-wave amplitude shortening in every other beat. The T-peak T-end interval (TpTe) represents the terminal ventricular repolarization duration and is explained by differences in phase-3 repolarization times from ventricular epicardial to endocardial fibers. Whether TpTe alternation occurs in patients with MTWA is unknown. This work investigated whether terminal ventricular repolarization duration alternation is present in MTWA patients. Seven 12-lead ECG records from Physionet T-wave Alternans Database were analyzed. The TpTe alternans was assessed in even and odds beats, in a series of 128 consecutive beats. In even and odd beats, TpTe intervals were ensembled and averaged, and the magnitude of TpTe alternans was defined as the difference between averages ($\Delta$TpTe). The MTWA was quantified by the difference between the average peak of the T-wave in even and odd beats. $\Delta$TpTe and MTWA were compared by Pearson's coefficient. Even and odd beats on TpTe alternans and MTWA were compared with non-paired Student t-tests. $\Delta$TpTe positively correlated with MTWA ($r = 0.5$; $p < 0.05$). In TpTe anal-ysis, averaged oddly and even beats showed significant differences ($p < 0.05$). Also, MTWA showed differences in T-wave peaks by comparing average even and odd beats. The present study provides evidence that beat-to-beat TpTe alternation is present among subjects with MTWA and is a potential marker for life-threatening ventricular arrhythmia.

**Keywords:** MTWA · T-peak T-end · Alternans · Sudden Cardiac Death

## 1 Introduction

Sudden cardiac death (SCD) is well defined as the death that occurs within one hour of onset of symptoms in witnessed cases and within 24 h of last being seen alive when it is unwitnessed [1, 2]. In 2018, SCD and unexpected cardiac death was the most common cause of death worldwide, accounting for 25% of all deaths [1], with impacts on

economic, social, and health systems. Overall, the most observed underlying mechanism related to SCD is ventricular fibrillation [3–5]. This scenario fosters the search for improved risk stratification tools and effective preventative strategies.

Microvolt T-wave alternans (MTWA) is a promising SCD risk stratification tool [6]. MTWA is thought to reflect the alternation of ventricular action potential phase-3 duration transmurally [7] and, thus, is a marker of transmural ventricular repolarization heterogeneity. MTWA is defined as a beat-to-beat alternation in the amplitude of the T-wave [7]. Several studies have shown the potential of MTWA to predict implantable cardioverter-defibrillator shocks and ventricular tachyarrhythmia events in diverse patient populations [8–11].

Prolongation of Tpeak-Tend interval (TpTe) is also considered a marker of transmural ventricular repolarization heterogeneity [12, 13]. It has been associated with an increased risk of mortality in congenital and acquired long-QT syndromes [10, 14], in hypertrophic cardiomyopathy with troponin I mutations [10, 13, 15], and in patients undergoing primary percutaneous coronary intervention for myocardial infarction [10]. TpTe is defined as the interval between the peak and the end of the T-wave and may be explained by differences in phase-3 repolarization times from ventricular epicardial to endocardial layers.

Thus, TpTe interval and MTWA represent different approaches to assessing the same physiological mechanism, phase-3 transmurally ventricular action potential duration, and both have shown potential to be used as a tool in SCD risk stratification. Nonetheless, it is unclear whether patients with MTWA also present alternations in TpTe and, furthermore, if they are correlated to each other. This work investigated if a terminal ventricular repolarization duration alternation, as represented by TpTe alternans is present in patients with T-wave amplitude alternation.

## 2   Materials and Methods

### 2.1   Database

The present study analyzed seven 12-lead ECG records from T-wave Alternans Database [16], sampled at 500 Hz, with 16-bit resolution at ±32 mV range. Overall, 77 ECG signals with different duration were analyzed. Considering a uniform analysis, signals used had 128 consecutive sinus beats (MTWA criteria).

### 2.2   Pre-Processing

To standardize and make ECG signal quality uniform across the analyzed databank signals, data pre-processing was carried out in all signals. Butterworth second-order zero-phase low-pass filtering at 30 Hz was applied to reduce electrical interference and muscle noises. The absolute first-order method [17] was employed to detect R-waves, and baseline drifting correction was accomplished by adjusting a spline function to the midpoint of two seven-point windows on the T-P segment, with windows duration set proportionally to the duration of the immediately preceding RR interval. This procedure was repeated throughout the whole filtered ECG, and the resulting function was subtracted.

## 2.3  Alternans Assessment

**MTWA.** Microvolt T-wave alternans (MTWA) was quantified by detecting T-wave peaks in 128 consecutive sinus beats. For monophasic T-waves, the T-wave peak was assessed as the absolute maximum within a search window, starting at 100 ms after R-wave and ending at the point representing 55% of the previous R-R interval duration (Fig. 1(a)). MTWA alternans was quantified by the difference between the average peak of the T-wave in even and odd beats (Fig. 1(b)).

**TpTe Alternans.** The TpTe is a well know interval in ECG, defined as the distance between the T-wave peak (maximum of T-wave search window) and the endpoint (detected the TpTe interval was calculated in the same 128 consecutive beats employed for MTWA. In even and odd beats, respectively, TpTe intervals were averaged. The magnitude of TpTe alternans was defined as the difference between even and odd TpTe averages ($\Delta$TpTe) (Fig. 1 (c)).
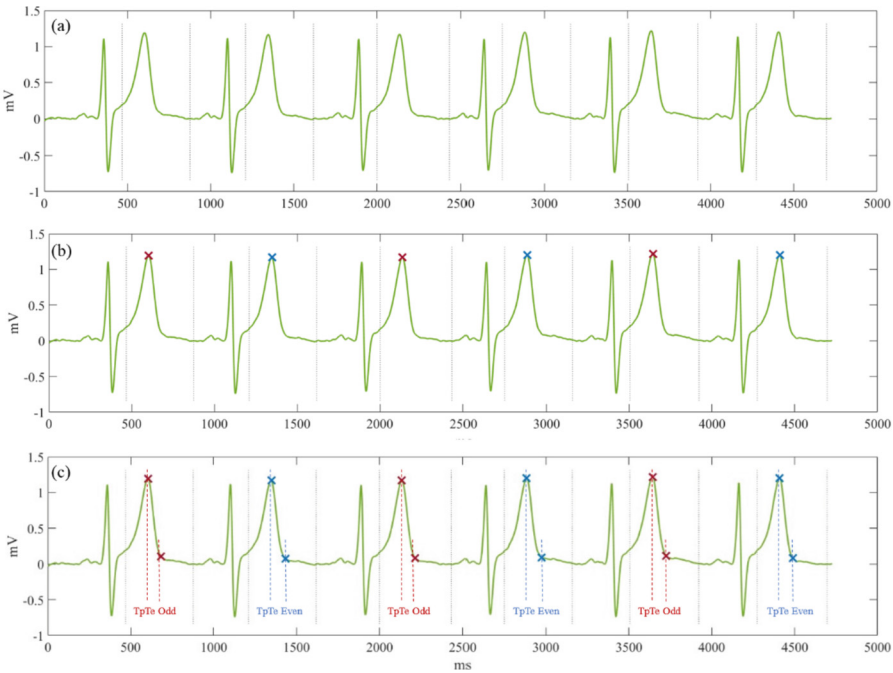


**Fig. 1.**  Alternans Assessment for both methods. A six-beat strip of a TWA36 in lead V3, in green line. (a) Result of proper T-wave assessment in T-wave search window (starting at 100 ms after R-wave and ending at the point representing 55% of the previous R-R interval duration) in dashed gray. (b) T-wave peaks were taken as the maximum of the T-wave search window. In red the odd peaks and in blue the even peaks that were further averaged for MTWA quantification. (c) Result of TpTe assessment. In red the odd TpTe and in blue the even TpTe that were further averaged for $\Delta$TpTe quantification.

## 2.4  Statistical Analysis

Data were presented as mean $\pm$ standard deviation. MTWA and $\Delta$TpTe were compared by Pearson's coefficient. Even and odd beats on TpTe alternans and MTWA were compared with non-paired Student t-test. The alpha error level was 0.05.

## 3  Results

ECG signals were successfully pre-processed in all leads, and movement artifacts and electromyogram interference in baseline were removed. Baseline drifting significantly removed the baseline oscillations, allowing an adequate signal for MTWA and $\Delta$TpTe assessment.

   The MTWA quantification was successfully performed, all T-waves peaks were found and quantified and the difference between the averaged oddly and even beats were assessed to all signals. TpTe alternans also performed successfully. Averaged TpTe was strongly correlated with heart rate (HR) in Fig. 2. Averaged even and odd beats were compared for both methods (Tables 1 and 2). Noteworthy, in even and odds beats, average
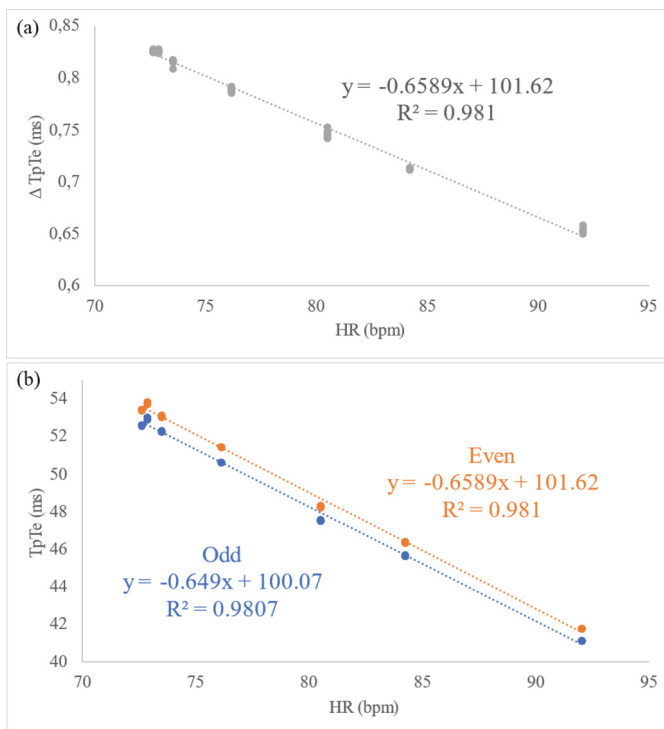


**Fig. 2.** Association of TpTe and HR. (a) Result of all beats. It is possible to notice that larger TpTe were associated with lower HR and there is a high correlation between HR and TpTe ($R^2 = 0.98$). (b) Results for even and odd groups. The same behaviour is observed and the TpTe of the odd group is slightly lower than the even group.

TpTe was significantly different across all leads. Regarding T-wave peak, in only V3, V4 and V5 even and odds beats showed significant T-wave amplitude differences $\Delta$TpTe vs. MTWA Pearson coefficient was 0.50 ($p < 0.05$).

**Table 1.** TpTe of averaged even and odd beats.

|       | Averaged Even    | Averaged Odd     | p-value      |
|-------|------------------|------------------|--------------|
| All   | 49.89 ± 4.17     | 49.12 ± 4.11     | $p < 0.05$   |
| D1    | 49.97 ± 4.87     | 49.20 ± 4.80     | $p < 0.05$   |
| D2    | 50.27 ± 4.63     | 49.50 ± 4.56     | $p < 0.05$   |
| D3    | 49.72 ± 4.48     | 48.96 ± 4.42     | $p < 0.05$   |
| aVR   | 49.72 ± 4.48     | 48.95 ± 4.42     | $p < 0.05$   |
| aVL   | 49.11 ± 4.58     | 48.35 ± 4.52     | $p < 0.05$   |
| aVF   | 49.72 ± 4.48     | 48.95 ± 4.41     | $p < 0.05$   |
| V1    | 49.72 ± 4.49     | 48.95 ± 4.43     | $p < 0.05$   |
| V2    | 51.05 ± 3.04     | 50.26 ± 3.00     | $p < 0.05$   |
| V3    | 49.74 ± 4.48     | 48.97 ± 4.42     | $p < 0.05$   |
| V4    | 49.74 ± 4.49     | 48.97 ± 4.42     | $p < 0.05$   |
| V5    | 50.28 ± 4.64     | 49.50 ± 4.57     | $p < 0.05$   |
| V6    | 49.72 ± 4.95     | 49.50 ± 4.88     | $p < 0.05$   |

**Table 2.** T-wave peak ($\mu$V) of averaged even and odd beats.

|       | Averaged Even    | Averaged Odd     | p-value      |
|-------|------------------|------------------|--------------|
| All   | 270.8 ± 169.3    | 270.3 ± 168.6    | $p < 0.05$   |
| D1    | 259.3 ± 51.3     | 258.6 ± 51.3     | $p = 0.45$   |
| D2    | 267.2 ± 129.5    | 267.7 ± 127.9    | $p = 0.70$   |
| D3    | 268.4 ± 131.1    | 268.3 ± 131.2    | $p = 0.93$   |
| aVR   | 187.1 ± 91.7     | 187.1 ± 91.0     | $p = 0.99$   |
| aVL   | 231.3 ± 111.1    | 230.6 ± 111.2    | $p = 0.32$   |
| aVF   | 234.0 ± 99.7     | 234.4 ± 98.6     | $p = 0.68$   |
| V1    | 144.1 ± 88.0     | 143.5 ± 86.5     | $p = 0.54$   |

(*continued*)

**Table 2.** (*continued*)

|  | Averaged Even | Averaged Odd | p-value |
|---|---|---|---|
| V2 | 450.7 ± 243.2 | 450.5 ± 241.8 | p = 0.75 |
| V3 | 437.4 ± 277.8 | 435.5 ± 277.1 | p < 0.05 |
| V4 | 322.1 ± 194.6 | 320.7 ± 193.7 | p < 0.05 |
| V5 | 222.2 ± 145.6 | 221.1 ± 145.4 | p < 0.05 |
| V6 | 222.9 ± 115.7 | 222.9 ± 116.3 | p = 0.96 |

## 4  Discussion

In patients with documented MTWA, the occurrence of TpTe alternans was observed, and both MTWA and TpTe alternans bear a significant correlation.

The decision to use Physionet's T-wave Alternans Data-base database choice aimed at a well-characterized dataset for MTWA information, although it was developed to setup part of requirements of a worldwide competition in the field [16].

In this study, the difference between averaged even and odd beats was chosen to assess MTWA because it conceptually best represents the idea of alternation (alternating beats). The concept of T-wave amplitude alternation was deemed an equivalent of alternation in the time duration of the phase-3 ventricular action potential. Thus, considering that TpTe conceptually represents ventricular transmural phase-3 gradient as well, it seemed rational to compare TpTe between consecutive alternating beats, and correlate TpTe duration alternation with MTWA.

The TpTe was associated with variations in heart rate (Fig. 2(a)) and larger TpTe was related with lower HR, as described in the literature [12, 18]. Also, in even and odd beats this phenomenon is still observed, as demonstrated in Fig. 2 (b). This is an indicator that the TpTe was properly assessed.

By comparing all TpTe in averaged even and odd beats, it was possible to compare two distinct groups of beats. Interestingly, it was found that even beats were larger than odd beats (on average, 0.79 ms). It indicates that there was a ventricular repolarization duration alternans in this dataset. Also, in all individual leads, significant differences between TpTe in even and odd beats were observed ($p < 0.05$).

Comparing MTWA in averaged beats showed that the T-wave amplitudes of even beats also differ from odd beats ($p < 0.05$). By comparing lead, individually, V3, V4, and V5 demonstrated significant differences between even and odd beats, suggesting that MTWA was lead-dependent. The lead V3, V4, and V5 leads are the ones that represented the larger amplitude of the ventricular repolarization vector amplitude; thus, it showed the larger amount of MTWA.

In subjects at high risk for SCD, T-wave amplitude alter-nation positively correlated with the magnitude of beat-to-beat alternation of the terminal ventricular repolarization duration (0.50; $p < 0.05$). As far as we know, this is the first time this association has been observed.

In conclusion, using different methods to assess phase-3 ventricular action potential duration, this study showed that a terminal ventricular repolarization duration alternation represented by TpTe alternans is present and correlated with micro T-wave amplitude alternation, in high-risk patients. Further studies are needed to investigate this phenomenon in other datasets.

# References

1. Srinivasan, N.T., Schilling, R.J.: Sudden cardiac death and arrhythmias. Arrhythmia Electrophysiol. Rev. **7**(2), 111–117 (2018)
2. Adabag, A.S., Luepker, R.V., Roger, V.L., Gersh, B.J.: Sudden cardiac death: epidemiology and risk factors. Nat. Rev. Cardiol. **7**(4), 216–225 (2010)
3. Haqqani, H.M., Chan, K.H., Kumar, S., Denniss, A.R., Gregory, A.T.: The contemporary era of sudden cardiac death and ventricular arrhythmias: basic concepts, recent developments and future directions. Hear. Lung Circ. **28**(1), 1–5 (2019)
4. Estes, N.A.M., III.: Predicting and preventing sudden cardiac death. Circulation **124**(5), 651–656 (2011)
5. Myerburg, R.J., Reddy, V., Castellanos, A.: Indications for implantable cardioverter-defibrillators based on evidence and judgment. J. Am. Coll. Cardiol. **54**(9), 747–763 (2009)
6. Tse, G., Wong, S.T., Tse, V., Lee, Y.T., Lin, H.Y., Yeo, J.M.: Cardiac dynamics: alternans and arrhythmogenesis. J. Arrhythmia **32**(5), 411–417 (2016)
7. Verrier, R.L., et al.: Microvolt T-wave alternans. J. Am. Coll. Cardiol. **58**(13), 1309–1324 (2011)
8. Yamada, S., et al.: Utility of heart rate turbulence and T-wave alternans to assess risk for readmission and cardiac death in hospitalized heart failure patients. J. Cardiovasc. Electrophysiol. **29**(9), 1257–1264 (2018)
9. You, T., Luo, C., Zhang, K., Zhang, H.: Electrophysiological mechanisms underlying T-wave alternans and their role in arrhythmogenesis. Front. Physiol. **12**, 614946 (2021)
10. Panikkath, R., et al.: Prolonged Tpeak-to-tend interval on the resting ECG is associated with increased risk of sudden cardiac death. Circ. Arrhythmia Electrophysiol. **4**(4), 441–447 (2011)
11. Gehi, A.K., Stein, R.H., Metz, L.D., Gomes, J.A.: Microvolt T-wave alternans for the risk stratification of ventricular tachyarrhythmic events. J. Am. Coll. Cardiol. **46**(1), 75–82 (2005)
12. Tse, G., et al.: The Tpeak − Tend interval as an electrocardiographic risk marker of arrhythmic and mortality outcomes: a systematic review and meta-analysis. Hear. Rhythm **14**(8), 1131–1137 (2017)
13. Bachmann, T.N., et al.: Electrocardiographic Tpeak-Tend interval and risk of cardiovascular morbidity and mortality: results from the Copenhagen ECG study. Hear. Rhythm **13**(4), 915–924 (2016)
14. Topilski, I., et al.: The morphology of the QT interval predicts torsade de pointes during acquired bradyarrhythmias. J. Am. Coll. Cardiol. **49**(3), 320–328 (2007)
15. Haarmark, C., et al.: The prognostic value of the Tpeak-Tend interval in patients undergoing primary percutaneous coronary intervention for ST-segment elevation myocardial infarction. J. Electrocardiol. **42**(6), 555–560 (2009)
16. Goldberger, A.L., et al.: PhysioBank, PhysioToolkit, and PhysioNet. Circulation **101**(23), e215–e220 (2000)
17. Kohler, B.-U., Hennig, C., Orglmeister, R.: The principles of software QRS detection. IEEE Eng. Med. Biol. Mag. **21**(1), 42–57 (2002)
18. Antzelevitch, C., Shimizu, W., Yan, G.-X., Sicouri, S.: Cellular basis for QT dispersion. J. Electrocardiol. **30**, 168–175 (1998)

# Study of Algorithms for Implementation of Brain-Computer Interfaces in Embedded Systems

Camila Ferrer[1]([✉]) [iD] and Marcus Vinícius Chaffim Costa[2] [iD]

[1] Electronic Engineering Graduate, University of Brasilia, Brasilia, Brazil
camilaferrer.eng@gmail.com
[2] Engineering College at Gama/Electronic Engineering, University of Brasília, Brasília, Brazil

**Abstract.** The implementation of a Brain-Computer Interface (BCI) system requires electroencephalography (EEG) signal processing that includes filtering, feature extraction, and classification algorithms. As such, the present work proposes to use the PocketBeagle embedded system to run algorithms for BCI using the Python language. This work aimed to explore some features of the embedded system to optimize performance and resource consumption, as well as the training time of the implemented algorithms, which used Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) classifiers, both with Common Spatial Patterns (CSP) filter. When comparing previous research results with the developed algorithms of this work embedded in the PocketBeagle, the training time increased by 42.98 s for LDA and 42.66 s for SVM. When analyzing the memory consumption of the implementations in the embedded system, the codes consumed less than half of the memory available in the 512 MB PocketBeagle. The consumption of the LDA classifier was 167 MB at its peak, and the SVM was 177 MB at the peak of its execution. Using the metrics resulting from the confusion matrix, it is clear that the SVM classifier had a better performance than the LDA since its accuracy is 83.14 % and its f-score is 0.8111, while for the LDA classifier, they are 66.29 % and 0.6940, respectively.

**Keywords:** Brain-Computer Interface · Embedded Systems · PocketBeagle · Python · System On Chip

## 1 Introduction

Motor imagery (MI) is the mental execution of a movement without any external physical action actually being performed. Several studies have shown that performing MI can result in the same brain areas being activated as performing the physical movement [10]. As such, brain-computer interfaces (BCI) can use brain activity associated with MI to transform those neurophysiological signals into commands for an external device [7]. These interfaces allow subjects with

motor disorders to control devices that assist them in carrying out activities, like robotic prostheses and wheelchairs.

However, the development of BCI is no longer restricted to people with diseases or undergoing treatment, for there is a shift to focus on healthy people as well. Especially the entertainment industry which is becoming a favorable market as users would likely adapt quickly to using EEG as a new modality [15]. By taking BCI to the entertainment market, the motivation to make the available systems easier to use, faster, and cheaper will take a higher priority, as current systems do not meet such standards.

The main motivation of this work is to explore the resources of the PocketBeagle embedded platform in the implementation of an interface, aiming at the use of some of the most important characteristics of this platform such as flexibility, ease of development, computational efficiency, reduced cost, and low energy consumption. In this perspective, the present work aims to develop training algorithms using Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) for execution on PocketBeagle. It seeks to understand the gains and losses related to training time, algorithm accuracy, memory consumption, and energy consumption, in order to verify the feasibility of implementing a BCI in this embedded system.

It is worth noting that this article is an excerpt from the final work for the Electronic Engineering Program at the University of Brasília, which seeks to disseminate the results of the study carried out [12].

## 2   Materials and Methods

The project is divided into four interacting parts: Embedded System, Data Set, Algorithm, and Results Validation.

### 2.1   Embedded System

PocketBeagle is an open-source, pocket-size, community-supported embedded Linux operational system. It also features a low-cost easy-to-use design, making it an ideal development board for beginners and professionals with a development environment directly in the web browser [2]. The availability of open-source software drivers also allows devices to interface with the platform.

The platform is formed by the integration of a high-performance microprocessor on a printed circuit board (PCB) and an extensive software ecosystem. Despite the impressive capabilities of this board, it does not fully expose all the features and interfaces of the ARM Sitara AM335x Cortex A8 microprocessor that it possesses. [8]. This processor at 1000 MHz executes 2000 million instructions per second. Despite not having built-in wireless connectivity, it is possible to add external modules to increase this functionality. Using PocketBeagle in a project is convenient when budget and small dimensions are considered important.

## 2.2   Data Set

For the tests and validations of this project, the Data Set IVa of the Berlin BCI Competition III was used [1]. We chose this Data Set because it is the same one used in Lotte and Guan's research [6], whose developed algorithms and results will be used as a basis for this article.

The Data Set's EEG signals files were divided based on the experiment's steps: the training and the testing of the subjects [3]. The Data Set only provides the visual cues *right* and *foot* from the subjects' training sessions. The Data Set format provides samples of continuous signals from 118 EEG channels rate (0.05–200 Hz) and markers that indicate the time points of 280 trials for each of the 5 subjects: A1, A2, A3, A4, and A5 [4].

## 2.3   Algorithm

EEG signal processing for BCI design seeks to translate the raw EEG signals into an estimation of the mental state of the user [5]. This translation is usually achieved by using a pattern recognition approach that is composed of two main steps. The first step is to extract the features in the EEG signals that are relevant to finding the desired mental state. After the extraction is done, the classification step begins where depending on the type of mental state identified a class is assigned to a set of characteristics extracted from the signals.

The initial processing of the EEG signal should remove noise and irrelevant information. As the Data Set was generated from IM, the Mu and Beta bands were taken into account to filter the brain waves recorded in the somatosensory and motor cortices. Thus, the passband between 7 Hz and 30 Hz was used in the first filter. The spatial filtering process must also be taken into account since for multichannel EEG signals, spatial filters are extremely useful for the analysis of a windowed assay and for improving the signal-to-noise ratio.

Two algorithms were implemented in Python that perform the filtering and classification of the EEG signal from the Data Set. The first algorithm used the LDA classifier and the second the SVM classifier. Both were executed on a personal computer, and then later on the PocketBeagle platform to acquire the relevant data for this work.

## 2.4   Results Validation

The first step to validate the results of this work was to reproduce the results found by Lotte and Guan [6] with the algorithm implemented by them and define the algorithm training time, establishing a basis for comparison. The main challenge when carrying out this testing phase was to obtain accurate and reproducible results.

The confusion matrix is a visualization tool commonly used to present the results obtained by a classifier. Each column of the matrix represents instances in a predicted class, while each row represents instances in an actual class. These

performance measures will be used between the two classifiers to validate the platform as a possible system for BCI implementations.

With only two classes, the choices are structured to predict the occurrence of a single event. In this case, the two possible errors are called false positive (FP) and false negative (FN). The number of correctly classified examples of one class is called true positive (TP) and for the other class is called true negative (TN). A benefit of the confusion matrix is that it's easier to check whether the classifier is mixing two classes by labeling one as the other.

Measures can be derived from the confusion matrix such as accuracy, precision, and sensitivity, also known as recall [9]. The most direct way to measure the performance of a classifier is to calculate its accuracy, which is defined in Eq. 1.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

However, in case of unbalanced data, observing only the accuracy measurement can lead to a wrong conclusion regarding the performance of the classifier used since the majority class can cover the low performance of the minority class [13]. By measuring the quality of an information retrieval process using recall and precision, one can have a better analysis of the performance of the classifier [11]. The recall describes the integrity of the retrieval, being defined as the share of positive examples achieved by the process versus the total number of existing positive examples as shown in Eq. 2. The precision is defined as the part of positive examples that exist in the total number of examples retrieved as shown in Eq. 3.

$$recall = \frac{TP}{TP + FN} \tag{2}$$

$$precision = \frac{TP}{TP + FP} \tag{3}$$

With these two equations, it is possible to calculate the measure called f-score which is a weighted average between precision and recall as shown in Eq. 4. When analyzing this measure, a result closer to 1 indicates a better performance of the classifier [13].

$$f - score = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

Therefore, for this work, accuracy and f-score measures will be used to validate the developed classifiers.

## 3   Results and Discussion

### 3.1   Reproduction of the Algorithm in Octave

Lotte and Guan's article [6] presents the classification of EEG signals by an LDA algorithm. To execute the algorithm in Octave, a DELL laptop was used

with an Intel Core i7-8550U CPU 4.0 GHz, 250 GB SSD, 8 GB of RAM, NVIDIA GeForce MX150 video card, and Windows 10 operating system.

Each type of BCI needs a calibration phase in terms of spatial or temporal filters and classifiers. This phase is fundamental for the functioning of the interface, as it is where the system defines the parameters to extract the relevant information from the EEG signal. However, as it usually demands a lot of user time, this is considered one of the diffusion limitations of the BCI system in real applications [14].

Thus, solutions must be sought with the shortest possible user calibration time, even if at some level it compromises the accuracy of the interface. For this purpose, the Common Spatial Patterns (CSP) filter was chosen as the basis for the implementation of the algorithms in Python, as it presents the shortest training time in the reproduction in Octave.

The accuracy and training time for each subject during the reproduction of the algorithm in Octave can be seen in Table 1. The accuracy results found in this execution were the same as those found by Lotte and Guan [6], validating the reproduction of the algorithm. The loading time of the Data Set files to memory was not taken into account in calculating the results, only the training time of the algorithm and the training of the classifier.

**Table 1.** Accuracy and training time of the reproduced CSP algorithm

| Subjects | Accuracy | Training Time (ms) |
|----------|----------|--------------------|
| A1 | 66.07 % | 215.36 |
| A2 | 96.43 % | 263.13 |
| A3 | 47.45 % | 117.96 |
| A4 | 71.87 % | 95.59 |
| A5 | 49.60 % | 74.49 |
| Mean | 66.28 % | 153.33 |

## 3.2   Algorithm with LDA Classifier

The researchers performed the Python code execution on the same machine and on the same operating system, as any divergence could lead to distorted values. As seen in Tables 2, 3, 4 5 and 6, the implementation of the algorithm with CSP and LDA generated the confusion matrices of each subject. The *right* and *foot* classes correspond to *Class 0* and *Class 1*, respectively.

From the confusion matrices of the LDA classifier, the researchers calculated values of accuracy, precision, recall, and f-score of each subject. With these values and the training times of each subject, we calculated the averages and organized the Table 7.

**Table 2.** Subject A1's confusion matrice for the LDA algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | Class 0 | Class 1 |
| Class 0 | 35 | 13 |
| Class 1 | 25 | 39 |

**Table 3.** Subject A2's confusion matrice for the LDA algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | Class 0 | Class 1 |
| Class 0 | 28 | 2 |
| Class 1 | 0 | 26 |

**Table 4.** Subject A3's confusion matrice for the LDA algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | Class 0 | Class 1 |
| Class 0 | 75 | 80 |
| Class 1 | 23 | 18 |

**Table 5.** Subject A4's confusion matrice for the LDA algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | Class 0 | Class 1 |
| Class 0 | 53 | 6 |
| Class 1 | 57 | 108 |

**Table 6.** Subject A5's confusion matrice for the LDA algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | Class 0 | Class 1 |
| Class 0 | 111 | 116 |
| Class 1 | 11 | 14 |

**Table 7.** LDA algorithm performance measures

| Subjects | Accuracy | Precision | Recall | f-score | Training Time (ms) |
|---|---|---|---|---|---|
| A1 | 66.07 % | 72.92 % | 58.33 % | 0.6481 | 1099.39 |
| A2 | 96.43 % | 93.33 % | 100 % | 0.9655 | 2419.87 |
| A3 | 47.45 % | 48.39 % | 76.53 % | 0.5929 | 442.81 |
| A4 | 71.88 % | 89.83 % | 48.18 % | 0.6272 | 325.09 |
| A5 | 49.60 % | 48.90 % | 90.98 % | 0.6361 | 143.61 |
| Mean | 66.29 % | 70.67 % | 74.81 % | 0.6940 | 886.15 |

The values found for accuracy were very close when comparing the results obtained in Python with the results of the reproduction in Octave. In Octave, there was an average of 66.28 %, and in Python of 66.29 %. Regarding the training time, it can be noted that the Octave program had a better performance compared to the Python program.

We performed a code profiler to identify which functions consume the most execution time. This means measuring the time spent on each function to provide data on where the program is spending the most time running and which area is worth optimizing.

The function with the longest accumulated time in the execution of the code is responsible for extracting the CSP matrix of the subjects from the training data and it corresponds to 52.4 % of the execution time of the program. The second largest execution time consumption is the function responsible for filtering EEG signals to remove noise and it corresponds to 31.2 % of the execution time.

### 3.3   Algorithm with SVM Classifier

For the implementation of the code with the SVM classifier, the same code elaborated with the LDA classifier was the basis. The researchers removed the discriminant analysis functions, inserting the support vector machine functions. However, the code structure remained similar. After elaborating the algorithm in Python, we executed it on the computer previously mentioned. Thus, Tables 8, 9, 10, 11 and 12 show the confusion matrices of each subject generated by the implementation of the algorithm with CSP and SVM.

**Table 8.** Subject A1's confusion matrice for the SVM algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | Class 0 | Class 1 |
| Class 0 | 60 | 0 |
| Class 1 | 2 | 50 |

**Table 9.** Subject A2's confusion matrice for the SVM algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | Class 0 | Class 1 |
| Class 0 | 28 | 0 |
| Class 1 | 0 | 28 |

**Table 10.** Subject A3's confusion matrice for the SVM algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | Class 0 | Class 1 |
| Class 0 | 67 | 31 |
| Class 1 | 13 | 85 |

**Table 11.** Subject A4's confusion matrice for the SVM algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | p1.2cmClass 0 | Class 1 |
| Class 0 | 83 | 27 |
| Class 1 | 16 | 98 |

**Table 12.** Subject A5's confusion matrice for the SVM algorithm

| Actual Values | Predicted Values | |
|---|---|---|
| | Class 0 | Class 1 |
| Class 0 | 105 | 17 |
| Class 1 | 86 | 44 |

From the confusion matrices of the SVM classifier, Table 13 was structured, which presents the data in relation to the performance measures of this classifier for each subject and the average between them.

When comparing LDA's training time with SVM's, we noticed a reduction of 1.08 % in relation to the training time with this second classifier. A high performance appeared for most of the Data Set subjects when analyzing the measures of the SVM classifier, which indicates an average accuracy of 83.14 % and an average f-score of 0.8111. When comparing the f-scores of both classifiers, it can be seen that the SVM performed better in the classification than the LDA, a difference of 14.44 %.

We made a code profiler to identify which functions consumed the most execution time in the SVM algorithm. As in the LDA's algorithm, the two functions that consumed the most execution time were the same as before: the function that calculates the CSP matrix and the function responsible for filtering the

**Table 13.** SVM algorithm performance measures

| Subjects | Accuracy | Precision | Recall | f-score | Training Time (ms) |
|---|---|---|---|---|---|
| A1 | 98.21 % | 100 % | 96.15 % | 0.9804 | 1078.48 |
| A2 | 100 % | 100% | 100% | 1.0000 | 2385.79 |
| A3 | 77.55 % | 73.28 % | 86.73 % | 0.7944 | 438.83 |
| A4 | 80.80 % | 78.40 % | 85.96 % | 0.8201 | 337.05 |
| A5 | 59.13 % | 72.13 % | 33.85 % | 0.4607 | 142.57 |
| Mean | 83.14 % | 84.76 % | 80.94 % | 0.8111 | 876.54 |

EEG signals to remove noise. The percentage of the execution time was 57.5% and 29.5 %, respectively.

However, when analyzing the classifier functions of each algorithm, the SVM functions consumed a total of 14.23 ms and the LDA functions 27.34 ms. This difference shows that it is faster to sort with SVM than with LDA on a personal computer.

### 3.4   Results of Running on PocketBeagle

After running the codes on the personal computer, they were executed on the PocketBeagle to verify memory consumption, training time, and code execution time. The PocketBeagle accuracy and f-score values remained the same as those when running on the personal computer. Comparing the algorithms, SVM had a 25.42 % increase in average accuracy compared to LDA. Regarding the average f-score, the SVM classifier had an increase of 16.87 % in relation to the LDA.

The execution time of the algorithms in PocketBeagle had very similar values: the execution of the LDA algorithm lasted 288.5 s and the SVM lasted 290.6 s. Given that the personal computer has greater computational capacity than the embedded system, this increase in execution time was expected. The average training time for the LDA algorithm was 42.98 s and for the SVM algorithm, it was 42.66 s.

Regarding PocketBeagle memory consumption, the LDA algorithm generated the graph shown in Fig. 1 and the SVM algorithm produced the graph shown in Fig. 2. In the embedded system, the LDA algorithm reached a peak of 175 MB of memory consumption while the SVM reached 185 MB.

To verify the PocketBeagle's power consumption, we used a USB tester to quantify the voltage and current consumed by the devices connected to it. The USB tester can measure a voltage from 4 V to 20 V with a resolution of 10 mV and a current from 0 to 3 A with a resolution of 10 mA. The voltage error range is ± 1 % and the current is ± 2 %. Using this information, we calculated the power of the embedded system at three different times: in an idle state, executing the LDA algorithm, and executing the SVM algorithm. The values found are in Table 14.
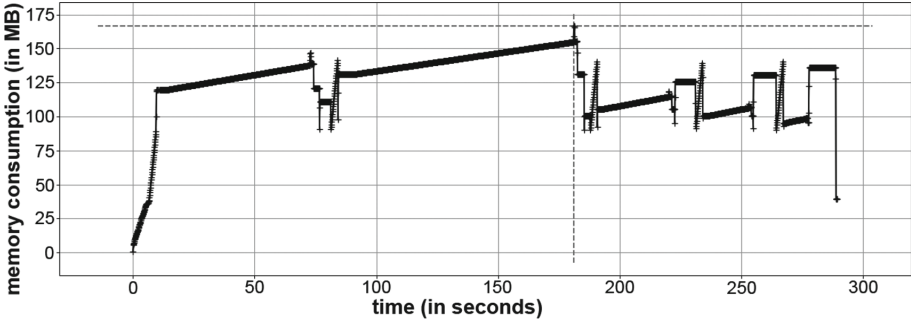
**Fig. 1.** Memory consumption as a function of time of the LDA algorithm on Pocket-Beagle
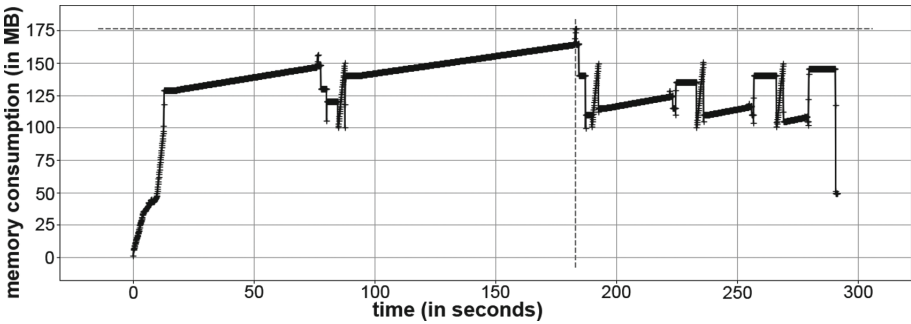


**Fig. 2.** Memory consumption as a function of time of the SVM algorithm on Pocket-Beagle

**Table 14.** PocketBeagle Power Consumption

| Measures | Idle State | LDA | SVM |
|----------|-----------|-----|-----|
| Voltage | 4.82 V | 4.81 V | 4.84 V |
| Current | 3.11 mA | 4.60 mA | 4.60 mA |
| Power | 15.0 mW | 22.1 mW | 22.3 mW |

By subtracting the PocketBeagle energy consumption while executing the algorithms with the idle state consumption, we verified that the LDA classifier in its execution used 7.1 mW and the SVM classifier used 7.3 mW. With these data, the algorithms had a power consumption very close to each other, and the LDA had a slightly better consumption.

## 4   Conclusion

As shown, this work is a study of the implementation of training algorithms for the LDA classifier and the SVM classifier in an embedded system. We compared

the results obtained using different program languages to establish a data basis for the executions in the PocketBeagle.

One of the goals was to obtain a training time in the Python implementation close to or less than in the Octave implementation. However, in relation to training time, the results presented show that on average the Python implementation was almost 6 times slower than the Octave. When embedding the Python implementations on PocketBeagle, this training time increased to an average of 42.98 s with the LDA classifier and 42.66 s with the SVM classifier. Comparing the results of running on an embedded system with running on a personal computer, the training time of the LDA classifier increased by 48.50 times and that of the SVM classifier by 48.14 times.

Due to this increase in PocketBeagle training time, it would not be recommended to execute in real-time applications the training part of the algorithms developed for this work. However, it might be possible to train a model with these algorithms beforehand and use the model already trained in real-time applications on the embedded system.

Another goal of this study was to analyze the memory consumption during the execution of the algorithms in the embedded system. The results show that the implemented interfaces consumed less than half of the available memory of 512 MB on the PocketBeagle. The consumption of the LDA classifier was 175 MB at its peak and the SVM was 185 MB.

We evaluated the performance of these algorithms using metrics derived from the confusion matrix. This method records the errors and hits in its rows and columns and generates measures such as accuracy, precision, recall, and f-score as validation metrics. It is clear when using the same Data Set that the SVM classifier has a better performance than the LDA since its accuracy is 83.14 % and its f-score is 0.8111, while for the LDA classifier, there were an accuracy of 66.29 % and f-score of 0.6940.

Observing the energy consumption of the algorithms, it is noteworthy that both have similar numbers, the LDA classifier consumed 7.1 mW and the SVM 7.3 mW. When compared to idle state power consumption, LDA and SVM increased by 47.33 %.

## References

1. BCI Competition III's Data set IVa. https://www.bbci.de/competition/iii/desc_IVa.html
2. Pocket Beagle System Reference Manual. https://github.com/beagleboard/pocketbeagle/wiki/System-Reference-Manual
3. Blankertz, B., et al.: The BCI competition iii: validating alternative approaches to actual BCI problems. IEEE Trans. Neural Syst. Rehabil. Eng. **14**(2), 153–159 (2006)
4. Dornhege, G., et al.: Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. IEEE Trans. Biomed. Eng. **51**(6), 993–1002 (2004)

5. Lotte, F.: A tutorial on EEG signal-processing techniques for mental-state recognition in brain–computer interfaces. In: Miranda, E.R., Castet, J. (eds.) Guide to Brain-Computer Music Interfacing, pp. 133–161. Springer, London (2014). https://doi.org/10.1007/978-1-4471-6584-2_7

6. Lotte, F., Guan, C.: Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. IEEE Trans. Biomed. Eng. **58**(2), 355–362 (2011)

7. Mokienko, O., et al.: Motor imagery and its practical application. Neurosci. Behav. Physiol. **44**(5), 483–489 (2014)

8. Molloy, D.: Exploring BeagleBone: Tools and Techniques for Building with Embedded Linux. Wiley, Hoboken (2019). ISBN 9781118935125

9. Monard, M.C., Baranauskas, J.A.: Conceitos sobre aprendizado de máquina. In: Sistemas Inteligentes Fundamentos e Aplicações, 1 edn, pp. 89–114. Manole Ltda, Barueri-SP (2003)

10. Mulder, T.: Motor imagery and action observation: cognitive tools for rehabilitation. J. Neural Transm. **114**(10), 1265–1278 (2007)

11. Pantic, M., et al.: Computer Based Coursework Manual Machine Learning (Course 395). Imperial College (2011)

12. Santos, C.C.F.: Estudo da implementação em sistemas embarcados de algoritmos para interfaces cérebro-máquina (2020). https://bdm.unb.br/handle/10483/27788

13. Santos, C.J.: Avaliação do uso de classificadores para verificação de atendimento a critérios de seleção em programas sociais. Master's thesis, Universidade Federal de Juiz de Fora, Departamento de Engenharia, Modelagem Computacional. Juiz de Fora, Março (2017)

14. Turi, F., Gayraud, N.T.H., Clerc, M.: Auto-calibration of c-vep bci by word prediction (2020). hal-02844024

15. Wang, A.I., Larsen, E.A.: Using brain-computer interfaces in an interactive multimedia application. In: Proceedings of the IASTED International Conference on Software Engineering and Applications (2012)

# ADHD Subtype Diagnosis Through Convolutional Neural Networks Evaluation of the Connectivity Networks in Brain fMRI

Guilherme Rodrigues Pedrollo[1](✉) 🔟, Alexandre Rosa Franco[2,3] 🔟, and Alexandre Balbinot[1] 🔟

[1] Post Graduate Program in Electrical Engineering, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 90020-000, Brazil
`guilherme.pedrollo@ufrgs.br`
[2] Child Mind Institute, New York, NY 10022, USA
[3] Nathan Kline Institute for Psychiatric Research, New York, NY 10962, USA

**Abstract.** Attention-deficit hyperactivity disorder (ADHD) affects at least 5% of the world population. ADHD typically starts in childhood and can affect normal development that can cause serious consequences during adulthood. Moreover, there are three subtypes of ADHD, each with its symptomatic differences that require particular treatment. The development of tools capable of detecting the brain differences between each subtype and a control group can help in directing future neuropsychological studies and guide treatments. Here we investigate the hypothesis of using Convolutional Neural Networks (CNN) as a tool to detect ADHD in connectivity networks of the brain. This research used the Yeo brain parcellation maps from 28 healthy volunteers, 36 combined subtype ADHD, and 26 inattentive type ADHD. The CNN achieved a 64.71% accuracy at detecting the multi-class diagnosis (Control, ADHD combined, or ADHD inattentive) when using the complete connectivity network matrix, surpassing the individual accuracies achieved when using only the within network connectivity measures. However, the separate analysis of each network showed that the Dorsal, Default and Frontoparietal networks have better accuracy of performing the classification compared to using other within brain connectivity measures. The Default Network has shown greater accuracy at identifying ADHD-I than the other within networks or whole brain connectivity measures. These results suggest that the neurological differences of the presence of ADHD and its subtypes may be present and, therefore, affect more than one connectivity network functionality.

**Keywords:** ADHD subtypes · Convolutional Neural networks · fMRI

## 1 Introduction

### 1.1 A Subsection Sample

Attention-deficit hyperactivity disorder (ADHD) affects about 51.1 million people worldwide [1]. This disease is typically first identified in children and affects between 5 and 7% according to DSM-IV [2]. Between 3 and 5% of adults have ADHD [3]. ADHD

is a perturbation of neurologic development characterized by a lack of attention, impulsivity, and hyperactivity [2]. However, each case of this disorder may affect patients with different symptom levels that limit their abilities to focus on details, hold attention, organize themselves, complete tedious or time-consuming tasks, and realize structured tasks [2]. Patients have a tendency to lose objects, pay less attention time than average, are easily distracted, and act on impulse [2]. Beyond disturbing the academic and working life of the afflicted individual, this disorder also elevates risks, such as suicide [4], traffic accidents [5], disease infection [6], substance abuse [7], or crime involvement [6]. Consequently, individuals with ADHD have a shorter life expectancy than healthy controls [8]. Therefore, it is fundamental for the ADHD population to get a diagnosis and adequate treatment.

ADHD is a disorder with three subtypes: the Inattentive (ADHD-I) presents symptoms linked to the lack of attention; the rarer hyperactive (ADHD-H) presents only the hyperactivity and lack of control over impulses; and, most commonly observed, the combined subtype (ADHD-C) [9]. The three have their particularities not only in symptoms but also in comorbidities [10, 11]. The study of Salvi et al. [11] shows that patients do not transit between subtypes. The treatments may differ between the subtypes since they have distinct brain alterations [12]. Therefore, identifying correctly the occurrence of the disorder is fundamental. Functional magnetic resonance image (fMRI) [13] can be an important tool not only to help in the diagnosis, but also to better identify the brain differences that characterize each subtype. This will help future studies to find more adequate treatments and pharmacological interventions for each subtype.

fMRI is modulated by the fluctuation of the oxygen levels in the blood (blood-oxygen-level dependent - BOLD) [13] within each brain region. The areas with an increase in the oxygen carried by the blood correspond to the areas of the brain where there is a greater neuronal activity [13]. Thus, measuring fMRI of the brain is a form of indirect measuring the functioning of the brain.

In 2011, the international neuroimaging data sharing initiative promoted ADHD-200, a competition with data from 700 participants, for which researchers were encouraged to create ADHD classifiers. The accuracies reported by the teams were within the range of 55 to 78% for the binary distinction between ADHD and control groups, though in an external hold-out testing none of these classifiers achieved more than a 61.54% accuracy [15, 18].

The research of Mao et al. [14] used fMRI data from 439 controls and 359 ADHD patients, binary classified by a space-time Convolutional Neural Network (CNN) with 71.30% of accuracy to detect dynamical differences in the brain behavior of ADHD patients. The work of Zou et al. [15] used CNN to binary classify 491 controls and 285 ADHD achieving an accuracy of 69.15%. Though, both of these works have not considered ADHD subtypes. Subtype classification has been accomplished by Qureshi et al. [9], which used data from 53 subjects of each class they worked with, namely ADHD-C, ADHD-I, and control groups from the ADHD-200 dataset. They achieved a hit rate of 76.19% at multiclass classification maximum accuracy (they did not report their average accuracy) using data from both structural MRI and fMRI using an Extreme Learning Machine. This work also analyzed the effects of age on the brain connectivity networks comparing adults and children. While previous ADHD classification studies

mostly used the ADHD-200 dataset, the present study uses the Healthy Brain Network (HBN) dataset currently being collected by Child Mind Institute, which will be further described in Sect. 2.

Section 2 also presents the preprocessing methods and classification techniques used. Sect. 3 presents the results from the application of the preprocessing and classification methodology in the dataset used and the analysis of these results. Section 4 presents the conclusions and discusses possibilities of future work.

## 2 Material and Methods

### 2.1 Healthy Brain Network Dataset

Researchers of the HBN [16] are currently collecting data for a very large database that includes MRIs EEG recordings, and medical assessments of 10,000 participants, aged between 5–21 years. Participants with other diagnostics, aside from ADHD subtypes and healthy volunteers have been excluded from the set used in this research. We have chosen to use just a subset of the HBN database to guarantee a balanced set since ANNs tend to give better results when their inputs have classification groups of similar sizes for preliminary tests.

Frequent head movement of the patients provokes distortions that greatly impairs the data. These impairments augment progressively as the patient moves more often his/her brain. The FrameWise Displacement Jenkinson is a commonly used index that measures the head movement of the subject along the exam. Individuals with an indexed Frame-Wise Displacement Jenkinson superior than 0.5 were excluded from the classification experiments. After the necessary exclusions, the present study used the fMRI data from 36 ADHD-C patients, 26 ADHD-I patients and 28 healthy volunteers from a subset of the HBN dataset acquired while each subject was watching predetermined two segments of children movies. Since the measures have been split by the database team in those two segments, the stages are treated as two different samples.

The HBN database divides the voxels measured through electromagnetic resonance according to the [17] 200-atlas, in 200 regions of interest (ROI) each with measures averaged from about 300 voxels.

### 2.2 Feature Extraction and Pre-processing

FMRI data was preprocessed using the C-PAC toolbox with its recommended settings [18]. Next a matrix reorganizing the 200 ROI using the Craddock 200 parcellation atlas [17]. These ROIs were reorganized according to the 7-segment version of the Yeo map [19]. The Yeo map divides the brain in 7 brain networks: Frontoparietal, Somatomotor, Dorsal Attention, Ventral Attention, Limbic, Default and Visual, as shown Fig. 1.
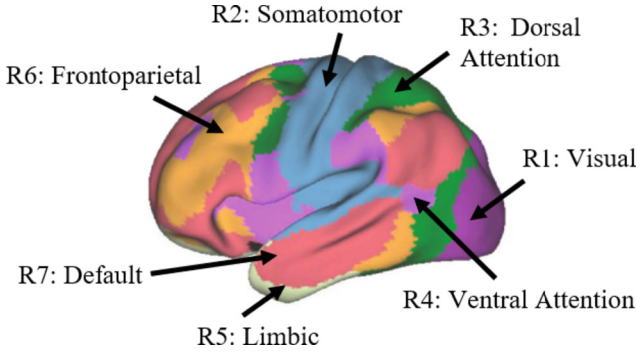
**Fig. 1.** Connectivity networks of the brain according to the Yeo 7-network cortical parcellation. SOURCE: Adapted from [19].

Next, we calculated the correlation between each ROI with and all other ROI, resulting in a 200 × 200 matrix for each subject, which corresponds to 40 thousand features. Figure 2 presents an example of these correlation matrices, whereas the squares along the main diagonal represent the correlation between the ROIs within each network, represented in the figure by R1 to R7. The correlation levels are represented in gray scale where the lighter cells represent highly positive correlations and the black cells represent no correlation. The negative correlation is not shown on this image. Whenever the value is less than zero, it is zero-clamped. However, the image is for illustrative purposes only. The negative values remain in the matrices evaluated by the classifier. The remaining rectangles show the correlation between the ROIs of one vertically adjacent network to the ROIs of a horizontal adjacent network. The eighth square in correlation matrix represented by R8 corresponds to the remaining voxels from the Craddock 200-atlas that do not belong to a defined network of the 7-segment Yeo map.

Each matrix was normalized according to (1)

$$X = \frac{X - \overline{X}}{\sigma(X)} \tag{1}$$

where X represents each correlation value, $\overline{X}$ is the average value of all correlations within the correlation matrix and $\sigma(X)$ is its standard deviation. The normalization helps in standardizing the values of the matrices for further comparison and transforms its values to a range where neural networks are known to perform better without losing its significance.

Due to some of the participants having a field of view of the MRI that misses the cerebellar areas of the brain, some regions of the CC200 needed to be removed. Therefore, 10 rows and columns from each matrix were excluded from the dataset, resulting in 190 × 190 matrices.

Next, since the correlation matrices can be divided by the main diagonal in two triangles that equally mirror each other, then sending redundant data from both triangles to the Convolutional Neural Network (CNN), which could impair its generalization potential, besides diminishing its computational efficiency. Therefore, we zeroed the upper triangle. It is noteworthy, that once the localization of each correlation value in
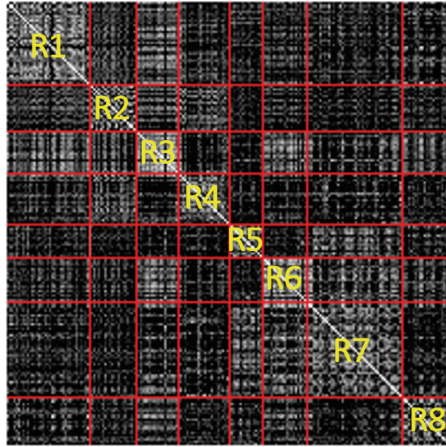
**Fig. 2.** Example of a correlation matrix in the form of an image with R1–7 denoting the Visual, Somatomotor, Dorsal Attention, Ventral Attention, Limbic, Frontoparietal and Default Networks respectively, while R8 corresponds to the remaining voxels areas (that are not characterized as part of the connectivity networks) from the Yeo 7-network cortical parcellation.
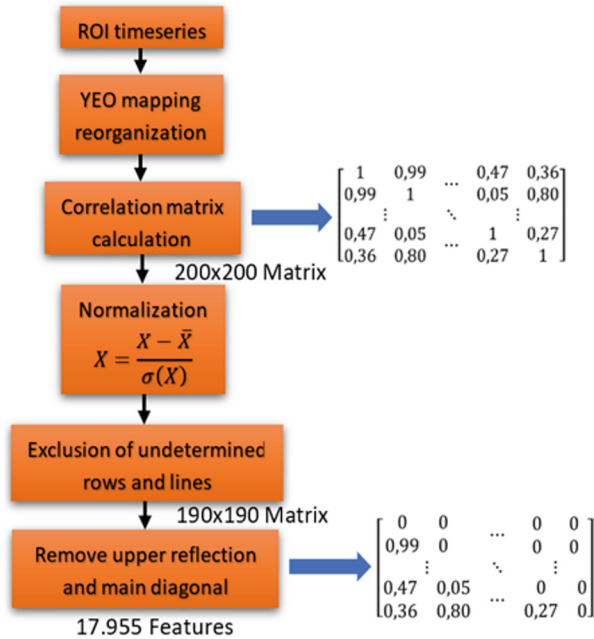


**Fig. 3.** Flowchart showing the feature extraction procedures used on preprocessed ROI timeseries prior to their use for ADHD subtype classification.

the matrix is a feature itself, reshaping the matrices into squares could reduce their relevant information. The workflow on Fig. 3 summarizes the preprocessing steps and their execution order. Therefore, the resulting matrices have 17.955 features that need to be analyzed by the classification model.

In the experiment where we consider the analysis of each connectivity network separately, there is an additional preprocessing step for this separation. As Figs. 1 and 2 show, each connectivity network has a different size that is proportional to the size it occupies in a correlation matrix, thus it will have a distinctly sized correspondent separated matrix. R1 through R8 are represented by matrices with $38 \times 38$, $21 \times 21$, $19 \times 19$, $23 \times 23$, $15 \times 15$, $20 \times 20$, $43 \times 43$ and $21 \times 21$ dimensions respectively.

### 2.3 Convolutional Neural Network

In this work, the preprocessed matrixes of each patient or volunteer are sent to a CNN that learns the patterns to identify and classify from which subtype the belongs.

Many CNN architectures were empirically tested for this purpose. The one that achieved the best results is detailed here and presented on Fig. 4(a). It starts with a pointwise convolutional layer [20] with ReLU activation, followed by a pooling layer with a $3 \times 3$ window. Next, a second convolution is applied to the data with a huge convolutional window of $50 \times 50$ with hyperbolic tangent activation (tanh) function, followed by a batch normalization and another pooling layer of $10 \times 10$. These steps have the purpose of extracting the most important information from the feature matrix and, consequently, establishing new relations between the features. These new relations matrices are then flattened. The deep neural network has two dense layers with 512 and 256 neurons and uses a dropout of 0.5 to avoid overfitting and exploding gradient problems. Finally, the processed data is sent to the output layer that defines which class the subject belongs to with 3 neurons and SoftMax function. Figure 4(b) presents the whole flowchart of the CNN.

All the results were averaged from 10 applications of the CNNs to subjects randomly sorted and divided in training (80%) and test (20%) sets.

To evaluate the activation function {sigmoid, ReLU, eLU and tanh}, batch size {2, 5, 7,10, 20, 30} and optimizer {Stochastic Gradient Descent (SGD), Adam, Adadelta, Adamax} that would produce the greatest accuracy in 10k epochs for the testing set classification, the CNN was tested with each combination through a grid search experiment. Figure 5 shows that the accuracy achieved by the CNN greatly varies, changing the optimizer and maintaining ReLU and a batch size of 7, which were the best parameters for the activation function and batch size respectively. The optimizer Adadelta surpassed the others in the majority of tests involving activation function and batch size variations. However, the use of Adamax and SGD had not always produced results as inferior as Fig. 5 might suggest and surpassed Adadelta for batch sizes of 5 and 30. ReLU and sigmoid showed better accuracy than the other activation functions, with sigmoid having less oscillation in the learning rate and ReLU providing faster learning for the CNN. A batch size = 7 provided greatest average classifying accuracy compared to other investigated sizes.

In order to evaluate the within networks connectivity classification, it was necessary to design another CNN topology, since the second layer of that CNN used filters that
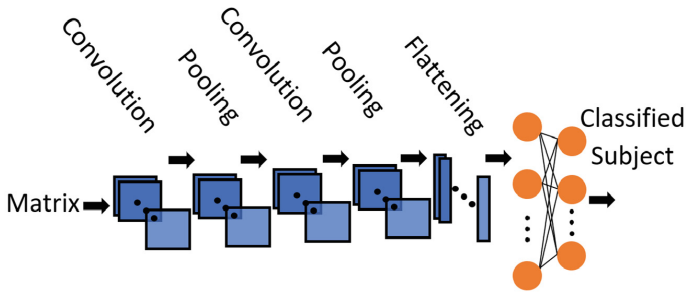
(a)
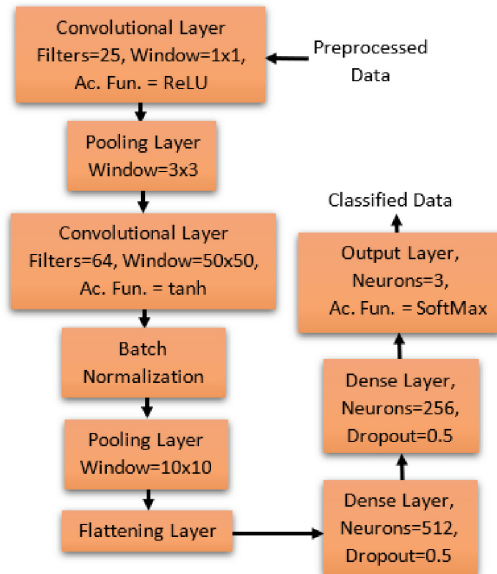


(b)

**Fig. 4.** (a) CNN Architecture (b) flowchart from the same CNN architecture showing configuration of each layer.

were bigger than the connectivity networks matrixes. Also, for a CNN to work properly its convolutional filters must be smaller than the convolved matrix. Thus, this new architecture is similar to the one described previously but has $3 \times 3$ convolutional windows in both convolutional layers and in its only pooling layer. It has two dense layers, first with 512 and second with 256 neurons, plus an output layer with 3 neurons, one for each class, and softmax activation. All remaining activations functions used in this CNN are ReLU aside from the one preceding the batch normalization that is a hyperbolic tangent.
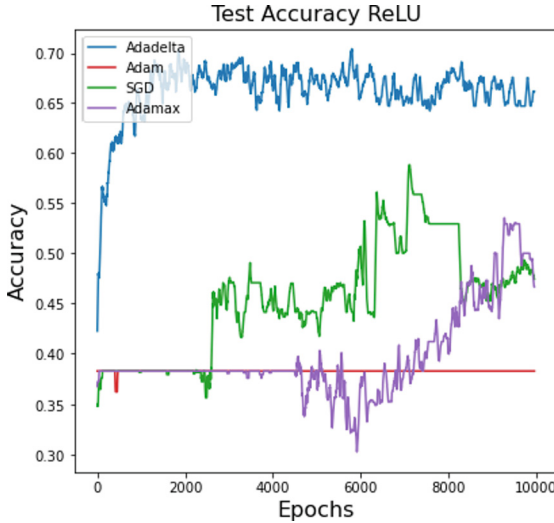
**Fig. 5.** Learning rate curve of ADHD subtype classification total accuracy at validation set achieved by the evaluation of the entire correlation matrix by the CNN with the jointly connectivity networks from Yeo 7-network cortical parcellation. Batch size = 7 and ReLU activation function.

## 3   Results and Discussion

CNN's first architecture, presented in Fig. 4, achieved 64.71% of average accuracy and 70% of maximum accuracy in classifying the validation samples. Figure 6 shows the confusion matrix of the average results achieved by the first CNN architecture at the task of classifying the complete correlation matrix. The confusion matrix shows that the CNN has 71.21%, 76.92%, and 41.67% of average accuracy at classifying control, ADHD-C, and ADHD-I groups. This difference between the classification accuracies is due to the unbalancing of the groups. The CNN classifies some ADHD-I samples as control, which infers a greater group similarity. Most ADHD-C misclassified samples are placed in the ADHD-I group, pointing to greater similarity of ADHD-C with ADHD-I than with control.

Figure 7 shows the confusion matrices achieved by evaluating the CNN with ROIs within a connectivity network. The CNN presented a better average classification accuracy for ADHD-C when using the Dorsal Attention Network (78.46%), while the greatest average accuracy for the control group was attained by the CNN classifying the Ventral Attention Network (60.91%). The Frontoparietal network also led to a considerable average accuracy in predicting ADHD-C (74.72%). Using the Default Network, CNN

```
Complete Matrix
        0      1      2

0 -   7.8    1.3    1.8                    0 – Control

1 -   1.0   10.0    2.0                    1 – ADHD-C

2 -   2.7    3.2    4.2                    2 – ADHD-I

      Predicted label
```
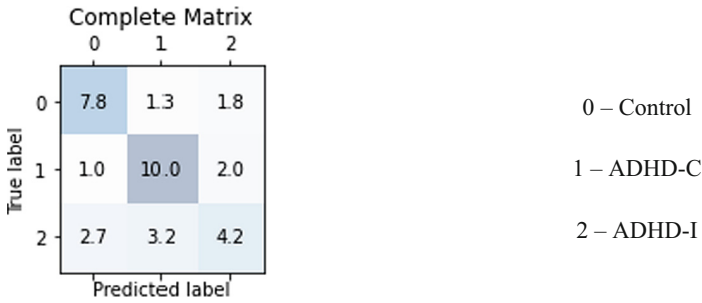
**Fig. 6.** Average of 10 confusion matrices achieved by the use of the complete correlation matrices for ADHD subtype classification with data of every connectivity network from Yeo 7-network cortical parcellation with the CNN architecture presented in Fig. 4 on the validation set.

obtained almost as good accuracy for the control group (60%) and greater accuracy for the ADHD-I group (48.45%). Considering there are three classes, an accuracy of 48.45% is higher than the average accuracy a classifier that randomly distributes the samples would get (33.33%). There is less confusion between ADHD-I and control groups when the CNN analyses the Default Network, which can indicate that this network is used differently in these groups.

The Somatomotor region has been shown to be useful for detecting controls. When CNN is wrong about predicting ADHD-C in Somatomotor Network, it classifies the sample as control. But when the CNN is using other regions, it tends to misclassify the sample as ADHD-I, thus, in the Somatomotor network ADHD-C is more similar to controls, but in the other regions ADHD-C is more like ADHD-I. The total average accuracy using the Limbic Network is the lowest, indicating less relation to the sub-types identification. Since the Dorsal, Ventral, and Frontoparietal Networks are linked to the processing of the spatial attention, reorientation of the attention upon receiving rare stimulus and sustained attention respectively, it is reasonable that the patterns of connectivity in these networks are different in ADHD and its subtypes.

Therefore, these results corroborate that the Dorsal and Ventral Attention networks are linked to ADHD occurrence as well as the subtype distinction, as reported in [9]. The results also add that the Default and Frontoparietal Networks are also important. CNN achieves its greatest accuracy using the complete matrix with the combination of all connectivity networks, rather than just one, reinforcing that multiple networks contribute to the diagnosis.
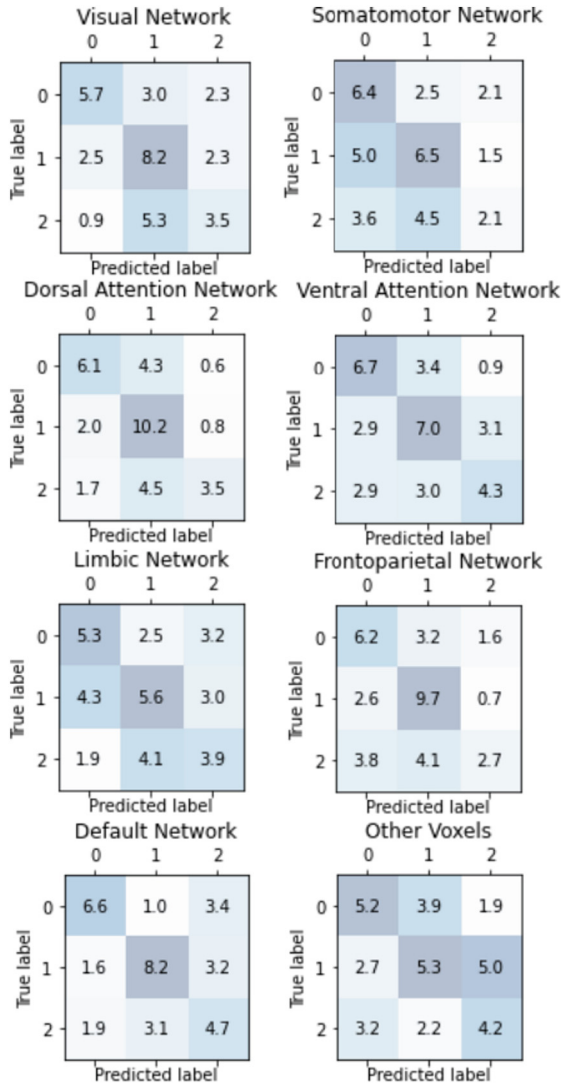
**Fig. 7.** Averages of 10 confusion matrices of the classification of ADHD subtypes based on each of the 7 connectivity networks by the second architecture of the CNN (last paragraph of Materials and Methods chapter). Legend: 0 – Control; 1 – ADHD-C; 2 – ADHD-I.

## 4  Conclusions

This manuscript has presented two approaches to classifying ADHD subtypes, namely combined and inattentive. One of the methods uses a CNN to distinguish correlation matrices of each subject's fMRI data. The other approach evaluates each connectivity network separately by a second CNN topology. The results of these analyses have shown that using Dorsal and Frontoparietal Networks has the greatest potential for identifying

the combined ADHD subtype. Using the Default Network granted the greatest average accuracy for identifying ADHD-I. Although the combination of the use of all networks simultaneously surpassed the other networks separate use in achieving average accuracy, the use of the separated connectivity networks shows the importance each network has for the diagnoses and which subtypes are more similar in which networks. That may help future studies to design more effective treatments for each ADHD subtype.

Although these results point out that the entirety of the connectivity networks may be linked to the ADHD subtyping distinction, further experiments increasing the number of samples used from the dataset are required to confirm the results. Future work may address the possibility of initial diagnosis errors, achievable through unsupervised methods.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th edn. American psychiatric association, Washington, DC (2013)
2. Faraone S., et al.: Attention-defcit/hyperactivity disorder. Nat. Rev. Dis. Primers 15020 (2015)
3. Fayyad, J., et al.: Cross-national prevalence and correlates of adult attention-deficit hyperactivity disorder. Brit. J. Psychiat. **190**(5), 402–409 (2007)
4. Furczyk, K., Tome, J.: Adult ADHD and suicide. ADHD Attent. Deficit Hyperact. Disord. **6**, 153–158 (2014)
5. Chang, Z., et al.: Serious transport accidents in adults with attention-deficit/hyperactivity disorder and the effect of medication: a population-based study. JAMA Psychiat. **71**, 319–325 (2014)
6. Barkley, R.A., Murphy, K.R., Fischer, M.: ADHD in Adults: What the Science Says, 1st edn. Guilford Press, New York (2010)
7. Dalsgaard, S., et al.: ADHD, stimulant treatment in childhood and subsequent substance abuse in adulthood - a naturalistic long-term follow-up study. Addict. Behav. **39**, 325–328 (2014)
8. Dalsgaard, S., et al.: Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. Lancet **385**, 2190–2196 (2015)
9. Qureshi, M.N., et al.: Multi-modal, multi-measure, and multi-class discrimination of ADHD with hierarchical feature extraction and extreme learning machine using structural and functional brain. Front. Human Neurosci. **11**, 157 (2017)
10. Saad, J.F., Griffiths, K.R., Korgaonkar, M.S.: A systematic review of imaging studies in the combined and inattentive subtypes of attention deficit hyperactivity disorder. Front. Integr. Neurosci. **14**, 31 (2020)
11. Salvi, V., et al.: (2019) ADHD in adults: clinical subtypes and associated characteristics. Riv. Psichiatr. **54**(2), 84–89 (2019)
12. Miranda, L., et al.: Functional MRI applications for psychiatric disease subtyping: a review. arXiv in press 2007.00126 (2020)
13. Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W.: Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. **87**(24), 9868–9872 (1990)

14. Mao, Z., et al.: Spatio-temporal deep learning method for ADHD fMRI classification. Inf. Sci. **499**, 1–11 (2019)
15. Zou, L., et al.: 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. IEEE Access **5**, 23626–23636 (2017)
16. Alexander, L., et al.: An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci. Data **4**, 170181 (2017)
17. Craddock, R.C., et al.: A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum. Brain Mapp. **33**(8), 1914–1928 (2012)
18. The ADHD-200 Global Competition. http://fcon_1000.projects.nitrc.org/indi/adhd200/results.html
19. Yeo, T., et al.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. **106**, 1125–1165 (2011)
20. Lin, M., Chen, Q., Yan, S.: Network in network. Neural and Evolutionary Computing. arXiv in press 1312.4400 (2014)

# Reconstructing Electrical Impedance Tomography 3D Brain Images with Anatomical Atlas and Total Variation Priors

Roberto G. Beraldo[1(✉)] , Leonardo A. Ferreira[1] , Fernando S. Moura[2] , André K. Takahata[1] , and Ricardo Suyama[1]

[1] Federal University of ABC, Avenida dos Estados,5001, Santo André, (SP) 09210-580, Brazil
roberto.gutierrez@ufabc.edu.br
[2] Federal University of ABC, Alameda da Universidade, s/no, São Bernardo do Campo, (SP) 09606-045, Brazil

**Abstract.** Electrical Impedance Tomography (EIT) is an imaging modality that allows the visualization of internal resistivities of a region of interest from electrical measurements external to the same region. In this work, we reconstruct 3D static images using two regularization terms, an anatomical atlas with $\ell_1$-norm and a total variation (TV) term. We chose the iteratively reweighted least squares (IRLS) algorithm to approximate the $\ell_1$-norms by quadratic terms and the Gauss-Newton algorithm to perform the optimization of the resulting functional. Together with the anatomical atlas, using a traditional $\ell_2$-norm and a high-pass filter as the regularizer tends to reconstruct the target on the mesh elements near the region boundary. In comparison, the reconstructed targets with the proposed method are better located, especially when reconstructing multiple targets, in addition to detecting a higher resistivity variation with the same number of iterations.

**Keywords:** electrical impedance tomography · regularization · total variation · sample-based density · subspace prior

## 1 Introduction

Electrical impedance tomography (EIT) is an imaging technique that aims to reconstruct resistivity values $\boldsymbol{\rho}$ in a region of interest (ROI) from electrical potential measurements $\mathbf{v}_m$ on the surface of the same region. EIT clinical applications include head, breast, lung, and stomach imaging [1]. Specifically, obtaining brain images is relevant for ischemic and hemorrhagic stroke classification, leading to fast and specific patient care [2].

Methods for the reconstruction of the $\boldsymbol{\rho}$ values from a single set of measurements are classified as absolute or static imaging, and the resulting forward operator is nonlinear with respect to $\boldsymbol{\rho}$. Solving the forward problem results in simulated electrical potentials $\mathbf{v}_c(\boldsymbol{\rho})$. Furthermore, the EIT inverse problem is severely ill-posed and we need to incorporate prior information about the

parameters to regularize it. The inverse problem solution can be written as a minimization problem given by

$$\arg \min_{\boldsymbol{\rho}} \left[ L \left( \mathbf{v}_c(\boldsymbol{\rho}), \mathbf{v}_m \right) + \lambda_1 R_1 \left( \boldsymbol{\rho} \right) + \lambda_2 R_2 \left( \boldsymbol{\rho} \right) \right], \tag{1}$$

where $L$ is a loss function (or data misfit term), both $R_1$ and $R_2$ are regularizers and $\lambda_1$ and $\lambda_2$ are regularization parameters.

Consider a generalized Tikhonov regularizer given by

$$R \left( \boldsymbol{\rho} \right) = ||\mathbf{L}(\boldsymbol{\rho} - \overline{\boldsymbol{\rho}})||_2^2, \tag{2}$$

where $\mathbf{L}$ is a regularization matrix and $\overline{\boldsymbol{\rho}}$ is a reference value. A possible prior in this form is the anatomical atlas, which considers that the expected distribution of parameters in the human body is approximately known [3, pp. 6–7]. The so-called anatomical atlas [4, p. 70] is an example of sample-based density that can be calculated from images of Computed Tomography or Magnetic Resonance Imaging, and the values are converted to the magnitude of interest, which, in the case of EIT, is the resistivity. The anatomical atlas defines a generalized Tikhonov regularization term $R$ by its covariance matrix $\boldsymbol{\Gamma}$ and mean vector $\overline{\boldsymbol{\rho}}$ [5] in the form of

$$R \left( \boldsymbol{\rho} \right) = (\boldsymbol{\rho} - \overline{\boldsymbol{\rho}})^T \boldsymbol{\Gamma}^{-1} (\boldsymbol{\rho} - \overline{\boldsymbol{\rho}}). \tag{3}$$

When it is possible to decompose the covariance matrix as $\mathbf{L}^T \mathbf{L} = \boldsymbol{\Gamma}^{-1}$, we write (3) in the form of (2).

If the samples used to calculate the anatomical atlas are similar to each other, $\boldsymbol{\Gamma}$ is ill-conditioned, or singular from a practical point of view [4, p. 70]. When this is the case, an alternative is to add an identity matrix $\mathbf{I}$ to $\boldsymbol{\Gamma}$ to guarantee the positive definiteness of $(\boldsymbol{\Gamma} + c\mathbf{I})$ [3, p. 8]. In a Maximum A Posteriori (MAP) interpretation, this is equivalent to adding Gaussian white noise to the original distribution [4, p. 79]. If we consider that $\boldsymbol{\Gamma}$ is a linear span of a subspace, adding $\mathbf{I}$ implies giving weight to all vectors of the basis, including the uninformative ones (whose eigenvalues are null).

In the literature, the authors in [6] developed an anatomical atlas of the thorax and in [7] of the head to solve the respective inverse problems. Regarding (1), the authors in both works chose the following component:

– $L \left( \mathbf{v}_c(\boldsymbol{\rho}), \mathbf{v}_m \right) = ||\mathbf{v}_c(\boldsymbol{\rho}) - \mathbf{v}_m||_2^2$, a quadratic loss function;
– $R_1 \left( \boldsymbol{\rho} \right) = (\boldsymbol{\rho} - \overline{\boldsymbol{\rho}})^T \left( \boldsymbol{\Gamma} + c\mathbf{I} \right)^{-1} (\boldsymbol{\rho} - \overline{\boldsymbol{\rho}})$ is the anatomical atlas, where $c$ is a constant and $\mathbf{I}$ is an identity matrix;
– $R_2 \left( \boldsymbol{\rho} \right) = ||\mathbf{L}_{HPF} \left( \boldsymbol{\rho} - \overline{\boldsymbol{\rho}} \right)||_2^2$, a generalized Tikhonov regularization term with regularization matrix $\mathbf{L}_{HPF}$ calculated from a high-pass filter (HPF) [8, p. 115].

Solutions obtained with the HPF and $\ell_2$-norm in every term of (1) are smoothed and abrupt transitions can be lost. On the other hand, $\ell_1$-norm promotes the sparseness of the solution. Hypothetically, if the changes in the ROI

caused by the pathology are small relative to the whole region, approximately sparse in that sense, the $\ell_1$-norm is an adequate option to regularize the problem [4, pp. 62–3]. In this work, we propose the following modifications to reconstruct static images of the brain using EIT:

– We use a new anatomical atlas of upper brain resistivity developed in [5] for EIT imaging. This is a high-resolution atlas that includes spatial information in three dimensions and dynamic information on how blood flow variations change resistivities over time. Instead of inverting its covariance matrix explicitly, we use a subspace prior [4,9,10, p. 71]. To this end, consider that $\mathbf{\Gamma}$ can be represented in terms of its eigenvalues and eigenvectors as

$$\mathbf{\Gamma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}, \tag{4}$$

where $\mathbf{Q}$ is an orthogonal square matrix, in which its columns are the eigenvectors of $\mathbf{\Gamma}$, and $\mathbf{\Lambda}$ is a diagonal matrix, where the non-zero elements are the eigenvalues of $\mathbf{\Gamma}$. When $\mathbf{\Gamma}$ is singular or ill-conditioned, some eigenvalues are null or close to zero, being possible to truncate them. Thus, the subspace prior is calculated as

$$R\left(\mathbf{\rho}\right) = ||(\mathbf{I} - \mathbf{Q}_M\mathbf{Q}_M^T)(\mathbf{\rho} - \overline{\mathbf{\rho}})||_2^2, \tag{5}$$

in the form of (2) when $\mathbf{L}_\Gamma = \mathbf{I} - \mathbf{Q}_M\mathbf{Q}_M^T$, where the columns of $\mathbf{Q}_M$ are the eigenvectors of $\mathbf{\Gamma}$ associated to positive eigenvalues higher than machine precision.
– We define the regularizers of (1) with the $\ell_1$-norm, instead of using the $\ell_2$-norm, and approximate them by using the Iteratively Reweighted Least Squares (IRLS) algorithm. Using IRLS with nonlinear inverse problems to approximate the $\ell_1$ norm was already proposed in [11, pp. 219–20]. It brings the possibility of using the anatomical atlas together with the total variation (TV) regularizer, defined as $||\mathbf{L}_d\left(\mathbf{\rho} - \overline{\mathbf{\rho}}\right)||_1$, where $\mathbf{L}_d$ is a matrix of first derivative [12, pp. 195–196]. In [13], the authors used the IRLS to obtain the TV regularizer, optimizing the resulting functional with the Newton-Raphson algorithm. In the present work, we use Gauss-Newton iterations.

## 2   Materials and Methods

No experimental data were collected in this work. We used the anatomical atlas calculated in [5] and MRI images from [14]. All simulations were performed using Python v3.7.

### 2.1   Forward Problem Solution and Data Simulation

We solved the forward problem using the finite element method (FEM). Two different models are used to generate the electric potentials, based on two different meshes of the human head. Both meshes were created after a segmented head model available online[1]. The first model was used to generate $\mathbf{v}_m$[V],

---

[1] https://www.pedeheadmod.net/pediatric-head-atlases/.

which are accurate (simulated) electrical potential measurements to replace experimental data. A second model is then used to produce $\mathbf{v}_c(\boldsymbol{\rho})$ [V], with a different discretization level, thus avoiding the inverse crime [4]. We added the electrodes positioned around the ROI using *blender*[2]. We generated the meshes by discretizing the domain in tetrahedral elements using *gmsh*[3].

Table 1 shows the characteristics of the two meshes. The anatomical atlas in [5] was calculated for different current injection frequencies and we chose the frequency of 1000[Hz] to define all resistivities.

**Table 1.** General mesh information

| Mesh used to simulate | | $\mathbf{v}_m$ | $\mathbf{v}_c(\boldsymbol{\rho})$ |
|---|---|---|---|
| Total | elements | 119792 | 24835 |
| | nodes | 23994 | 6204 |
| Scalp | elements | 54459 | 14540 |
| | $\boldsymbol{\rho}[\Omega m]$ | 3.11 | 3.11 |
| Skull | elements | 51401 | 6165 |
| | $\boldsymbol{\rho}[\Omega m]$ | 30.84 | 30.84 |
| Brain | elements | 12786 | 2984 |
| | $\boldsymbol{\rho}[\Omega m]$ | Single subject | Atlas |
| Electrodes | elements | 1146 | 1146 |
| | $\boldsymbol{\rho}t[\Omega m^2]$ | 0.02 | 0.02 |

Concerning the brain resistivities, the atlas provided in [5] was calculated using all the 109 MRI images from the original dataset in [14]. In this case, we chose a single subject from the same dataset to simulate $\mathbf{v}_m$ in the finer mesh, which would mean that there was a previous MRI exam available from that subject before the pathology. To solve the forward problem and calculate $\mathbf{v}_c(\boldsymbol{\rho})$, we projected the mean brain resistivities vector from the anatomical atlas available at [5] in the coarser mesh. In both forward and inverse problems, we set the same resistivities for the scalp and the skull, as well as the same electrode positions. We obtained the resistivities of the subject's brain tissues from a Gaussian distribution with the mean and standard deviation as defined in [5]. The obtained values were 10.20[$\Omega m$], 14.25[$\Omega m$], and 0.55[$\Omega m$] for the grey matter, white matter, and CSF, respectively.

We calculated the local conductivity matrices of the domain and the electrodes according to [15, Appendix A] and the global stiffness matrix according to [8, p. 222]. We considered a skip-16 current injection pattern, with the current value set to $10^{-3}[A]$.

To simulate the pathologies, we used spherical targets, with radii equal to 2[cm] or 1.5[cm], in different positions within the ROI, but always at the height

---

[2] https://www.blender.org/.
[3] https://gmsh.info/.

of the electrodes. Their resistivities were lower than that of the brain ($1.43[\Omega m]$) to simulate hemorrhagic stroke or higher than that of the brain ($18.52[\Omega m]$) as an ischemic stroke. We also considered both simultaneously, with opposite resistivity variation. Figure 1a-d) presents all these cases, where we projected the resistivities from the $\mathbf{v}_m$ simulation mesh to the coarser mesh. Also, Fig. 1e) presents the initial atlas resistivity. It is possible to see that there are differences in their resistivities beyond the targets themselves. In all cases, we added a white Gaussian noise $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I})$ to $\mathbf{v}_m$, where $\sigma = 0.001$[V] is the standard deviation.

## 2.2  Approximating $\ell_1$-Norms Using the IRLS

Let $\mathbf{r} = \mathbf{L}\left(\boldsymbol{\rho} - \overline{\boldsymbol{\rho}}\right)$ be a residual vector. We restrict ourselves to the case of using the IRLS algorithm to approximate the $\ell_1$-norm in the regularizers. This is done iteratively from a quadratic term [16, p. 193], [12, pp. 195–196]

$$||\mathbf{r}||_1 \approx ||\sqrt{\mathbf{W}}\mathbf{r}||_2^2, \tag{6}$$

where $\mathbf{W}$ is a diagonal reweighting matrix calculated by

$$W_{i,i}^k = \begin{cases} 1 \backslash |r_i^k|, & \text{for } |r_i^k| \geq \varepsilon \\ 1 \backslash \varepsilon, & \text{for } |r_i^k| < \varepsilon, \end{cases} \tag{7}$$

where $W_{i,i}$ is a value from $\mathbf{W}$, $k$ is the iteration number (not an exponent), and $\varepsilon$ is a tolerance we set to $10e - 5$. Note that the matrix $\mathbf{W}$ depends on $\boldsymbol{\rho}$, so both $\mathbf{W}$ and $\boldsymbol{\rho}$ need to be updated in each iteration.

## 2.3  Total Variation Regularization Using the IRLS

Total variation regularization is edge-preserving, hence suitable to reconstruct abrupt transitions. TV is defined by using $\ell_1$-norm in the regularizer together with a first derivative regularization matrix $\mathbf{L} = \mathbf{L}_d$ [12, pp. 195–196]. It is possible to implement the TV regularizer using the IRLS [13]. In the case of 3D EIT using the FEM, we calculated the regularization matrix representing the discretization of the differential operators as the 3D analogous to [17, p. 3].

## 2.4  Gauss-Newton Iterations with the IRLS

Table 2 shows the two cases we consider in this study.

Table 2. Reconstruction functional terms with $L\left(\mathbf{v}_c(\boldsymbol{\rho}), \mathbf{v}_m\right) = ||\mathbf{v}_c(\boldsymbol{\rho}) - \mathbf{v}_m||_2^2$.

| Method | $R_1(\boldsymbol{\rho})$ | $R_2(\boldsymbol{\rho})$ |
|---|---|---|
| Tikhonov* | $||\mathbf{L}_{HPF}\left(\boldsymbol{\rho} - \overline{\boldsymbol{\rho}}\right)||_2^2$ | $||(\mathbf{I} - Q_M Q_M^T)\left(\boldsymbol{\rho} - \overline{\boldsymbol{\rho}}\right)||_2^2$ |
| Proposed | $||\mathbf{L}_d\left(\boldsymbol{\rho} - \overline{\boldsymbol{\rho}}\right)||_1$ | $||(\mathbf{I} - Q_M Q_M^T)\left(\boldsymbol{\rho} - \overline{\boldsymbol{\rho}}\right)||_1$ |

*Based on [6,7], but using the subspace prior.

Iterative algorithms are necessary to reconstruct static EIT images. To minimize them, we use Gauss-Newton iterations [18, p. 36] and the IRLS algorithm

$\rho[\Omega m]$
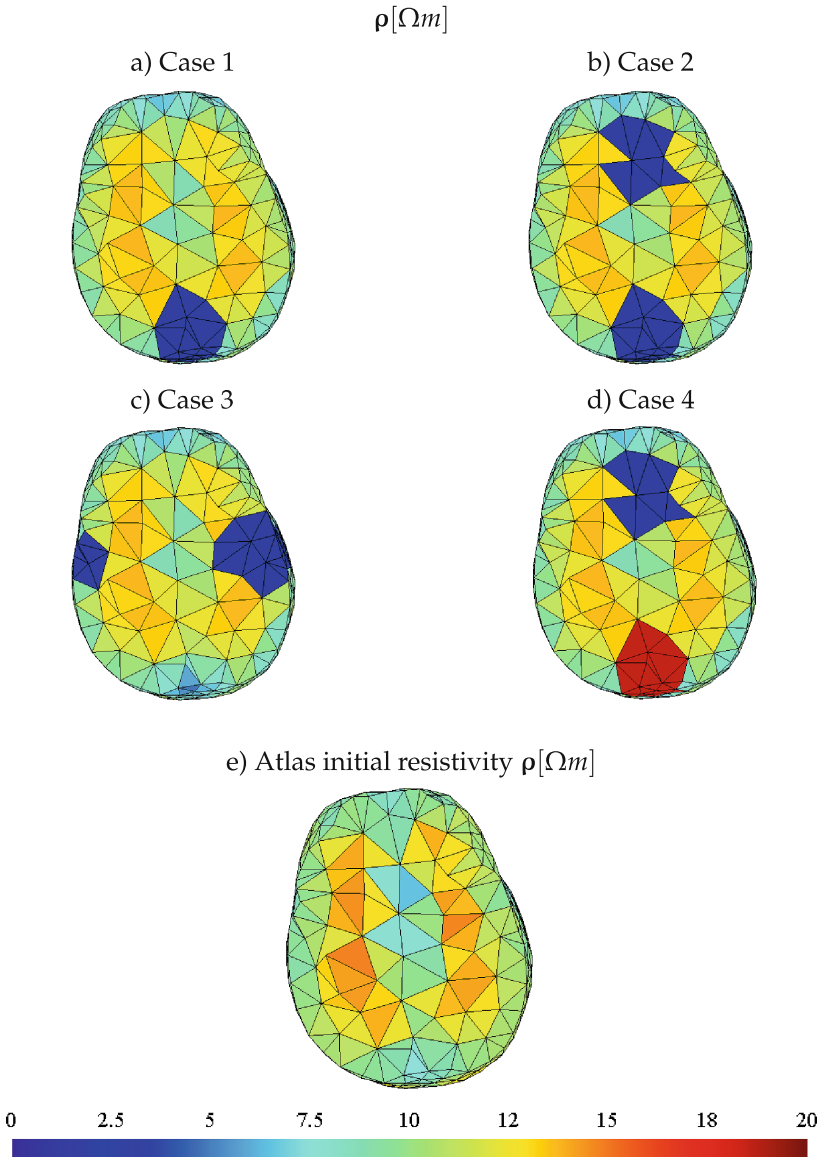


Fig. 1. a-d) Resistivities $\rho$ of the brain, used to simulate $\mathbf{v}_m$. Here, they were projected to the coarser mesh as a ground truth reference to the reconstruction. e) Atlas resistivities $\rho$ of the brain, used as the initial values in the iterative algorithm and as $\bar{\rho}$ in the regularizers.

to approximate the $\ell_1$-norm when necessary. The first step in static imaging is to linearize the forward operator around an initial resistivity $\boldsymbol{\rho}_0$ from the first-order Taylor series expansion [18, 2.49] as

$$\mathbf{v}_c(\boldsymbol{\rho}) \approx \mathbf{v}(\boldsymbol{\rho}_0) + \mathbf{J}(\boldsymbol{\rho}_0)\,(\boldsymbol{\rho} - \boldsymbol{\rho}_0)\,, \tag{8}$$

where $\mathbf{J}(\boldsymbol{\rho}_0)$ is the Jacobian of the forward problem calculated with respect to the initial resistivity $\boldsymbol{\rho}_0$ [8, p. 55]. From now on, we denote $\mathbf{J}(\boldsymbol{\rho}_0)$ by $\mathbf{J}$ only. Then, we approximate the data misfit as

$$||\mathbf{v}_c(\boldsymbol{\rho}) - \mathbf{v}_m||_2^2 \approx ||\mathbf{v}'_m - \mathbf{J}\boldsymbol{\rho}||_2^2, \tag{9}$$

where $\mathbf{v}'_m = \mathbf{v}_m - \mathbf{v}_c(\boldsymbol{\rho}_0) + \mathbf{J}\boldsymbol{\rho}_0$. It is necessary to perform a linearization step with each iteration of the algorithm, that is, the Jacobian is always recalculated for the $\boldsymbol{\rho}$ of the previous iteration.

The second step is to perform Gauss-Newton iterations. To use the IRLS in this framework, we must include the reweighting matrix for each term: $\mathbf{W}_\Gamma$ is related to the anatomical atlas and $\mathbf{W}_d$ to the TV regularizer. From (1), (6) and Table 2, the functional can be rewritten to include all cases as

$$\arg\min_{\boldsymbol{\rho}} \left|\left| \begin{pmatrix} \mathbf{v}'_m \\ \lambda_1 \mathbf{L}_d \sqrt{\mathbf{W}_d}\overline{\boldsymbol{\rho}} \\ \lambda_2 \mathbf{L}_\Gamma \sqrt{\mathbf{W}_\Gamma}\overline{\boldsymbol{\rho}} \end{pmatrix} - \begin{pmatrix} \mathbf{J} \\ \lambda_1 \mathbf{L}_d \sqrt{\mathbf{W}_d} \\ \lambda_2 \mathbf{L}_\Gamma \sqrt{\mathbf{W}_\Gamma} \end{pmatrix} \boldsymbol{\rho} \right|\right|_2^2, \tag{10}$$

where we truncated the $\Gamma$ eigendecomposition at 412 eigenvectors to obtain $\mathbf{L}_\Gamma = \mathbf{I} - Q_M Q_M^T$.

Let $k$ be the iteration number. Using Gauss-Newton iterations to solve (10) results in

$$\begin{aligned}
\mathbf{H}^k &= (\mathbf{J}^k)^T\mathbf{J}^k + \lambda_1^2 \mathbf{L}_d^T \mathbf{W}_d^k \mathbf{L}_d + \lambda_2^2 \mathbf{L}_\Gamma^T \mathbf{W}_\Gamma^k \mathbf{L}_\Gamma \\
\mathbf{U}^k &= (\mathbf{J}^k)^T(\mathbf{v}_m - \mathbf{v}_c(\boldsymbol{\rho}^k)) \\
\mathbf{S}_1^k &= \lambda_1^2 \mathbf{L}_d^T \mathbf{W}_d^k \mathbf{L}_d(\boldsymbol{\rho}^k - \overline{\boldsymbol{\rho}}) \\
\mathbf{S}_2^k &= \lambda_2^2 \mathbf{L}_\Gamma^T \mathbf{W}_\Gamma^k \mathbf{L}_\Gamma(\boldsymbol{\rho}^k - \overline{\boldsymbol{\rho}}) \\
\Delta\boldsymbol{\rho}^k &= (\mathbf{H}^k)^{-1}\left(\mathbf{U}^k - \mathbf{S}_1^k - \mathbf{S}_2^k\right) \\
\boldsymbol{\rho}^{k+1} &= \boldsymbol{\rho}^k + \alpha\Delta\boldsymbol{\rho}^k,
\end{aligned} \tag{11}$$

similar to the solution found in [18, p. 36], but now the algorithm includes $\mathbf{W}_L^k$ and $\mathbf{W}_\Gamma^k$ that are also updated with each iteration. We set the maximum number of iterations to 10 in all cases. To choose $\lambda_1$ and $\lambda_2$, we varied them from 1 to $1e-7$ in powers of ten, evaluating the results and choosing the adequate ones.

## 3    Results

Let $\hat{\boldsymbol{\rho}}$ be the reconstructed resistivity obtained after the last Gauss-Newton iteration. For better visualization, all figures in this section present the difference between $\hat{\boldsymbol{\rho}}$ to those in the initial iteration given by the anatomical atlas, that is, $\hat{\boldsymbol{\rho}} - \overline{\boldsymbol{\rho}}[\Omega m]$.

Although not shown here, it was not possible to solve the inverse problem with the anatomical atlas as the only regularizer. In this case, we compared

the regularization with and without the anatomical atlas to see its contribution. Figure 2 shows the results from the Tikhonov and the proposed methods, from Table 2, in Case 1, with just one target in the domain. It is possible to see that including the atlas results in more well-behaved reconstructions, with fewer artifacts.

Figure 3 shows the reconstructions when there were two targets in the brain. The left column refers to the Tikhonov method. We were able to identify the targets, but these mesh elements with the altered values are in the region of the border of the domain (the brain), although we see in Fig. 1 that the anterior targets (cases 2 and 4) are not in the border. Using the TV regularizer (right column) these changes, as we can see in the reconstructed anterior targets.

Also, $\hat{\boldsymbol{\rho}} - \overline{\boldsymbol{\rho}}$ values are lower with the Tikhonov method, a smoother solution. Using the proposed method, the variations $\hat{\boldsymbol{\rho}} - \overline{\boldsymbol{\rho}}$ are higher, with abrupt transitions. For example, in Fig. 3c), the minimum value was approximately $-5.6[\Omega m]$, while in Fig. 3d) the minimum value was approximately $-8.1[\Omega m]$.

## 4   Discussion

When the inverse problem is linear, such as the dynamic (or difference) EIT imaging, there are one-step solutions, such as the generalized Tikhonov regularization, that are appropriate for real-time applications. In this case, the iterative character of the IRLS can be seen as a negative side. Still, using the IRLS with Gauss-Newton iterations to reconstruct EIT static images makes it possible to consider both the nonlinear characteristic of the forward operator and the presence of the norm $\ell_1$ in the same iterative algorithm. With each iteration, the forward operator is linearized and all reweighting matrices **W** are calculated.

The results from the proposed and the Tikhonov methods from Table 2 yields different solutions, but the final results are similar to a degree since both include the anatomical atlas. While we observed an expected resistivity variation $\hat{\boldsymbol{\rho}} - \overline{\boldsymbol{\rho}}$ in the targets region, regions that are far from the electrodes, such as the base of the brain, also presents a high $\hat{\boldsymbol{\rho}} - \overline{\boldsymbol{\rho}}$. Because of this, we adjusted the scale range in each plot. Also, we observed more artifacts in the solution using further iterations, in addition to eventually reconstructing negative $\boldsymbol{\rho}$ values. For this reason, we fixed the iteration numbers to 10, resulting in an image reconstruction per 30 s with the Tikhonov method and an image per 2 min with the proposed method, both using our personal notebooks (Windows 10, Intel i7-8550U, 16Gb RAM).

Concerning the anatomical atlas, we needed it in the form of (2) because of the IRLS framework. Considering $R(\boldsymbol{\rho}) = (\boldsymbol{\rho} - \overline{\boldsymbol{\rho}})^T (\boldsymbol{\Gamma} + c\mathbf{I})^{-1} (\boldsymbol{\rho} - \overline{\boldsymbol{\rho}})$, it would be necessary to use a matrix decomposition, such as the Cholesky decomposition, to obtain a regularization matrix $\mathbf{L}_\Gamma$ that $||\mathbf{L}_\Gamma (\boldsymbol{\rho} - \overline{\boldsymbol{\rho}})||_2^2$, where $\mathbf{L}_\Gamma^T \mathbf{L}_\Gamma = (\boldsymbol{\Gamma} + c\mathbf{I})^{-1}$. Although this is a possible option, using the subspace prior is more straightforward, because it is easier to truncate the zero eigenvalues of $\boldsymbol{\Gamma}$ than choosing the constant $c$ to invert $(\boldsymbol{\Gamma} + c\mathbf{I})^{-1}$. We tested them both, but we followed with the subspace prior since we did not notice significant differences between them.

$$\hat{\rho} - \overline{\rho}[\Omega m]$$

a) $\lambda_1 = 0, \lambda_2 = 1e - 6$ 　　　　 b) $\lambda_1 = 1e - 5, \lambda_2 = 1e - 6$



Resistivity

-5　　-4　　-3　　-2　　-1　　0　　1　　2　　3

c) $\lambda_1 = 0, \lambda_2 = 1e - 5$ 　　　　 d) $\lambda_1 = 1e - 5, \lambda_2 = 1e - 5$



Resistivity

-8　　-6.2　　-4.5　　-2.8　　-1　　0.75　　2.5　　4.2　　6
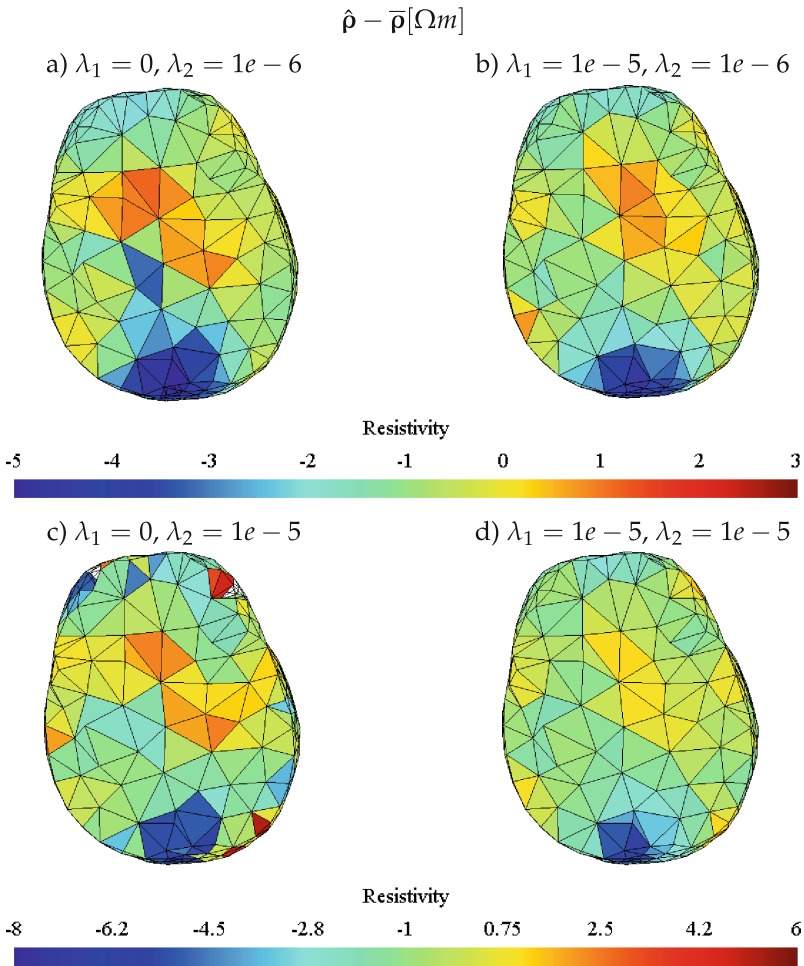
**Fig. 2.** Case 1 reconstructions to observe the influence of the anatomical atlas. a-b) Tikhonov reconstructions with HPF. c-d) Proposed reconstructions with TV.

Using $\ell_1$-norm promotes sparsity and $\ell_2$-norm in a Tikhonov framework results in smoother solutions. It is worth noting that both priors (TV and HPF) are limited in the sense that the brain is a complex structure, with smoother regions of resistivities, as well as non-smooth regions. On the other hand, both methods include the same anatomical atlas (with the difference in the norm), so it was expected that the solutions would have a degree of similarity.
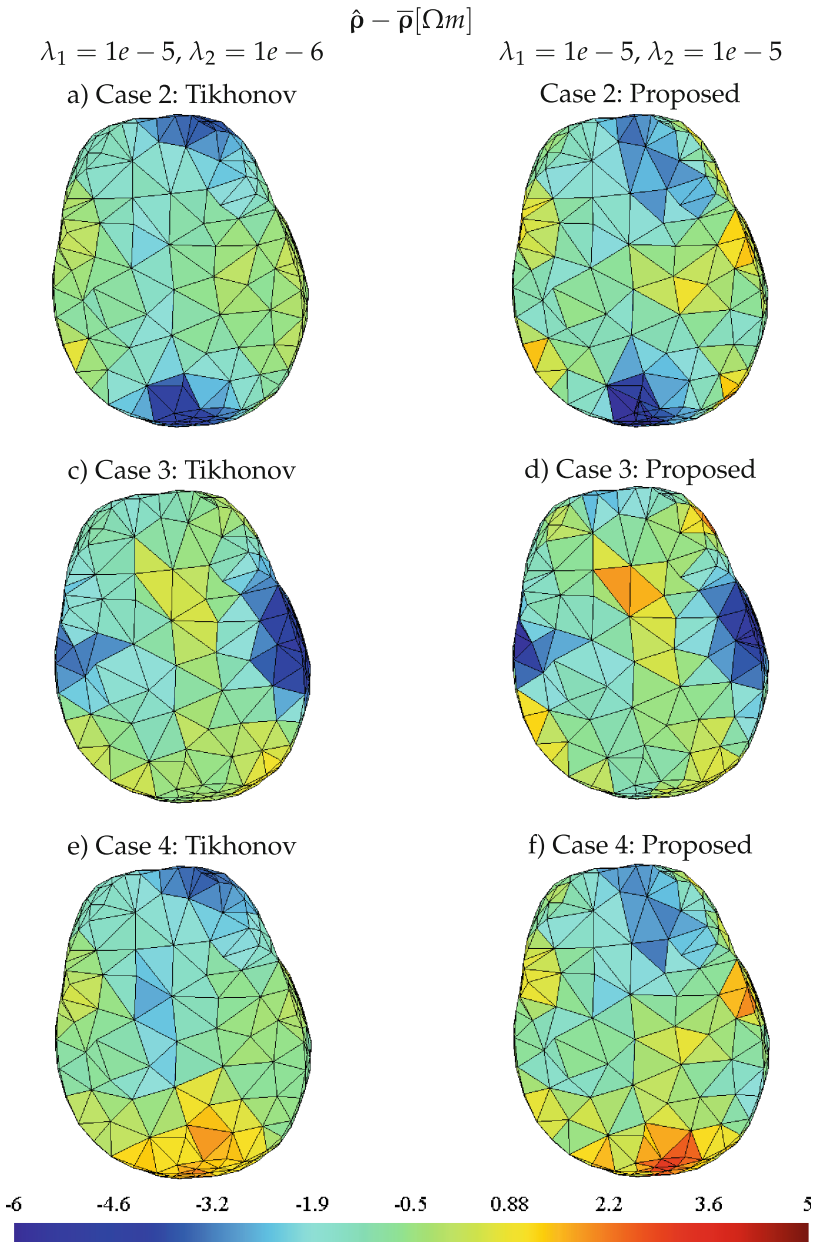
**Fig. 3.** Left column: Reconstructions with the Tikhonov method. Right column: Reconstructions with the proposed method.

## 5 Conclusion

In this work, we reconstructed brain images with EIT using an anatomical atlas together with a total variation regularizer, both with $\ell_1$-norm, leading to better-located targets and higher resistivity changes detection than the conventional HPF regularizer with the $\ell_2$-norm.

Future works include using the total generalized variation [19], a variation of TV that considers higher-order derivatives of the parameters, without presenting the staircasing effect of TV; using the IRLS to also approximate the $\ell_1$-norm in the data misfit term; using other algorithms to implement the $\ell_1$-norm; and using the method with experimental EIT data, such as those available at [2].

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Khan, T.A., Ling, S.H.: Review on electrical impedance tomography: artificial intelligence methods and its applications. Algorithms **12**(5), 88 (2019). https://doi.org/10.3390/a12050088

2. Goren, N., et al.: Multi-frequency electrical impedance tomography and neuroimaging data in stroke patients. Sci. Data **5**(1), 180112 (2018). https://doi.org/10.1038/sdata.2018.112

3. Calvetti, D., Somersalo, E.: Inverse problems: from regularization to bayesian inference. Wiley Interdisc. Rev. Comput. Stat. **10**(3), e1427 (2018). https://doi.org/10.1002/wics.1427

4. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems, pp. 978–0387220734. Springer, New York (2005). https://doi.org/10.1007/b138659

5. Moura, F.S., Beraldo, R.G., Ferreira, L.A., Siltanen, S.: Anatomical atlas of the upper part of the human head for electroencephalography and bioimpedance applications. Physiol. Meas. **42**(10), 105015, (2021). https://doi.org/10.1088/1361-6579/ac3218

6. Camargo, E.D.L.B., Moura, F.S., Luppi, O.L., Martins, F.P.R., Lima, R.G.: Converting CT scan images into impeditivity measurements to form an anatomical atlas for electrical impedance tomography. In: Proceedings of the 21st International Congress of Mechanical Engineering (COBEM2011). ABCM (2011)

7. Ferreira, L.A., Beraldo, R.G., Camargo, E.D.L.B., Moura, F.S.: Anatomical atlas of the human head for electrical impedance tomography. In: Bastos-Filho, T.F., de Oliveira Caldeira, E.M., Frizera-Neto, A. (eds.) CBEB 2020. IP, vol. 83, pp. 1693–1699. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-70601-2_247

8. Moura, F.S.: Estimação não linear de estado através do unscented kalman filter na tomografia por impedância elétrica. PhD thesis, Escola Politécnica da Universidade de São Paulo (2013)

9. Vauhkonen, M., et al.: Subspace regularization method for electrical impedance tomography. In: Proceedings of the 1st International Conference on Bioelectromagnetism. Ragnar Granit Foundation (1996)

10. Karjalainen, P.A., Kaipio, J.P., Koistinen, A.S., Vauhkonen, M.: Subspace regularization method for the single-trial estimation of evoked potentials. IEEE Trans. Biomed. Eng. **46**(7), 849–860 (1999). https://doi.org/10.1109/10.771195
11. Farquharson, C.G., Oldenburg, D.W.: Non-linear inversion using general measures of data misfit and model structure. Geophys. J. Int. **134**(1), 213–227 (1998). ISSN 0956-540X. https://doi.org/10.1046/j.1365-246x.1998.00555.x
12. Aster, R.C., Borchers, B., Thurber, C.H.: Parameter Estimation and Inverse Problems, 3rd edn., p. 9780128134238. Elsevier, Amsterdam, Netherlands (2019)
13. Shi, Y., Rao, Z., Wang, C., Fan, Y., Zhang, X., Wang, M.: Total variation regularization based on iteratively reweighted least-squares method for electrical resistance tomography. IEEE Trans. Instrum. Meas. **69**(6), 3576–3586 (2020). https://doi.org/10.1109/tim.2019.2938640
14. Bullitt, E., et al.: Vessel tortuosity and brain tumor malignancy. Acad. Radiol. **12**(10), 1232–1240 (2005). https://doi.org/10.1016/j.acra.2005.05.027
15. Beraldo, R.G.: Desenvolvimento de um modelo dinâmico da circulação cerebral para tomografia por impedância elétrica. Master's thesis, Universidade Federal do ABC, São Bernardo do Campo (2019)
16. Bovik, A.: Handbook of Image and Video Processing. Elsevier Academic Press, Amsterdam Boston, MA (2005). ISBN: 0-12-119792-1
17. Borsic, A., Brad, G.M., Adler, A., Lionheart, W.R.B.: In vivo impedance imaging with total variation regularization. IEEE Trans. Med. Imaging **29**(1), 44–54 (2010). https://doi.org/10.1109/TMI.2009.2022540
18. Vauhkonen, P.: Image reconstruction in three-Dimensional electrical impedance tomography. PhD thesis, University of Kuopio (2004)
19. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM J. Imag. Sci. **3**(3), 492–526 (2010). https://doi.org/10.1137/090769521

# Development of an Intelligent System for Detection of Chronic Stress from Biological Signal Processing

Luis Junqueira[1]([envelope]) [ORCID] and Marta Pina[2] [ORCID]

[1] Biomedical Engineering, Laboratory of Signal Processing and Medical Image, São Paulo, SP, Brazil
junqueira.umc@gmail.com

[2] Federal Institute of Education, Science and Technology, IFSP, São Paulo, SP, Brazil

**Abstract.** Our objective in this work was to develop an intelligent system able to perform an automated detection of chronic stress, based on biological signals processing and features extraction, confronted with Hans Selye clinical model of stress phase's assessment. We recorded biological signals of blood pressure, skin surface temperature, galvanic skin resistance and heart rate of 120 health adult volunteers. Also the heart rate variability parameters were extracted for the time-domain and frequency-domain analysis. A Multi-Layer Perceptron Artificial Neural Network with a supervised learning approach was applied to generate the mathematical model. The system classification produced a precision index of 89% for identifying the stressed class, when applying the skin temperature and the heart rate variability parameters as input of the neural network, providing a satisfactory initial performance in the discrimination of stressed individuals. A worth training set with more examples mighty potentially increases the precision for identifying the stressed individuals. Despite study limitations, we consider that the use of intelligent systems to classify biological signals and identify the long term stress presence in human organism could contribute to a more objective analysis of such physiologic characteristics associated to the chronic stress and its implications in human health and performance. Future works should include different settle of physiological parameters and different machine learning techniques for analysis and classification of signals.

**Keywords:** Chronic Stress · Biological Signal · Automated Detection of Stress · Intelligent System · Psychophysiological Assessment

## 1 Introduction

Stress has been considered the health epidemic of the 21st Century by the World Health Organization (WHO), which defines stress as the body's response to any type of change that causes physical, emotional or psychological strain. Informed by evidence and field testing, the WHO provides information and practical skills to help people cope with stress, especially in this COVID-19 pandemic era [1]. Hans Selye (1907–1982), a physician and researcher, established the concept of stress in biology and medicine as a General

Adaptation Syndrome [2], the body's resistance to maintain homeostasis. He developed a clinical model for stress evaluation in human organism [3]. This model distributes the stress presence in three phases: Alert, Resistance and Exhaustion. The Alert phase is related to the short-term stress, while Resistance and Exhaustion phases are related to an extended presence of stress. The assessment of stress presence is generally based on an inventory of symptoms, fulfilled by the individual. The Lipp Inventory of Stress Symptoms for Adults (ISSL) is an inventory for stress assessment based on the clinical model of stress phases by Hans Selye and covers both physical and psychological symptoms. A phase of Quasi-Exhaustion, between Resistance and Exhaustion phases, is also considered by ISSL, a quadriphasic classification instrument. It is available in English, Spanish, and Portuguese languages and has been used in clinical practice and stress research [4–9].

Chronic stress is characterized by some symptoms that individuals experience throughout their usual lives, during a more extensive period of time, which causes decreasing of their physical resistance and intellectual performance [10]. If extreme enough, chronic stress can result in the immune system impairment and psychological disorders [2]. The chronic stress is related to Resistance, Quasi-Exhaustion and Exhaustion phases and can be identified by ISSL [9]. However, the results have a component of subjectivity, since the diagnosis is based only on the answers provided, which depend on the individual's perception of their symptoms at the time of filling out the inventory [4]. Physiologically, it is possible to measure the concentration of the hormone cortisol in blood samples to determine cortisol levels over time [2]. Higher cortisol levels are related to chronically stressed individuals [9]. Alternatively, it is possible to use biological signals as indicators of stress presence [11–13], due to the physiological changes produced by the autonomic nervous system in this condition [2]. Some of the biological signals commonly used are skin temperature, galvanic skin response, blood pressure, respiration rate, heart rate, and the heart rate variability [14]. This method has the advantage of being non-invasive and low cost when compared to cortisol method, and the use of computerized systems for acquiring biological signals has extended the possibilities for the physiological variables analysis aimed to stress detection [15].

Nevertheless, the usual studies has been pointing only to identification of acute stress, by applying a physical or psychological stimulus to elicit a short-term peak of stress, that are not useful to detect the long term presence of stress, representing a challenge to establish mathematical models for analysis of physiological parameters associated to chronic stress [15]. The objective of this work was to develop an intelligent system able to perform an automated detection of chronic stress, based on biological signals processing and features extraction, confronted with Hans Selye clinical model of stress phase's assessment.

## 2 Materials and Methods

### 2.1 Ethics and Subjects

This research was approved by the ethics committee (CAAE-28201014.2.0000.5497) and adhered to the Declaration of Helsinki. We recruited 120 health adult volunteers for the study (average age $24 \pm 6$ yrs.). A consent form was assigned by the participants.

## 2.2 Psychophysiological Assessment

The psychophysiological stress assessment was made by application of ISSL. This instrument allows identifying the presence or absence of stress and, in case of stress presence, classifying it in four different phases. Additionally, ISSL presents the results in percentage terms of the individual's stress level, which can vary from 0 to 100%, depending on the presence of physiological and psychological symptoms in each phase. The use of the inventory was supervised by a psychologist. Volunteers were received in a calm and comfortable environment. The research objectives were informed and a consent form was read and signed. Subsequently, the ISSL was completed by the individuals and an additional period of five minutes was provided for relaxation and stabilization of their physiological signals.

## 2.3 Biological Signals Acquisition

After taking the measurement of systolic (SIST) and diastolic (DIAS) blood pressure of each volunteer, respecting the directives of the Brazilian Society of Cardiology [16], the biological signals relating to skin surface temperature (ST), galvanic skin resistance (GSR), and heart rate (HR) were simultaneously recorded during 6 min.

The first 30 s and the last 30 s were excluded, providing a useful 5 min short-term dataset accordant to standards of measurement for short-term recording procedures [17]. All biological signals were recorded using only non-invasive sensors.

## 2.4 Signal Processing

The block diagram for the biological signals processing is presented in Fig. 1.
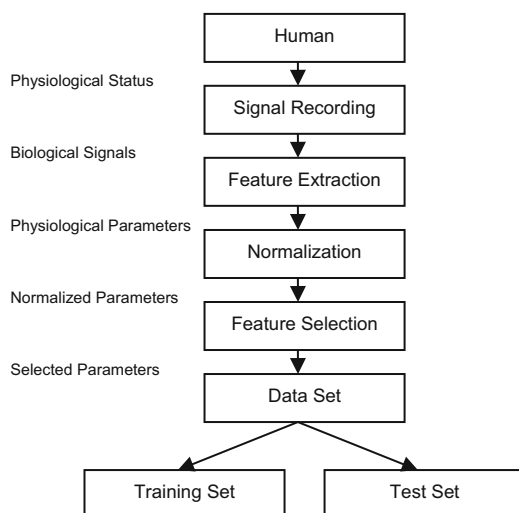


**Fig. 1.** Biological signals processing.

The current physiological status of each volunteer is captured by the sensors and the respective biological signals are recorded. The feature extraction produces the physiological parameters of interest and the normalization step produces the normalized parameters for selection and data set composition. The total data set is divided in two parts, providing the training set and the test set with a proportional number of stressed and non-stressed individuals in each set.

## 2.5   Data Normalization

A global normalization produces a linear transformation [18], by applying the Eqs. (1) to (3).

$$\bar{x}_i = \frac{1}{N} \sum_{n=1}^{N} x_i^n \tag{1}$$

$$\sigma_i = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (x_i^n - \bar{x}_i)^2} \tag{2}$$

$$z_i^n = \frac{x_i^n - \bar{x}_i}{\sigma_i} \tag{3}$$

where n = 1 to N labels the patterns. Each variable is designed by $x_i$, and each re-scaled variable is designed by $z_i$.

## 2.6   Feature Extraction

The feature extraction was applied to heart rate signals and produced the parameters of the heart rate variability (HRV) for the time-domain (Table 1) and frequency-domain analysis (Table 2).

**Table 1.**  Parameters of HRV for the time-domain analysis.

| Parameter | Description |
|-----------|-------------|
| mHR | Mean of HR |
| SDHR | Standard Deviation of HR |
| mRR | Mean of RR Intervals |
| SDNN | Standard deviation of all NN intervals |
| rMSSD | The square root of the mean of the sum of the squares of differences between adjacent NN intervals |
| NN50 | Number of pairs of adjacent NN intervals differing by more than 50 ms |
| pNN50 | NN50 count divided by the total number of all NN intervals |

**Table 2.** Parameters of HRV for the frequency-domain analysis.

| Parameter | Description |
| --- | --- |
| VLF | Power in Very Low Frequency range |
| LF | Power in Low Frequency range |
| HF | Power in High Frequency range |
| PWR | Total Spectral Power |
| LFnu | Low Frequency in normalized units |
| HFnu | High Frequency in normalized units |
| VLFp | Very Low Frequency relative percent |
| LFp | Low Frequency relative percent |
| HFp | High Frequency relative percent |
| LF/HF | Low per High Frequency Ratio |

A treatment of artifacts (corrupted intervals) and ectopic beats was performed, including the identification, removal and interpolation of signals by the Cubic Spline interpolation method. Frequency domain analysis was performed using the Fast Fourier Transform technique (FFT) to estimate the Power Spectrum Density (PSD). The frequency bands adopted for the power spectrum include the ranges of VLF (0–0.04 Hz), LF (0.04–0.15 Hz) and HF (0.15–0.40 Hz), attending to standards for measurement and physiological interpretation of these parameters [17].

## 2.7 Artificial Neural Network

The Artificial Neural Network (ANN) applied to build both the training and classifier modules is a Multi-Layer Perceptron (MLP) with backpropagation algorithm and a three layers architecture [19–21], as showed in the Fig. 2.
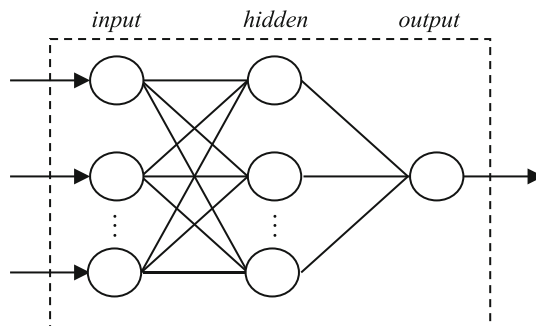


**Fig. 2.** MLP ANN with three layers.

Each input neuron corresponds to a physiological parameter extracted from the selected biological signals. The number of hidden neurons is determined by experiment

and the best results were achieved with half of input neurons number in our case. The output neuron is responsible to identify the presence or absence of stress. The nonlinear model of the ANN processing functions is presented in the Fig. 3.
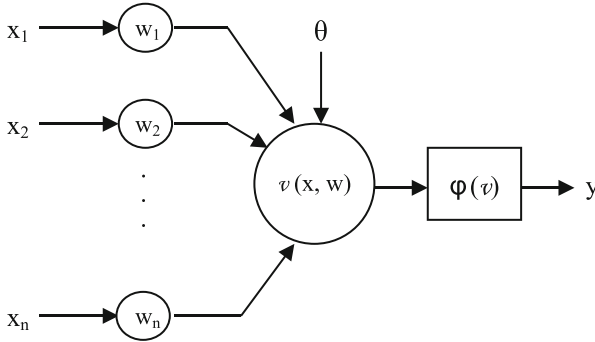


**Fig. 3.** Nonlinear model of the ANN processing functions.

The variables from $x_1$ to $x_n$ represent the input signals; the variables from $w_1$ to $w_n$ represent the synaptic weights associated to each input signal; the variable $\theta$ represents the bias; the $v(x, w)$ represents the transference function, given by Eq. (4); the $\lambda$ represents the gain, and the $\varphi(v)$ represents the sigmoid activation function, given by Eq. (5); and y represents the output. The computation of the mean square error (MSE), used as a stopping criterion for the ANN training, is given by Eq. (6), where $t_i$ is the desired target vector, $o_i$ is the attained output, and n is the number of training examples presented to network.

$$v(x, w) = \sum_{i=1}^{n} x_i.w_i = (x_1.w_1 + x_2.w_2 \cdots + x_n.w_n) \tag{4}$$

$$\varphi(v) = \frac{1}{1 + e^{\lambda.v}} \tag{5}$$

$$MSE = \frac{1}{n} \sum_{i=0}^{n-1} (t_i - o_i)^2 \tag{6}$$

A learning rate ($\eta$) controls the percentage of weight update (amount of change) and the momentum ($\alpha$) indicates the percentage of the previous weight that will remain (amount of inertia) for the next step of adjustments performed by backpropagation algorithm. Considering x the number of input neurons and z the number of hidden neurons, the final ANN training module configuration was given by $z = x/2$, $\lambda = 1$, $\eta = 0.7$ and $\alpha = 0.3$. The output (y) provides one of two classes: normal or chronically stressed.

## 2.8  Automated Stress Detection

A block diagram for the automated stress detection system is showed in the Fig. 4.
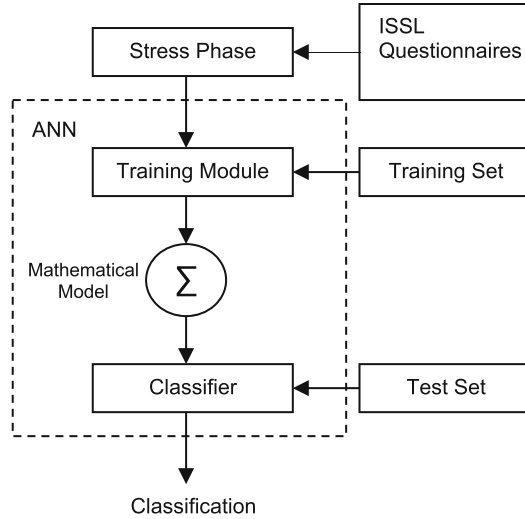
**Fig. 4.** Block diagram for automated stress detection.

The ANN in the training module is responsible for establishing the mathematical model that relates the stress presence with the physiological parameters. After training the ANN with a training data set, we applied a test data set for classification of stressed person with the classifier module.

### 2.9   Performance Evaluation

We chose the metrics for evaluating classifier performance according to precision index, sensitivity index, accuracy index, specificity index and ROC (Receiver Operating Characteristic) curves [22]. Given the True Positives (TP), the True Negatives (TN), the False Positives (FP), the Positives (P) and the Negatives (N), we can find the precision index by Eq. (7), the sensitivity index by Eq. (8), the accuracy index by Eq. (9) and the specificity index by Eq. (10).

$$precision = \frac{TP}{(TP + FP)} \tag{7}$$

$$sensitivity = \frac{TP}{P} \tag{8}$$

$$accuracy = \frac{(TP + TN)}{(P + N)} \tag{9}$$

$$specificity = \frac{TN}{N} \tag{10}$$

## 3   Results

### 3.1   Stress Presence According to ISSL

The stress presence distribution according to ISSL can be seen in Table 3. No individuals were classified in Exhaustion phase.

Table 3.  Frequency and stress level in each stress phase.

| ISSL Phase | Frequency % | Stress Level % |
| --- | --- | --- |
| Absence | 30.83 | 0.00 |
| Alert | 1.67 | 11.11 |
| Resistance | 55.83 | 28.11 |
| Quasi-Exhaustion | 11.67 | 67.26 |

### 3.2   Biological Signals Statistics

According to D'Agostino-Pearson test we verified that DIAS (p = 0.2559), ST (p = 0.0559) and HR (p = 0.4096) have normal distribution, while GSR (p < 0.0001) and SIST (p < 0.0001) don't have. The mean values for blood pressure in each stress class can be seen in Table 4.

Table 4.  The mean values of SIST and DIAS for each phase.

| ISSL Phase | SIST (mmHg) | DIAS (mmHg) |
| --- | --- | --- |
| Absence | 120.16 | 71.30 |
| Alert | 115.50 | 73.50 |
| Resistance | 118.06 | 73.61 |
| Quasi-Exhaustion | 121.36 | 77.07 |

We applied Pearson's linear correlation test and found a correlation between diastolic pressure and stress level (r = 0.9671, p < 0.05). We found that diastolic pressure values tend to increase with the phase and level of stress in which the individual is.

The mean values for ST, GSR and HR in each stress class can be seen in Table 5.

**Table 5.** The mean values of ST, GSR and HR for each phase.

| ISSL Phase | ST ($^{O}$C) | GSR (KΩ) | HR (bpm) |
| --- | --- | --- | --- |
| Absence | 24.48 | 459.84 | 80.62 |
| Alert | 24.30 | 799.23 | 89.23 |
| Resistance | 24.12 | 448.56 | 82.63 |
| Quasi-Exhaustion | 23.96 | 515.62 | 78.81 |

Applying Pearson's linear correlation test between skin temperature values and stress levels by phase, we obtained the results of r $= -0.9554$ and p $= 0.0446$ (p $< 0.05$), verifying the existence of a negative and statistically significant correlation between the variables. It is possible to observe a decrease in temperature as the stress phase in which the individual is found increases.

### 3.3   Stress Identification by the Intelligent System

The stress presence class adopted here is related specifically to chronic stress and was arranged by the set of individuals classified in Resistance and in Quasi-Exhaustion phases of ISSL. Individuals classified in Alert phase were excluded since this phase is usually related to short-term stress. No individuals were classified in Exhaustion phase.

We preferably selected the physiological parameters regarding ST, HR and blood pressure since the associated biological signals presented normal distribution.

For the first experiment (A) of system classification we utilized all the parameters of HRV in time-domain and frequency-domain (Tables 1 and 2), more ST and blood pressure values, including DIAS and SIST.

For the second experiment (B) we excluded the blood pressure values (both DIAS and SIST). The results are show in the Table 6.

**Table 6.** Performance of ANN classification for each experiment A and B.

| Index | A | B |
| --- | --- | --- |
| Precision | 80% | 89% |
| Sensitivity | 80% | 85% |
| Accuracy | 72% | 83% |
| Specificity | 56% | 78% |

The ROC curves for the system performance analysis are showed in Fig. 5.
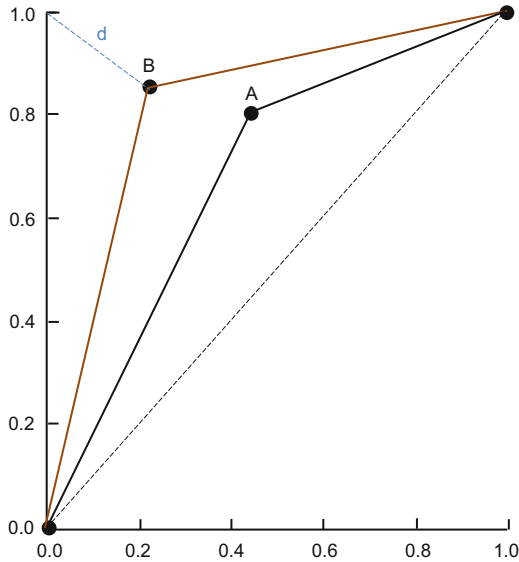
**Fig. 5.** ROC curves for each experiment A and B, where vertical axis represents the true positive ratio and the horizontal axis represents the false positive ratio.

## 4 Discussion

The results of stress assessment with ISSL in our sample are consistent with findings in the literature for lager datasets [4]. Performing the biological signals processing, features extraction and physiological parameters classification by the automated stress detection system we could find a set of synaptic weights for the ANN that was capable to produce a precision index of 89%, when applying ST and HRV parameters as input signals. In the ROC graphic the vertical axis represents the true positive ratio (sensitivity) and the horizontal axis represents the false positive ratio (1 - specificity). The line "d" represents the best performance, closer of the gold standard, i.e. true positive ratio $= 1$ and false positive ratio $= 0$. The best result is for the experiment B, with distance $= 0.27$ and the area $= 0.815$, confronted with the experiment A, with distance $= 0.48$ and the area $= 0.680$. The inclusion of blood pressure parameters decreased the system performance.

The study presents some considerations such as the limitation of sample size, the physiological parameters specificity in the input layer of the ANN and the fact that only the ISSL assessment was applied to identify chronically stressed individuals. Complementary applying of the cortisol levels method could be considered for future works as well a larger sample size for better discrimination between classes. A worth training set with more examples mighty potentially increases the precision for identifying chronically stressed individuals.

# 5 Conclusions

As far as our knowledge is concerned, no other study has attempted to relate biological signals to the presence of chronic stress in individuals, applying no stimuli, confronted with Hans Selye clinical model of stress phase's assessment by a quadriphasic instrument. The use of statistical analysis of physiological variables to identify stress presence in the quadriphasic classification presented difficulties to attain a predictable mathematical model.

We developed an intelligent system, based on features extraction and signals classification, by applying the supervised learning approach of an ANN, getting an initially satisfactory discrimination of stressed individuals in the group of volunteers. Results indicated the possibility to establish a mathematical relation, represented by the set of synaptic weights got by the ANN training, between Hans Selye clinical model and physiological parameters. Despite study limitations, we consider that the use of intelligent systems to classify biological signals and identify the long term stress presence in human organism could contribute to a more objective analysis of such physiologic characteristics associated to the chronic stress and its implications in human health and performance. Future works should include different settle of physiological parameters and different machine learning techniques for the biological signals processing, analysis and classification.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. WHO: Doing what matters in times of stress. World Health Organization, Geneva (2020)
2. Fink, G.: Stress: Concepts, Cognition, Emotion, and Behavior. Elsevier, London (2017)
3. Selye, H.: Stress and the general adaptation syndrome. Br. Med. J. **1**(4667), 1383–1392 (1950)
4. Lipp, M.: The Lipp Inventory of Stress Symptoms for Adults (ISSL). Casa do Psicólogo, São Paulo (2005)
5. Cavalcante, M., et al.: Assessment of sleep and stress level in individuals with chronic pain. Sleep Sci. **15**(2), 210–215 (2022)
6. Cardozo, A., et al.: Degree of stress in nursing residents in the pandemic. Rev. Recien **12**(38), 229–237 (2022)
7. Mazariolli, A., et al.: Stress and impact on the mental health of military police workers. Rev. Rebesp **15**(1), 165–189 (2022)
8. Wottrich, S., et al.: Gender and the manifestation of stress in hypertensive patients. Est. Psi. **28**(1), 27–34 (2011)
9. Teixeira, R., et al.: Chronic stress induces a hyporeactivity of the autonomic nervous system in response to acute mental stressor and impairs cognitive performance in business executives. PLoS ONE **10**(3), e0119025, 1–14 (2015)

10. Selye, H.: The Stress of Life. Rev McGraw-Hill, New York (1978)
11. Sierra, A., et al.: A stress detection system based on physiological signals and fuzzy logic. IEEE Trans. Ind. Electron. **58**(10), 4857–4865 (2011)
12. Karthikeyan, P., et al.: Detection of human stress using short-term ECG and HRV signals. J. Mech. Med. Biol. **13**(3), 1–29 (2013)
13. Singh, R., et al.: A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. Biomed. Signal Process. Control **8**, 740–754 (2013)
14. Giannakakis, G., et al.: Review on psychological stress detection using biosignals. IEEE Trans. Affect. Comput. **13**(1), 440–460 (2022)
15. Sharma, N., Gedeon, T.: Objective measures, sensors and computational techniques for stress recognition and classification: a survey. Comput. Methods Programs Biomed. **108**, 1287–1301 (2012)
16. SBC: Brazilian guidelines of hypertension. Brazilian society of cardiology. ABC Cardiol. **95**(1), 1–51 (2010)
17. Malik, M., et al.: Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task force of the European society of cardiology. Eur. Heart J. **17**, 354–381 (1996)
18. Bishop, C.: Input normalization and encoding. In: Neural Networks for Pattern Recognition. Oxford University Press, New York (1995)
19. Samarasinghe, S.: Neural Networks for Applied Sciences: From Fundamentals to Complex Pattern Recognition. Auerbach Publications, Boca Raton (2007)
20. Zaknich, A.: Neural Networks for Intelligent Signal Processing. World Scientific, New Jersey (2003)
21. Haykin, S.: Neural Networks: A Comprehensive Foundation. Macmillan College Publishing, New York (1994)
22. Han, J., Kamber, M., Pei, J.: Metrics for evaluating classifier performance. In: Data Mining: Concepts and Techniques, 3rd edn. Morgan Kaufmann Publishers, San Francisco (2012)

# An Exploratory Study on Powell Optimization Method for Block Matching Evaluation on Ultrasound Images

Maurício Devicentis, Carolina Benetti, and Anderson Gariel Santiago[(✉)]

Center for Engineering, Modelling and Applied Social Sciences,
Federal University of ABC, São Bernardo do Campo, Brazil
`gabriel.santiago@ufabc.edu.br`

**Abstract.** This paper presents the application of an elastography model based on block matching using ultrasound mathematical phantoms. Elastography can be defined as the visualization of differences in the biomechanical properties of healthy and diseased tissues. The differences are seen and measured from comparisons of the tissue examined in two states: equilibrium and after an applied perturbation. Therefore, an algorithm was developed for application in two ultrasound images, one with deformation and a non-deformed one. The proposed algorithm uses the Powell optimization method to evaluate the displacements between the two images. It also uses cubic-spline interpolation to evaluate sub-pixel displacements. As this research has an exploratory bias, only a part of the image was selected for a first analysis of the results. The results proved promising, showing opportunities for future optimizations and applications of the proposed model presenting root-mean-squared errors and absolute errors of order $10^{-3}$ and relative errors of 4%, but still need improvements on the computational time.

**Keywords:** Elastography · Ultrasound · Block Matching · Optimization

## 1 Introduction

Elastography is an image-based technique which employs cross-correlation algorithms and it is mainly applied to medical ultrasound exams, for example, in breast, thyroid and liver exams [4,5]. It aims to visualize the differences between healthy and injured tissues given a Region of Interest (RoI) in two different stages: before and after an external disturbance.

Doyley & Parker [2] describe the Elastography in three steps: (i) application of a disturbance to a tissue using a quasi-static, harmonic, or transient external mechanical excitation source; (ii) measurement of response in terms of displacements, deformations, or phase and amplitude of vibrations; (iii) estimation of the mechanical properties of the imaged tissues. Shiina et al. [6] describe the

Elastography as a technique based on the change in the elasticity of soft tissues, as in the detection of differences in mechanical properties between healthy tissues and solid tumors.

As mentioned before, one of the main steps for the elastographic study is to estimate the displacements and their gradients so a strain profile can be obtained, and Block Matching algorithms can be used to perform this task. Liu et al. [3] present this concept in a study that uses as input data, not the ultrasound images, but the Radio Frequency (RF) signal.

This paper presents an exploratory study based on the paper published by Liu et al. using images of ultrasound mathematical phantoms, since images are more commonly available than RF signals. The algorithm proposed aims to minimize the mean Sum of Squared Differences (SSD) using optimization algorithms and, in this paper, Powell Method was the algorithm of choice since it does not need the derivative of the objective function, performing successive one-dimensional searches in each dimension of the problem. In order to evaluate the inter-pixel values and improve the accuracy of the results, the reference non-deformed image was interpolated using cubic splines.

Since the interpolation is an expensive computational procedure, and the main goal of this paper is to evaluate and provide a better understanding of the application of the algorithm proposed by Liu et al. on ultrasound images, a portion of the considered images were chosen and compared to a Finite Element simulation of the same model. The metrics considered are the relative error map ($\varepsilon_\%$), the mean squared error ($\varepsilon_{RMS}$), the absolute error map ($\varepsilon_{ABS}$), and the computational time of the proposed algorithm, so it can be optimized in future studies.

## 2    Methodology

### 2.1    Block Matching

In order to perform the block matching, two ultrasound images are used: the non-deformed one, taken as a reference, and a deformed image, taken after the application of a pressure field. Since the procedure to generate an ultrasound image implies in signal degradation [7], a cubic spline interpolation was used in order to fill the inter pixel values and improve the searching algorithm.

The first step in the proposed method is to perform a cropping in both images $I_{ref}$ and $I$, resulting in sub-images $\tilde{I}_{ref}$ and $\tilde{I}$, the first with size $M \times M$ and the latter with size $N \times N$ with $M < N$. $\tilde{I}_{ref}$ is interpolated and along $\tilde{I}$ used as input for the algorithm.

The block matching model consists in minimizing the cost function in Eq. 1, where $\bar{x}$ and $\bar{y}$ are the Taylor expansion of $x$ and $y$ given by Eqs. 2 and 3 and the summation is evaluated over $\Psi = M - N + 1$ pixels. Considering a fixed coordinate (x, y) taken at the center of the deformed image, the variables are the rigid displacements u and v in X and Y directions respectively and their derivatives $u_x$, $u_y$, $v_x$, $v_y$. The optimization algorithm chosen to perform this

task is the Powell Method [1] which does not requires the derivative of the cost function, performing a line search across all the variables.

$$C = \sum_{k=1}^{N} \sum_{l=1}^{N} [\tilde{I}_{ref}(\bar{x}_k, \bar{y}_l) - \tilde{I}(x_k, y_l)]^2 \tag{1}$$

$$\bar{x} = x + u + xu_x + xv_x \tag{2}$$

$$\bar{y} = y + v + yv_y + yu_y \tag{3}$$

The proposed algorithm was implemented in Python 3.9 Anaconda distribution and *interp2d* and *fmin_powell* functions from the *Scipy* library were used.

The interpolation is performed only once due to the fact that it is a computationally expensive process, being responsible for almost 90% of the execution time. Thus, this algorithm also looks for an ideal window size looking for a less expensive process of cubic-spline interpolation.

### 2.2   Finite Elements and Mathematical Phantom

In order to provide a gold standard for the block matching model, a finite element model was implemented in Python 3.9. For this simulation, a mesh consisting of 9728 linear triangular elements and 4929 nodes was used to describe both the geometry and the field variables. The material properties are listed below:

– Background:
  • Elastic modulus: 200.0 kPa;
  • Poisson coefficient: 0.495;
– Lesion:
  • Elastic modulus: 1.80 MPa;
  • Poisson coefficient: 0.495;

An ultrasound B-mode mathematical phantom was evaluated using FieldII package and the deformed image was obtained by applying the Finite Element Model with the same parameters described.

The resulting image is a $400 \times 400$ pixel image with two circular hyperechoic lesions located at $(x, y) = (200, 70)$ and $(x, y) = (200, 330)$ with radius 15 pixels. The parameters used for the ultrasound simulation are listed below:

- Speed of sound: 1,540 m/s;
- Attenuation: 0.0 dB;
- Number of scatters: 1000 mm$^{-2}$;
- Background scatter amplitude: 1.0;
- Lesion scatter amplitude: 1.7 (hyperechoic);
- Central frequency: 3.5 MHz;
- Sampling frequency: 100 MHz;
- Phantom dimension: $100 \times 100$ (mm $\times$ mm);

For the present study, a constant pressure field $p = 2.0 kPa$ applied in the -Y direction resulting in a vertical strain of 0.1%. Figure 1a presents the physical model and its boundary conditions, and Fig. 1b presents the mathematical phantom with a red region used to evaluate the Block Matching model, consisting of the background plus lesion, allowing the verification of the model when a discontinuity of tissue is present.

## 2.3   Model Evaluation

In order to proceed with the analysis, $\tilde{I}_{ref}$ was chosen with a fixed squared window size of length $M = 120$ pixels, while the window size for $\tilde{I}$, $N = 61$ pixels, resulting in a total of $60 \times 60$ displacement map for $u_{BM}$ and $v_{BM}$, respectively, the displacements according X and Y evaluated using Block Matching (BM).

Equation 4 to 6 presents the chosen metrics used to evaluate the Block Matching algorithm, where $\beta$ represents the displacements $u$ and $v$ in X and Y axis respectively and $\Psi = M - N + 1$. In order to provide a coherent comparison, Finite Element (FEM) results were converted to pixels.

$$\varepsilon_{ABS} = |\beta_{BM} - \beta_{FEM}| \tag{4}$$

$$\varepsilon(\%) = 100 \frac{|\beta_{BM} - \beta_{FEM}|}{(\beta_{FEM} + 10^{-10})} \tag{5}$$

$$\varepsilon_{RMS} = \sqrt{\frac{1}{\Psi^2} \sum_{n=1}^{\Psi^2} (\beta_{BM}^n - \beta_{FEM}^n)^2} \tag{6}$$

## 3   Results

Considering a window of $N = 61$ pixels, Figs. 2a to 2d and Figs. 3a to 3d present the Finite Element simulation and Block Matching results for X and Y axis along their respective results for absolute error ($\varepsilon_{ABS}$) and relative error ($\varepsilon_\%$).

Table 1 presents, for $u_{BM}$ and $v_{BM}$, the Root Mean Squared Error ($\varepsilon_{RMS}$), Mean Absolute Error ($\bar{\varepsilon}_{ABS}$) and its Standard Deviation ($\sigma_{\bar{\varepsilon}_{ABS}}$) (all in $mm$), the Mean Relative Error ($\bar{\varepsilon}_\%$) and its Standard Deviation ($\sigma_{\bar{\varepsilon}_\%}$).

The approximate time spent for the model simulation was 27.2 minutes, with 24.0 minutes used to interpolate $\tilde{I}_{ref}$.

(a)



(b)
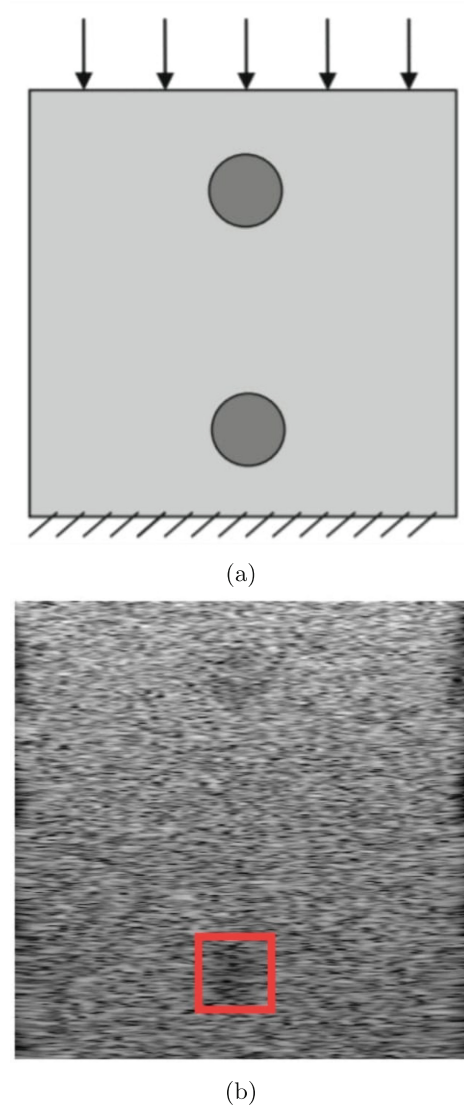
**Fig. 1.** a) Physical model; b) Mathematical phantom used.
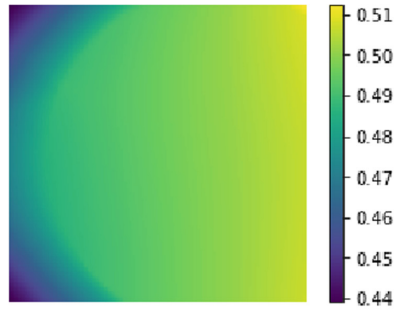
(a)
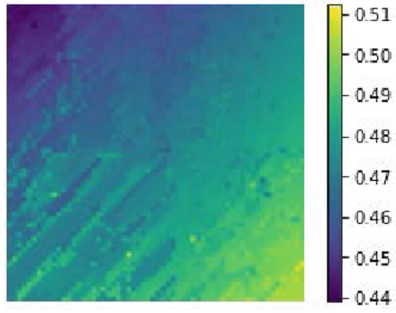


(b)
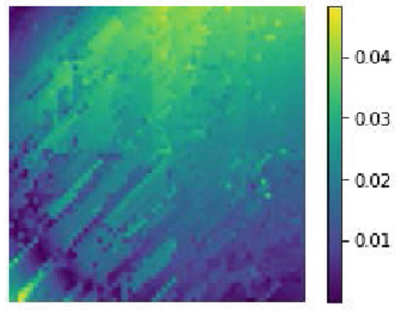


(c)
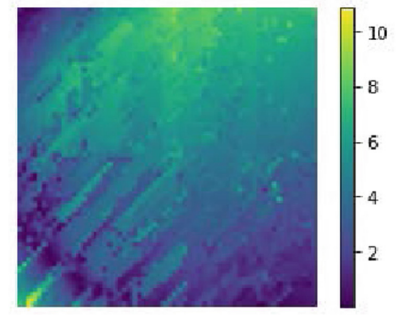


(d)

**Fig. 2.** Results for X displacements. a) Finite element displacement (in $mm$); b) Block matching displacement (in $mm$); c) $\varepsilon_{ABS}$ (in $mm$); d) $\varepsilon(\%)$.
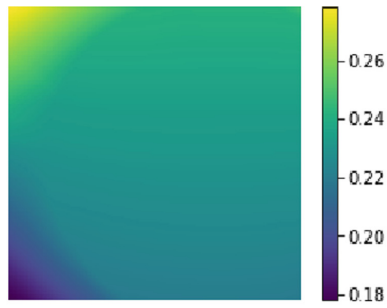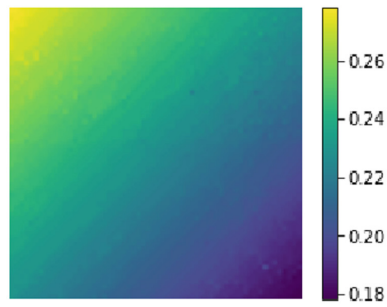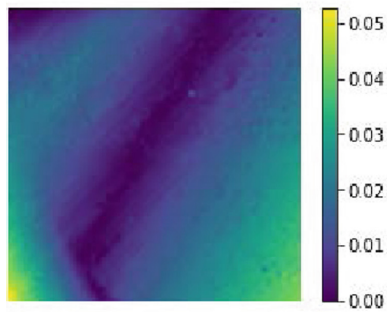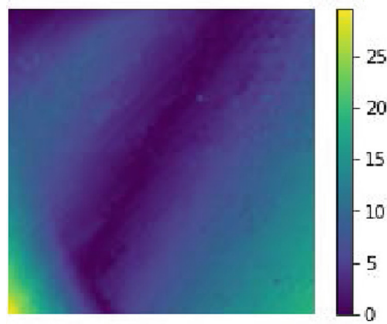
**Fig. 3.** Results for Y displacements. a) Finite element displacement (in $mm$); b) Block matching displacement (in $mm$); c) $\varepsilon_{ABS}$ (in $mm$); d) $\varepsilon(\%)$.

**Table 1.** Comparison between Block Matching and Finite Element simulation.

|          | $\varepsilon_{RMS}$ | $\bar{\varepsilon}_{ABS}$ | $\sigma_{\varepsilon_{ABS}}$ | $\bar{\varepsilon}_\%$ | $\sigma_{\varepsilon_\%}$ |
|----------|---------------------|---------------------------|------------------------------|------------------------|---------------------------|
| $u_{BM}$ | 1.912               | $20.790 \times 10^{-3}$   | $9.953 \times 10^{-3}$       | 4.214                  | 2.002                     |
| $v_{BM}$ | 1.097               | $14.540 \times 10^{-3}$   | $9.670 \times 10^{-3}$       | 6.426                  | 4.549                     |

## 4    Discussion

As quantitative results, Table 1 it is possible to note that the results obtained by Block Matching correspond to the same order of displacements as those evaluated using the Finite Element Method. The mean relative error is less than 10% while the absolute error is smaller than the $1mm$. On the other hand, the qualitative results presented in Fig. 2b and 3b do not allow the identification of the lesion. Observing the error maps in Figs. 2c, 3c, 2d, and 3d it is possible to note an increase in error values in the discontinuity between the two tissues. A possible solution for this issue is decrease the value of $M$ and $N$.

The presented results suggest that the applied model obtained satisfactory quantitative results concerning the error metrics used. Two major issues encountered during the experiments were the computational cost of the interpolation algorithm and the qualitative results, observed when comparing the Block Matching and Finite Element Method displacements maps. Also, the results presented by Liu et al [3] presented a smoother displacement map that can be justified by the fact that US images were used instead of RF signals, which are richer in information.

The next step in this research is to investigate and improve the qualitative results in order to allow the identification of the lesion in the ultrasound phantom and explore new ways to interpolate and evaluate inter-pixel displacements.

## 5    Conclusion

This paper presented an exploratory study of the implementation of a Block Matching model based on ultrasound images using Powell Method as an optimization algorithm. The results shows potential of the method, however, as presented in the Results, it is necessary to improve the computational cost of the interpolation algorithm and the qualitative results.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Arora, R.K.: Optimization: Algorithms and Applications. CRC Press (2015)

2. Doyley, M.M., Parker, K.J.: Elastography: general principles and clinical applications. Ultrasound Clin. **9**(1), 1–11 (2014). https://doi.org/10.1016/j.cult.2013.09.006. https://www.sciencedirect.com/science/article/pii/S1556858X13001060
3. Ke, L., Pengfei, Z., Shao, J., Xinjian, Z., Jing, B.: A 2D strain estimator with numerical optimization method for soft-tissue elastography. Ultrasonics **49**(8), 723–732 (2009). https://doi.org/10.1016/j.ultras.2009.05.004. https://www.sciencedirect.com/science/article/pii/S0041624X09000651
4. Szabo, T.L.: Diagnostic Ultrasound Imaging. Academic Press (2004)
5. Boctor, E.M., Choti, M.A., Hager, G.D.: Ultrasound elastography using multiple images. Med. Image Anal. **18**(2), 314–329 (2014). https://doi.org/10.1016/j.media.2013.11.002
6. Shiina, T., et al.: WFUMB guidelines and recommendations for clinical use of ultrasound Elastography: Part 1: basic principles and terminology. Ultrasound Med. Biol. **41**(5), 1126–1147 (2015). https://doi.org/10.1016/j.ultrasmedbio.2015.03.009. https://www.sciencedirect.com/science/article/pii/S0301562915002227
7. Schiefler, N.T., Maia, J.M., Schneider, F.K., Zimbico, A.J., Assef, A.A., Costa, T.: Generation and analysis of ultrasound images using plane wave and sparse arrays techniques. Sensors **18**(11), 3360 (2018)

# Evaluation of Classifiers for the Identification of Multiple Sclerosis Lesions in Neural MRI Scans with Attributes Extracted from Pre-trained Neural Networks

D. A. Vital[1]([✉]) [iD], M. S. Kudo[2] [iD], L. P. Marconatto[2] [iD], M. C. Moraes[2] [iD], and N. Abdala[1] [iD]

[1] Department of Diagnostic Imaging, Paulista School of Medicine, Federal University of São Paulo, UNIFESP, São Paulo, Brazil
`daniel.vital@unifesp.br`

[2] Laboratory of Image and Signal Processing, Institute of Science and Technology, Federal University of São Paulo, UNIFESP, São José dos Campos, Brazil

**Abstract.** About 2.8 million people worldwide have multiple sclerosis, and around 250,000 new cases are diagnosed annually. This disease damages the myelin sheath of neurons, injuring electrical signal conduction, and causing impairment and loss of senses, movement, and other neurological functions. Although the disease has no cure, early diagnosis is essential for the initiation of adequate treatment, controlling eventual outbreaks, delaying the advance, and improving the quality of life of patients. Neuroaxis magnetic resonance imaging is used to investigate this disease, identifying and following up lesions in the brain and spinal cord tissues; however, the diagnosis only through the visual evaluation of these exams may lack information for a quick and accurate analysis. Recent studies present computational methods based on artificial intelligence that allow the identification of lesions caused by the disease, aiming to overcome these visual limitations, but still have limitations in terms of accuracy and scope. The objective of this work was to evaluate the potential of classifiers based on machine learning algorithms in the identification of multiple sclerosis lesions in brain tissue. Axial FLAIR MRI brain exams were used from two databases combining filtering, normalization, and enhancement image pre-processing methods to extract the exam attributes. The algorithms evaluated were Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Machine and Logistic Regression, through training with attributes extracted from 3 pre-trained neural models: SqueezeNet, Inception V3 and VGG-19. The classifiers were evaluated by the classification of images from the studies as "with lesion" and "without lesion", and the SVM classifier trained with attributes extracted from the pre-trained neural model Inception V3 provided the best result, obtaining AUC = 0.988, accuracy = 0.980, sensitivity = 0.990 and F1-Score = 0.979.

**Keywords:** Multiple Sclerosis · Cerebral Magnetic Resonance · Computer Vision · Machine Learning · Deep Learning

# 1   Introduction

Multiple Sclerosis (MS) is a demyelinating disease of the central nervous system (CNS), being the major cause of nontraumatic neurological disability in young adults, that affects 2.8 million people worldwide, with around 250 thousand new cases annually [1]. MS is a chronic inflammatory disease of the CNS of presumed autoimmune etiology, in which inflammatory demyelination of axons causes damage to the myelin sheath of neurons and focal lesions in CNS White Matter (WM) [2, 3]. Although the cause of MS is still unknown due to the complexity of the disease, the studies indicate that many genes increase disease susceptibility, in addition to several environmental, immunological, and conditionals factors [3].

Magnetic Resonance Imaging (MRI) is the most used imaging technique to assist diagnosis and support of medical treatment of MS because of the high sensitivity to showing lesions in WM [2]. The sequence of MRI FLAIR (Fluid-Attenuated Inversion Recovery) is one of the most used to visualize these lesions, appearing hyperintense, discriminating lesions from the rest of the tissue [4]. The detection and evaluation of the evolution of MS lesions using MRI has been fundamental for clinical trials, used to detect the onset of the disease and track its progression [5]. However, the exam per se, without any processing, may lack information for accurate analysis and lesion identification, which may lead to subjective diagnosis and thus, poor pathological discrimination due to the high variation of lesions in terms of size, shape, intensity, and location.

The early identification of MS lesions allows the initiation of appropriate treatment, controlling eventual outbreaks, delaying the advance, and improving the quality of life of patients [5]. Recently, researchers tend to use computer vision with artificial intelligence to support radiologists in the diagnosis and segmentation of MS lesions with semi-automatic and automatic tools. Although these studies have shown results that add value to the clinical follow-up routine, they still have limitations, since MS diagnosis may be confused with other white matter diseases, there is a risk of clinical misinterpretation [4]. Despite being important ideas for monitoring the evolution of the disease, primary detection is the crucial factor in the treatment, and it supports professionals in selecting and reading possible urgent series of exams. Thus, the development of algorithms to automate the identification the MS lesions based on MRI data, before individual analysis and segmentation, would make a valuable contribution in this regard.

The state of the art presents different identification methods based on Convolutional Neural Networks (CNN) and machine learning classifiers with pre-processing methods of Data Augmentation, Histogram Stretching and Patch Extraction. In Zhang et al. [6], they compared classifier methods for stationary wavelet entropy-based multiple sclerosis detection with a decision tree, k-nearest neighbors, and support vector machine. In the work of Wang et al. [7], authors performed lesion identification in brain MRI exams, it was based on combining 14-layer CNN with batch normalization, dropout, and stochastic pooling. Already, in Zhang et al. [8] was used a 3D CNN with dropout and parametric ReLU for identification. Also, in the work by Siar and Teshnehlab [9] make use of CNN with Data Augmentation to diagnose and classify tumors and MS Simultaneously in brain MRI.

These works present refined solutions in which accuracy, precision and practicability have gradually evolved with each work, and may be useful in the diagnostic routine.

Although, some limitations such as lack of robustness with the evaluation in different equipment from distinct institutions and the high computational cost for improving accuracy could be found. The lack of identification of some MS lesions can be critical, especially the smaller ones in the onset of the disease, and in the emergence of new lesions, indicating the evolution and follow-up of the treatment. Thus, the objective of this work is to present an evaluation of different solutions for the identification of MS lesions in brain MRI scans from machine learning classifiers to overcome the limits of effectiveness and robustness of current methods.

## 2   Materials and Methods

The objective of this work is to present and evaluate the potential of machine learning classifiers in the identification of demyelinating lesions proved by multiple sclerosis in brain MRI scans.

The material used consists of FLAIR sequence brain MRI images obtained in the axial plane from two different databases described below. The training and evaluation algorithms of the classifiers were developed in Python programming language, using two computers. A notebook with an Intel® CORE i7 processor and 2.2 GHz microprocessor, 16 GB of RAM, and a 4 GB NVIDIA® GeForce GTX™ 1650 GPU. And a workstation with Intel® CORE i7 processor and 4.9 GHz microprocessor, 128 GB of RAM, and 8 GB Pci-express Geforce RTX 2070 Galax GPU.

**MICCAI 2016 Database**

The 2016 Medical Image Computing and Computer Assisted Intervention (MICCAI) challenge dataset includes brain MRI images of 53 MS patients [10]. The images in this dataset were obtained from two different centers in France with four different MR scanners [10]. This dataset contains exams from 53 MS patients, where the MR images were also manually segmented by seven specialists, with the respective consensus masks [10]. The exams were obtained in 3D FLAIR sequences and are available for research and education purposes.

**Department of Diagnostic Imaging (UNIFESP) Database**

The database referring to the Department of Diagnostic Imaging (DDI) of the Federal University of São Paulo (UNIFESP) refers to a set of retroactive brain MRI exams performed and made available by the institution. The dataset contains exams of 110 patients, aged between 25 and 60 years, diagnosed with MS, of which 23 patients had two MRI exams performed at two different points in time with intervals of 12 to 36 months, totaling 133 studies. The exams were obtained in two MRI scans of the institution. The use of the retroactive test set was approved by the ethics committee under protocol number 03830718.9.0000.5505, and the segmentation of MS lesions was performed manually by 3 specialists.

The methodology is divided into 4 steps: Pre-processing, Attribute Extraction, Classifier Training and Evaluation (Fig. 1). In the Pre-processing step (Fig. 1 in yellow), the original axial FLAIR MRI (IO) went through normalization processing, followed by filtering combining a Gaussian filter and edge sharpening to increase the discrimination

of MS lesions. Subsequently, in the Attribute Extraction step (Fig. 1 in blue) the normalized and filtered images (IF) are vectorized using three pre-trained deep learning neural networks, Inception V3 [11], SqueezeNet [12] and VGG-19 [13], obtaining vectors with characteristics that will be used in the next step. Then, in the Classifier Training step (Fig. 1 in green) the MRI FLAIR studies used from the available databases were divided into training, validation, and test sets, separating the attribute vectors obtained in the previous step, and were used in the training of 5 classifiers based on machine learning algorithms, Decision Tree [14], Random Forest [15], Knn (K-Nearest Neighbors) [16], SVM (Support Vector Machines) [17], Logistic Regression [18], with the vectors of each of the three pre-trained neural models used. Finally, in the Evaluation step, each of the classifiers developed is performed regarding the ability to classify the images of the brain FLAIR MR series regarding the presence or absence of MS lesions (Fig. 1 in gray).
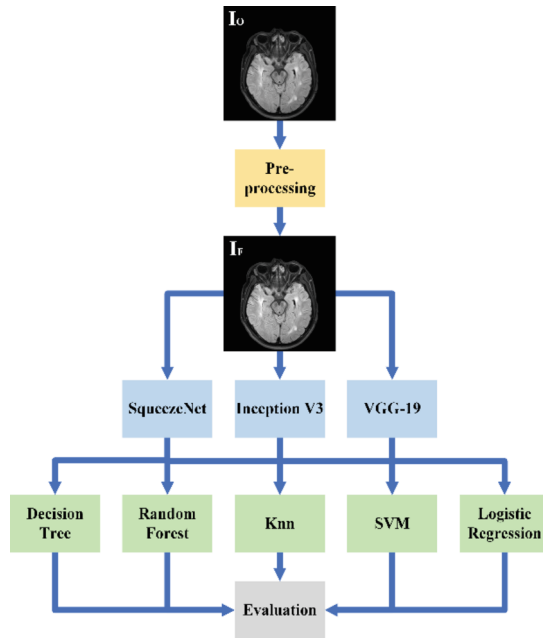


**Fig. 1.** Proposed Methodology for the training and evaluation of machine learning classifiers regarding the identification of MS lesions in axial FLAIR MRI exams.

*Pre-processing*
In the first step of the methodology, pre-processing operations were performed to increase the enhancement and discrimination of demyelinating lesions. The first operation of this stage was the normalization of the exams according to the histogram adjustment method by the Histogram Stretching (HS) method defined in Eq. 1 [7].

$$\varphi_{(x,y)} = \frac{\mu_{(x,y)} - \mu_{min}}{\mu_{max} - \mu_{min}} \tag{1}$$

where $\mu$ is the original image, $\varphi$ is the normalized image, (x,y) represents each pixel coordinate of the images, and $\mu_{min}$ and $\mu_{max}$ represent the minimum and maximum intensity values of the original brain image $\mu$. This method increases the contrast of images by stretching the range of intensity values from different sources into a single range [7]. This stage was carried out because the exams used are from two databases from different institutions, with images obtained from 6 different MRI scans. Adopting this normalization in all exams, they can be combined into a single dataset for the development of classifiers. Thus, each of the original axial FLAIR MRI ($I_O$) (Fig. 2 (a)) was passed by HS normalization (Fig. 2 (b)), obtaining images on a scale of 0 to 1 normalized ($I_N$) (Fig. 2 (c)).
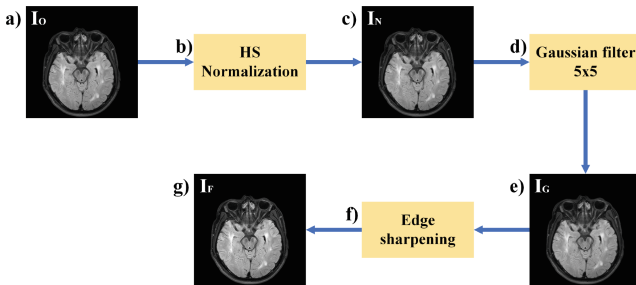


**Fig. 2.** Pre-processing methodology classification of exams for the presence of lesions. (a) Original Image ($I_O$). (b) Image normalization. (c) Normalized image ($I_N$). (d) $5 \times 5$ Gaussian Filter. (e) Filtered Image ($I_G$) (f) Edge sharpening. (g) Enhanced Image ($I_F$).

Subsequently, the $I_N$ were filtered using a Gaussian filter followed by an edge sharpening procedure. The Gaussian filter is a 2D convolution filtering method that is used to smooth images and remove high-frequency noise, which can make the image a little blurry [19]. And edge sharpening is a methodology that increases the prominence of the inner edges of the image by adding the original image to its discriminated edges [20]. In this way, the image obtained presents the initial characteristics, but with an enhancement of the edges of the objects. MRI can present speckle noise in their formation, and the filtering by the Gaussian filter is ideal to remove it. And when combined with edge sharpening, it enhances the tissue contour, resulting in sharper images and better visualization of MS lesions. Thus, the $I_N$ were filtered by a $5 \times 5$ Gaussian filter (Fig. 2 (d)), resulting in filtered images ($I_G$) (Fig. 2 (e)). And $I_F$ was passed by edge sharpening procedure (Fig. 2 (f)) resulting in the filtered and enhanced image ($I_F$) (Fig. 2 (g)) that will be used in the next steps.

*Attribute Extraction*
In the attribute extraction step, methods were designed to extract features from axial FLAIR MRI to classify the presence or absence of MS lesions for the development of classifiers in the next step. As the specific objective of this proposed project was to individually classify each of the images of the FLAIR MRI sequences into two classes, "with lesion" and "without lesion", it was identified the need to evaluate analytical models that would allow identifying the class belonging of each slice of the exams. And the method

selected for the extraction of these attributes was the use of pre-trained CNN (convolutional neural network) classification models to vectorize the pre-processed images, IF, from the previous step, obtaining feature vectors at the end of their convolutional layers, which will be used in the training of classifiers in the next step.

The use of pre-trained neural models aims to obtain, in a differentiated way, attributes obtained and identified in convolutional neural networks already proposed in the literature, and which have had efficient results in the area of action of the proposed objectives [21]. The use of this mechanism optimizes the feature extraction process, allowing the acquisition of a series of image descriptors, without the need to extract each one of them individually. In addition, these proposed method results in a reduction in computational cost, without the need to perform the training and optimization of filters and parameters of the convolutional layers, which have already been evaluated with thousands of images in the architectures presented by the literature [21]. In this context, once the feature vectors are already extracted, it is possible to focus on evaluating the best classifier to weight the weights of these attributes, similar to focusing specifically on the FCL (Fully Connected Layer) of deep learning models [21]. Thus, 3 pre-trained networks were used, Inception V3 [11], SqueezeNet [12] and VGG-19 [13] (Fig. 1 in blue), so that they could be compared regarding the potential for classifying the presence of EM in the FLAIR MRI exams. Inception V3 returns for each of the transformed images a vector with 2048 features. SqueezeNet returns a vector with 1000 features. And the VGG-19 network returns vectors with 4,096 descriptors for each image processed. The pre-trained models were imported from the Python Tensorflow computer language library version 2.7. These vectors with attributes will be the input of the machine learning algorithms developed in the next step, where they were evaluated for the potential of identifying the images in the 2 classes indicated.

*Classifier Training*

In the classifier training stage, 5 analytical classification methods based on machine learning algorithms were evaluated: Decision Tree [14], Random Forest [15], Knn [16], SVM [17] and Logistic Regression [18] (Fig. 1 in green). These classifiers were trained and evaluated from the feature vectors extracted from each of the 3 pre-trained neural networks obtained in the previous step. For the proposed work, the axial FLAIR MRI exams were divided into sets for training, validation, and testing, in the proportion of 80%, 10%, and 10%, respectively, with each set separated into 2 classes, "with lesion" and "without lesion". The exams were randomly separated between the three sets, respecting three previous rules:

- All images from the same FLAIR MRI series were placed in the same set;
- The 23 patients from DDI-UNIFESP database who have 2 MRI exams obtained at different moments in time, had both exams allocated to the same set;
- The proportion of exams between sets was followed in the distribution of exams by equipment.

Table 1 describes how many exams of each equipment were distributed in each of the sets. The number of images in each sequence of exams is different, and there is no homogeneity between the studies. But the proportion in each set was close to the division performed between the exams as shown in Table 2.

**Table 1.** Distribution of the number of FLAIR MRI exams in the training, validation and test sets, broken down by database and resonance scan.

| Database | Scan | Set | | |
|---|---|---|---|---|
| | | *Train* | Val | Test |
| MICCAI 2016 | GE Discovery 3T | 6 | 2 | 2 |
| | Philips Ingenia 3T | 12 | 2 | 1 |
| | Siemens Aera 1.5T | 12 | 1 | 2 |
| | Siemens Verio 3T | 13 | 1 | 2 |
| DDI-UNIFESP | Philips Achieva 3T | 69 | 9 | 9 |
| | Siemens Skyra 3T | 36 | 5 | 5 |

**Table 2.** Distribution of the number of axial FLAIR MRI images in the training, validation and test sets in the "with lesion" and "without lesion". classes.

| | Train | Val | Test |
|---|---|---|---|
| Images with lesions | 14,129 | 1,724 | 1,827 |
| Images without lesions | 16,170 | 1,959 | 2,074 |

The division of the axial FLAIR MRI exams and images was performed between the sets as described in Tables 1 and 2, using the respective feature vectors obtained in the previous step of each of the images. Training and validation data were used in the optimization of each of the 5 machine learning classifiers employed. The hyperparameters of the classifiers that resulted in higher performance were evaluated using a grid search with the training and validation data for each of the evaluated algorithms, from each of the three pre-trained neural networks from which the feature vectors were obtained.

*Evaluation*

The last step of the classification methodology of FLAIR MRI images regarding the presence of MS lesions was the evaluation of the classifieds trained in the previous step (Fig. 1 in gray). In this step, the 15 optimized classifiers developed were analyzed, 5 from each of the implemented machine learning algorithms, for each of the three pre-trained neural models from which the attribute vectors were extracted. The comparison metrics used were accuracy, area under the classification curve (AUC), sensibility, specificity, precision, and F1-score.

Accuracy is the most used metric in the evaluation of a classifier, it is defined as the ratio between the number of samples correctly classified and the total number of samples. And considering a binary classification problem ("with lesion" and "without lesion" as in the proposed objective), the evaluation also measured and analyze parameters such as: true positive (TP) that represents the number of samples of the positive class correctly classified; false negative (FN), or the number of samples of the positive class incorrectly classified; true negative (TN), which represents the number of samples of the negative

class correctly classified; and false positive (FP), or the number of samples of the negative class classified incorrectly. Sensibility (Eq. 2) is used to assess the number of positive class predictions that are actually positive. Specificity (Eq. 3) is used to assess the number of samples of the negative class that were correctly classified. Precision (Eq. 4) assesses the number of samples classified as positive that is actually positive. The F1-Score (Eq. 5) is a harmonic mean between precision and sensitivity. And the AUC is a metric related to the power of the classifier to separate two classes, related to the ROC curve (Receiver Operating Characteristics), which is a graph that allows the analysis of the sensitivity and specificity of a binary classification.

$$\text{Sensibility} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision} * \text{Sensibility}}{\text{Precision} + \text{Sensibility}} \tag{5}$$

## 3  Results and Discussions

As provided in the methodology, the present objective was the elaboration and evaluation of classifiers based on machine learning algorithms trained from the characteristics extracted from pre-trained CNNs. Three pre-trained neural models were evaluated: Inception V3, SqueezeNet and VGG-19, and 5 analytical models, decision tree, random forest, Knn, SVM, and logistic regression, for the classification of FLAIR MRI images in "with lesion" and "without lesion". Lesion". The classifiers were optimized through the variation of the specific hyperparameters in a grid with the training and validation sets according to Tables 1 and 2. Tables 3, 4, and 5 present respectively the AUC and Accuracy values of each of the 5 classification models trained with the characteristics extracted by the pre-trained Inception V3, SqueezeNet and VGG-19 networks, respectively. In the tables, the values of the metrics are presented both for the best performance of the validation data, as the numbers obtained by the optimized classifiers with the data of the test set.

The results presented in Tables 3, 4 and 5 show that the best classifiers for the characteristics extracted from the 3 pre-trained neural models were SVM and Logistic Regression (LR), presenting higher AUC and accuracy values. Table 6 presents the *sensibility*, specificity, precision and F1-score values obtained by these 6 best classifiers using data from the Test set.

Seeing Tables 3, 4, 5 and 6, the most sensitive classifier for the classification of axial FLAIR RM images was the SVM with attributes of Inception V3, which obtained accuracy of 0.980, sensibility of 0.990, and F1-Score of 0.979. And the most specific classifier was the Logistic Regression with Inception V3 attributes, obtained specificity

**Table 3.** AUC and accuracy results of the validation and test sets of the classifiers trained with the attribute from pre-trained Inception V3 model.

|                     | AUC   |       | Accuracy |       |
|---------------------|-------|-------|----------|-------|
|                     | Val   | Test  | Val      | Test  |
| Decision tree       | 0.880 | 0.863 | 0.921    | 0.903 |
| Random forest       | 0.972 | 0.963 | 0.953    | 0.944 |
| Knn                 | 0.936 | 0.921 | 0.936    | 0.928 |
| SVM                 | 0.997 | 0.988 | 0.986    | 0.980 |
| Logistic regression | 0.998 | 0.991 | 0.988    | 0.974 |

**Table 4.** AUC and accuracy results of the validation and test sets of the classifiers trained with the attribute from pre-trained SqueezeNet model.

|                     | AUC   |       | Accuracy |       |
|---------------------|-------|-------|----------|-------|
|                     | Val   | Test  | Val      | Test  |
| Decision tree       | 0.833 | 0.829 | 0.892    | 0.887 |
| Random forest       | 0.960 | 0.952 | 0.936    | 0.932 |
| Knn                 | 0.699 | 0.682 | 0.614    | 0.597 |
| SVM                 | 0.988 | 0.982 | 0.961    | 0.951 |
| Logistic regression | 0.993 | 0.984 | 0.973    | 0.962 |

**Table 5.** AUC and accuracy results of the validation and test sets of the classifiers trained with the attribute from pre-trained VGG-19 model.

|                     | AUC   |       | Accuracy |       |
|---------------------|-------|-------|----------|-------|
|                     | Val   | Test  | Val      | Test  |
| Decision tree       | 0.849 | 0.839 | 0.903    | 0.897 |
| Random forest       | 0.972 | 0.965 | 0.954    | 0.947 |
| Knn                 | 0.796 | 0.779 | 0.703    | 0.693 |
| SVM                 | 0.995 | 0.983 | 0.978    | 0.971 |
| Logistic regression | 0.993 | 0.985 | 0.977    | 0.965 |

values of 0.978 and precision of 0.975. These results demonstrate that the combination of these models, with the pre-processing step, resulted in classifiers with high precision and robustness for the classification of axial FLAIR MR images regarding the presence or absence of demyelinating lesions caused by EM. When compared to the best values in

**Table 6.** *Sensibility*, specificity, precision and F1-Score of the classification of the images from Test set regarding the presence of MS lesions from the SVM and Logistic regression classifiers.

|  |  | Sens | Spec | Prec | F1 |
|---|---|---|---|---|---|
| Inception V3 | SVM | **0.990** | 0.971 | 0.968 | **0,979** |
|  | LR | 0.969 | **0.978** | **0.975** | 0.972 |
| SqueezeNet | SVM | 0.969 | 0.946 | 0.940 | 0.954 |
|  | LR | 0.968 | 0.957 | 0.952 | 0.960 |
| VGG-19 | SVM | 0.981 | 0,962 | 0.958 | 0.969 |
|  | LR | 0.962 | 0.967 | 0.963 | 0.962 |

the literature, Siar and Teshnehlab [9] obtained accuracy and sensibility of 0.968, Zhang et al. [8] obtained 0.982 for the same metrics, and Wang et al. [7] obtained 98.7.

## 4  Conclusions

In Conclusion, the proposed approach presented significant values regarding accuracy and sensibility for an automatic Axial MRI brains MS lesion identification. The methodology of the proposed approach is based on the combination of attribute extraction by pre-trained neural models for the elaboration of the classifiers. As well as the outcome, this methodology presented that the SVM classifiers presented greater potential in the proposed classification, with the Inception V3 as the pre-trained neural model being the most efficient for the extraction of attributes. The classifier presented similar or superior results to the literature, with the advantage that it was developed with exams from two datasets from different institutions with six different MRI scans. Finally, for future works, we will seek new collaborators to increase our database to expand the analyses of the potential applications.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Walton, C., King, R., Rechtman, L., et al.: Rising prevalence of multiple sclerosis worldwide: insights from the Atlas of MS. Mult. Scler. J. **26**(14), 1816–1821 (2020)
2. Roy, S., Butman, J.A., Reich, D.S., et al.: Multiple sclerosis lesion segmentation from brain MRI via fully convolutional neural networks. arXiv preprint arXiv:1803.09172 (2018)
3. Dobson, R., Giovannoni, G.: Multiple sclerosis–a review. Eur. J. Neurol. **26**(1), 27–40 (2019)

4. Louapre, C.: Conventional and advanced MRI in multiple sclerosis. Revue Neurologique **174**(6), 391–397 (2018)
5. de Arruda, A.L.C., Vital, D.A., Kitamura, F.C., et al.: Multiple sclerosis segmentation method in magnetic resonance imaging using fuzzy connectedness, binarization, mathematical morphology, and 3D reconstruction. Res. Biomed. Eng. **36**(3), 291–301 (2020)
6. Zhang, Y., Lu, S., Zhou, X., et al.: Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and sup-port vector machine. Simulation **92**(9), 861–871 (2016)
7. Wang, S.H., Tang, C., Sun, J., et al.: Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. Front. Neurosci. **12**, 818 (2018)
8. Zhang, Y.D., Pan, C., Sun, J., et al.: Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. J. Comput. Sci. **28**, 1–10 (2018)
9. Siar, H., Teshnehlab, M.: Diagnosing and classification tumors and MS simultaneous of magnetic resonance images using convolution neural network. In: 2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), pp. 1–4. IEEE (2019)
10. Commowick, O., Istace, A., Kain, M., et al.: Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Sci. Rep. **8**(1), 1–17 (2018)
11. Szegedy, C., Vanhoucke, V., Ioffe, S., et al.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
12. Iandola, F.N., Han, S., Moskewicz, M.W., et al.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv preprint arXiv:1602.07360 (2016)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
14. Shalev-Shwartz, S., Ben-David, S.: Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, Cambridge (2014)
15. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
16. Peterson, L.E.: K-nearest neighbor. Scholarpedia **4**(2), 1883 (2009)
17. Gunn, S.R.: Support vector machines for classification and regression. ISIS Techn. Rep. **14**(1), 5–16 (1998)
18. Bishop, C.M., Nasrabadi, N.M.: Pattern Recognition and Machine Learning, vol. 4, no. 4, p. 738. Springer, New York (2006)
19. Cadena, L., Zotin, A., Cadena, F., et al.: Noise reduction techniques for processing of medical images. In: Proceedings of the World Congress on Engineering, vol. 1, pp. 5–9 (2017)
20. Ahmed, H.S., Nordin, M.J.: Improving diagnostic viewing of medical images using enhancement algorithms. J. Comput. Sci. **7**(12), 1831 (2011)
21. Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., et al.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016)

# Histopathological Analysis of Fine-Needle Aspiration Biopsies of Thyroid Nodules Using Explainable Convolutional Neural Networks

Matheus de Freitas Oliveira Baffa[1](✉) , Luciano Bachmann[2] ,
Thiago Martini Pereira[3] , Denise Maria Zezell[4] , Edson Garcia Soares[5] ,
Joel Del Bel Pádua[6] , and Joaquim Cezar Felipe[1]

[1] Department of Computing and Mathematics, Faculty of Philosophy, Science and
Letters of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, SP, Brazil
mbaffa@usp.br
[2] Department of Physics, Faculty of Philosophy, Sciences and Letters at Ribeirão
Preto, University of São Paulo, Ribeirão Preto, SP, Brazil
[3] Department of Science and Technology, Federal University of São Paulo,
São José dos Campos, SP, Brazil
[4] Nuclear and Energy Research Institute, São Paulo, SP, Brazil
[5] Department of Pathology, School of Medicine of Ribeirão Preto,
University of São Paulo, Ribeirão Preto, SP, Brazil
[6] Pathology Service, Clinical Hospital of Ribeirão Preto, University of São Paulo,
Ribeirão Preto, SP, Brazil

**Abstract.** Thyroid Cancer is a disease in which abnormal cells grow uncontrollably in the gland with the potential to invade other organs. Every year, almost 44,000 new cases are diagnosed worldwide. Histopathological diagnosis of fine-needle aspiration biopsies of thyroid nodules is the most precise exam to confirm the diagnosis and estimate the stages of the disease. The diagnostic process in such an exam involves detecting atypical signs, such as the presence of cell proliferation with irregular shape and texture. This task could be even harder once you consider that most thyroid biopsies might present multiple pathological states, such as inflammatory diseases and hyperplasia. Therefore, this paper addresses the development of a Computer Vision method to assist the histopathological diagnosis of normal, thyroid papillary carcinoma and goiter. The proposed method model and implement a Convolutional Neural Network to detect visual patterns to differentiate the three pathological states. Experiments following the Holdout Cross-Validation protocol reached an accuracy of 88.73% for the multiclass approach and 95.74% accuracy for the binary assessment. The results confirm the potential of the proposed method to assist pathologists in prescribing a more precise diagnosis.

**Keywords:** convolutional neural network · deep learning · histopathological analysis · thyroid cancer

## 1    Introduction

Thyroid Cancer is a type of malignant tumor that affects the thyroid gland causing difficulty in swallowing, a sore throat that does not get better, hoarseness, and swollen glands in the neck. This is a rare type of cancer that affects mainly people in their 30s, with the majority of cases registered in women, up to 3 times more frequently compared to men [1].

The process of diagnosis involves physical exams, imaging exams such as Ultrasound and CT scan, and biopsy. The fine-needle aspiration biopsy removes several tissue samples from the thyroid to better detect the type of cancer, the stage and to follow up with the treatment, to confirm whether the cancer was completely removed and the patient's general health [2]. A study conducted by Alves et al. [3] with 110 patients showed that among the tests for detecting cancers, the histological method is the best method for detecting thyroid tumors.

The tumors can be categorized as either being differentiated thyroid cancers, such as Papillary and Follicular Carcinoma (well-differentiated tumors), and Medullary thyroid cancer, which is a neuroendocrine tumor [2]. Another thyroid pathological condition involves inflammatory and autoimmune conditions such as Goiter, Thyroiditis, Graves' and Hashimoto's Disease [4]. The detection of the thyroid cancer type, as well as the inflammatory diseases, can be a very hard task that requires a very experienced and specialized pathologist. Hence there is a need for developing a method that does not require such a highly specialized professional and can assist the less experienced one.

Recent advances in Computer Vision have been proposed in the medical literature to assist pathologists to prescribe a better diagnosis. In 2017, for example, Araújo et al. [5] developed a computational method to assist in breast cancer diagnosis using histology images. Similarly, Siqueira et al. [6] proposed a method for automatic prostate cancer classification. Computer Vision methods are described as a three stages techniques. First, it is necessary to preprocess the data in order to highlight features and to set the region of interest for the analysis. Then, feature extraction methods are used to describe image content into a numerical vector. Finally, a Machine Learning algorithm is used to recognize patterns and to create a classification model for further analysis [7]. Such methods may assist doctors to detect abnormal signs related to different pathological conditions in a highly efficient way.

Therefore, the object of the present study was to develop a Computer Vision method to assist the pathologist in diagnosing Goiter, Thyroid Cancer e Normal tissues. A binary assessment is also provided to measure the efficiency of the proposed method to differentiate Normal and Tumor tissues. The proposed method is based on Convolutional Neural Networks (CNN), to extract patterns over a preprocessed sample of thyroid tissue and to create a classification model for further applications. The results obtained are the main contributions to the field, along with the interpretable results and the development of a custom CNN architecture for both multiclass and binary assessments of thyroid histopathological image classification.

The contents of this paper are developed as follows. First, related studies in the area are introduced to analyze how researchers have approached the thyroid classification problem. Then, the database composition and preprocessing are presented along with the classification methodology proposed. Finally, the experiments and results are discussed in detail.

## 2    Related Studies

Over the last years, the thyroid tissue classification problem has been addressed in the literature in many different ways. Most of them deal with binary classifications between healthy and tumor tissues, and the extraction of representative features for further pattern recognition stages. This section introduces some of the recent approaches available in the literature.

Guan et al. [8] developed a Computer Vision method based on Visual Geometric Group (VGG-16), a Very Deep Convolutional Neural Network architecture, to identify papillary thyroid carcinoma in cytological images. The authors used cytological-level images which present a small number of loose cells, without tissue organization. In total, the authors worked with 279 cytological images of thyroid nodules. The VGG-16 algorithm achieved an accuracy of 97.66% in differentiating benign nodules and papillary tumor cells.

Later, Halicek et al. [9] developed a Computer Vision method to assist pathologists in detecting head and neck tumors among whole-slide pathology images. In total, his image dataset had 381 samples acquired from 156 patients and was used to train a modified version of Inception-v4 Network, with around 141 convolutional layers. The authors reached an accuracy of 89.40%, a sensitivity of 89.60% and a specificity of 89.10% for thyroid cancer detection.

In 2020, Chen et al. [10] proposed a whole-slide analysis for thyroid cancer diagnosis but using a feature extraction approach. The authors used a dataset with 345 samples and extracted features using a Convolutional Neural Network for further evaluation using traditional Machine Learning methods, such as Decision Tree, Support Vector Machine (SVM), Random Forest, and Multilayer Perceptron (MLP). The combination of ResNet-50 features and an MLP reached the highest results, with an accuracy of 96.10%, a precision of 96.62%, and a recall of 98.40%.

Han et al. [11] proposed an automatic classification method for thyroid pathological images as well. The authors used a dataset of digitalized images of 16,500 thyroid samples, of which 7,928 are normal tissues and 8,572 are papillary thyroid carcinomas, collected on multiple regions of the sample from 55 patients. The samples are pre-selected using a similarity-based method and the most representative samples are used to train a VGG network. The authors report a precision of 95.70% and a sensitivity of 95.8%.

Duc et al. [12] proposed a Deep Learning ensemble method to automatically predict papillary thyroid cancer using fine-needle aspiration cytology images. The authors used a dataset with 367 images, of which 222 cases are papillary thyroid tumors and 145 are benign lesions. They evaluated several Very Deep

CNN models, such as ResNet-50, DenseNet, and Inception-v3, all pre-trained with ImageNet dataset. The DenseNet algorithm reached the highest results, with an average accuracy of 95.56%, a sensitivity of 97.34%, and a specificity of 94.05%.

Unlike the studies mentioned above, this work proposes a custom CNN architecture for thyroid histopathological classification, which reduces the network complexity and helps build a more reliable classification model. Also, we evaluate the thyroid tissue classification into a multiclass (normal, goiter, and cancer) and a binary (normal and cancer) approaches, providing a visualization of the inner patterns explaining how the CNN classified the samples.

## 3   Materials and Methods

According to the traditional Computer Vision approach for image classification, the proposed method can be described as a three-step process. First, a digitalized version of the sample is generated. After that, since our samples are not strongly stained, an image processing step is required in order to improve the image color distribution. Finally, a custom CNN is implemented in order to extract features, recognize patterns and create a classification model for further applications. An overview of the proposed method is shown in Fig. 1.
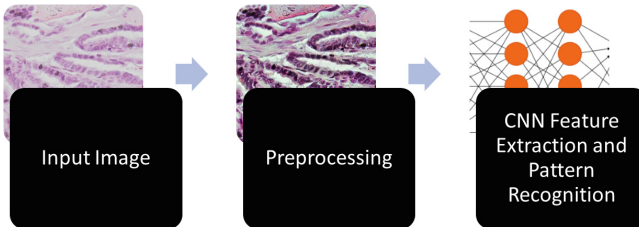


**Fig. 1.** Method Overview.

### 3.1   Data Acquisition and Preprocessing

To develop this work, the biological material was acquired from Biomax Inc., by researchers at the Brazilian Institute of Energetic and Nuclear Research (IPEN). Biomax Inc. focuses on sales, histochemical analysis, and other services using histological tissue. Three micro-array slides were acquired containing tissues of Normal, Goiter (Hyperplastic Goiter, Nodular Goiter, Hashimoto's disease, and Subacute Thyroiditis), and Papillary Thyroid Carcinoma, each one with 80 cases/cores, 240 in total. These micro-arrays are stained with Hematoxylin and Eosin (H&E) chemical compounds to enable further pathological studies.

After acquiring the micro-arrays, a digitalized image of each core is generated using a microscope Scan Scope Olympus BX61VS, under 40x magnification. The scan process considers a focus point on each core in order to obtain detailed and high-definition images.

All tissue slides are pre-labeled by two specialists from Biomax Inc. but in order to confirm the diagnosis, two pathologists specialized in thyroid tissue from the Pathology Service of the Clinical Hospital of Ribeirão Preto Medical School (FMRP-USP) analyzed the digital images to improve the diagnostic label before further feature extraction and pattern recognition stages.

Since each core may present more than one diagnosis, e.g., cores with normal cells and a small region of hyperplastic goiter, a small inner region of interest is manually selected in order to improve the representative of that core. This is an important step to ensure that the supervised Machine Learning method will find patterns over the correct labels and remove unwanted information, such as blood cells and colloids. Each inner region is selected with a window size of $1500 \times 950$ in pixel size. Only one region per core is selected to avoid multiple images per patient. The process of acquiring the image from the microscope and selecting the inner region were performed using the software QuPath - Quantitative Pathology & Bioimage Analysis v0.3.2. Figure 2 shows a digitalized core, the region of interest manually selected, and the resulting image.
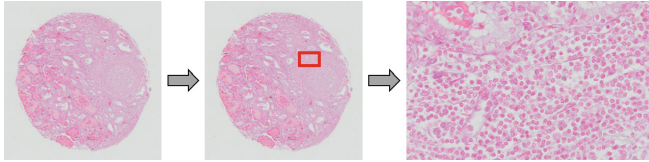


**Fig. 2.** Region of interest selection over the whole core.

After the acquisition process, some samples presented very poor staining, which turns the diagnosis a hard task. Therefore, to deal with such a problem, a preprocessing stage is proposed to improve the image quality and as well highlight the sample's color.

In this stage, the image is loaded using RGB color space. Then, each channel is split in order to be equalized and enhanced separately. The Contrast Limited Adaptive Histogram Equalization (CLAHE) technique is used to equalize each channel. This method is better than the traditional histogram equalization because we can manually limit the contrast intensity, enhancing the color distribution without adding noise or artifacts to the sample. Figure 3 contains an example of poorly stained samples and their respective preprocessed image.

Finally, to improve the model efficiency and generalization, a data augmentation step is utilized. In this stage, seven operations are applied for each image increasing the sample variability. The seven operations are: (i) rescale, (ii) shear, (iii) zoom, (iv) horizontal flip, (v) width shift, (vi) height shift, (vii) rotation.
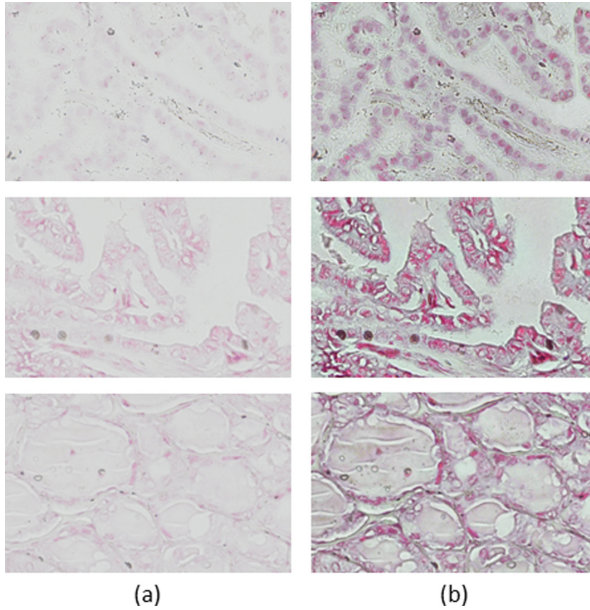
**Fig. 3.** H&E staining enhancement using Contrast Limited Adaptive Histogram Equalization. In (a) the original image and (b) the color enhanced version.

### 3.2   Classification Methodology

Convolutional Neural Network is a specialized type of Deep Neural Network for processing data in a grid-shape topology, such as images and videos. This type of network is capable of extracting feature maps through convolutional operations and developing a classification model based on such features [13]. CNN has successfully been applied to many Computer Vision problems which motivates its evaluation for thyroid histopathological image classification as well.

Many related studies use very deep architectures, such as Inception, ResNet, and VGG. The problem with such an approach is that they require a large amount of data for a better generalization and to avoid overfitting the model. On the other side, very short architectures, such as LeNet-5, aren't capable of finding enough patterns for the proposed problem. Therefore, a custom CNN architecture is developed in order to extract a relatively high amount of feature maps, without compromising the performance and the generalization.

The proposed CNN has 10 layers, of which one is the input layer, six are convolutional layers, followed by two dense layers and one output layer. The input layer receives the image with $64 \times 64$ in pixel size, with three color channels. The convolutional layers are organized in pairs interspersing the padding between "valid" and "same" options. They extract 16, 32, and 64 filters, respectively, with the size of $3 \times 3$. For every two convolutional layers, a MaxPooling operation is used to reduce the sample space by $2 \times 2$ window size. After the

convolutional layers, a flattening operation is used to reshape the data for the pattern recognition task using the dense layers. Each dense layer has 512 neurons and is followed by a dropout operation, which randomly turns off 20% of the neurons in the dense layer. Both convolutional and dense layers use the Rectifier Linear Unit (ReLU) activation function. Finally, the output layer has three neurons, representing the three classes dealt in this work (goiter, cancer and normal), and uses the Softmax activation function. The other parameters used in the neural network are set as follows. The Adam algorithm was used to optimize the network, along with the Categorical Cross-entropy as the loss function. We used a batch size of 64 and ran the algorithm for 1000 epochs. Figure 4 summarizes the proposed CNN custom architecture.
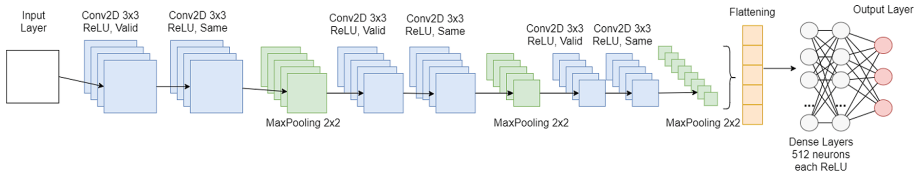


**Fig. 4.** Custom CNN architecture for thyroid histopathological classification.

An assessment for binary classification, between normal and cancer tissue is also provided. The same architecture is utilized except for some parameters adjusted for binary classification, such as the output activation function, which was the Sigmoid function, the number of neurons in the output layer was one, and the loss function was set to binary cross-entropy.

## 4  Experiments and Results

To develop this work, we used a Linux Workstation, with an Intel Xeon Silver CPU, 64 GB of RAM, and two NVIDIA RTX A4000 GPU. We used Python programming language to code the scripts, along with Tensorflow 2.7.0, OpenCV 4.5.5, and Scikit-Learn 1.1.1.

The experiment followed the Holdout Cross-Validation Protocol. The dataset was randomly split into two parts, in which 70% of the images were used to train the CNN and 30% to test. No data or patients were repeated in between the subsets.

This experiment was also validated quantitatively using four evaluation metrics: Accuracy, Precision, Recall (Sensitivity) and Area Under the ROC Curve (AUC). These metrics are calculated accordingly to the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) classifications. For the multiclass approach, we considered these metrics using the micro-average strategy, along with the One-Versus-Rest (OVR) for the AUC calculation.

The multiclass assessment, in which we classified normal, goiter, and cancer thyroid tissues, the proposed CNN architecture achieved a validation accuracy of 88.73%. This is a relatively high accuracy considering the complexity of finding enough visual patterns to differentiate goiter and normal tissue. These two pathological states have very similar structures which make the diagnosis difficult. Figure 5 displays the training and validation curve for the accuracy value over the epochs. The shape of the curve is very characteristic of a good fit Deep Learning model, which demonstrates that the model doesn't overfitted and that the testing data was large enough.
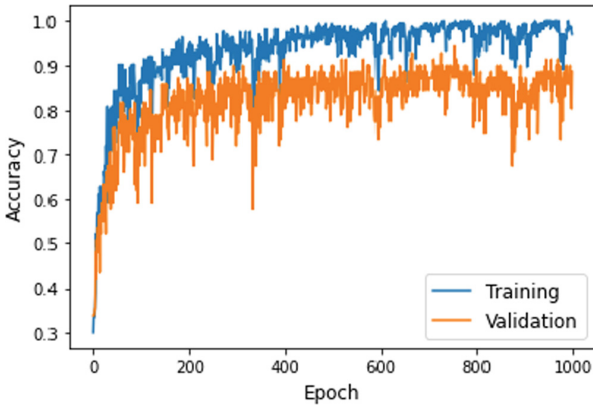


**Fig. 5.** Training and validation curve for accuracy over the epochs.

The recall, often referred to as sensitivity, achieved 88.73% which is also a very satisfactory and important result for medical imaging analysis, once it measures the proportion of actual positives was correctly identified among the classes. In Fig. 6 we display the training and validation curve for the recall values over the epochs. This curve is also well fit for the proposed problem. The method also achieved a precision of 88.73% and an AUC of 94.41% related to the overall efficiency of the method.

The structure of the CNN was developed empirically in order to maximize the accuracy and the recall. As mentioned before, we tested simpler architectures, such as LeNet-5, and very deep architectures, such as VGG, Inception, Xception, ResNet, and EfficientNet. The smaller architectures didn't have enough depth to create complex patterns capable of differentiating the three proposed classes, while the deep architectures didn't have enough data to generalize the model. Also, because these deep architectures are pre-trained on the ImageNet dataset, we were not able to modify it to add structures to avoid overfitting such as regularization and dropout.

The image preprocessing was also important to improve the overall efficiency of the method. We tested the original images, among with the poorly stained
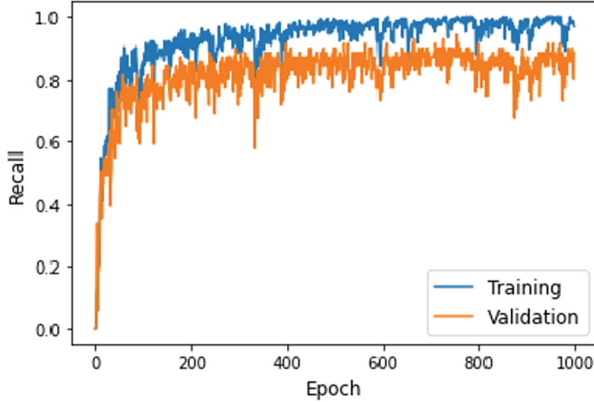
**Fig. 6.** Training and validation curve for recall over the epochs.

cores, and their respective grayscale images, but the color in the custom CNN showed as an extremely important feature to differentiate the three classes.

By analyzing our results, we observed the complexity to differentiate normal and goiter tissues. Inflammatory tissue may be very similar to normal tissue, but with subtle changes, e.g., the hyperplastic goiter, which appears as a dense cluster of nuclei. Since most of the thyroid literature doesn't cover inflammatory diseases in a multiclass approach, a second experiment was proposed with only cancer and normal tissues to measure the impact of this difficulty.

The binary assessment, without the goiter images, used the same CNN custom architecture but adjusted the multiclass parameters to binary. Compared to the multiclass performance, as expected, the binary assessment had a slightly higher result, achieving an accuracy of 95.74%, with a precision of 100.0%, a recall of 91.30%, and an AUC of 97.55%. This second experiment shows an increase in accuracy of 7% when the goiter class isn't considered. This illustrates the difficulty in separating the three classes indicating that there's a need for further investigations to improve their representation. Table 1 summarizes the multiclass and the binary assessments using the processed dataset and the custom CNN.

**Table 1.** Quantitative Results Summary.

|            | Accuracy | Precision | Recall | AUC |
|------------|----------|-----------|--------|-----|
| Multiclass | 88.73%   | 88.73%    | 88.73% | 94.41% |
| Binary     | 95.74%   | 100.0%    | 91.30% | 97.55% |

Using the Grad-CAM algorithm [14], we were able to plot heatmaps illustrating the pattern found in the CNN. Our goal was to check how the multiclass CNN

was interpreting the image and to visualize the normal and cancer patterns in a human-understanding way. To do so, Grad-CAM utilizes feature maps (filters) found in the very last convolutional layer and the classification model to infer the image class and point out as heat regions, the location the patterns were found. Figure 7 shows four sequences of images, each containing the heatmap illustrating the pattern detected on normal and cancer samples, its respective H&E sample, and the overlap between the heatmap and the sample. In normal samples, the heated area tends to be in a well-shaped region, clustered together while the cancer tissue presents in a scattered form, without a specific structure.
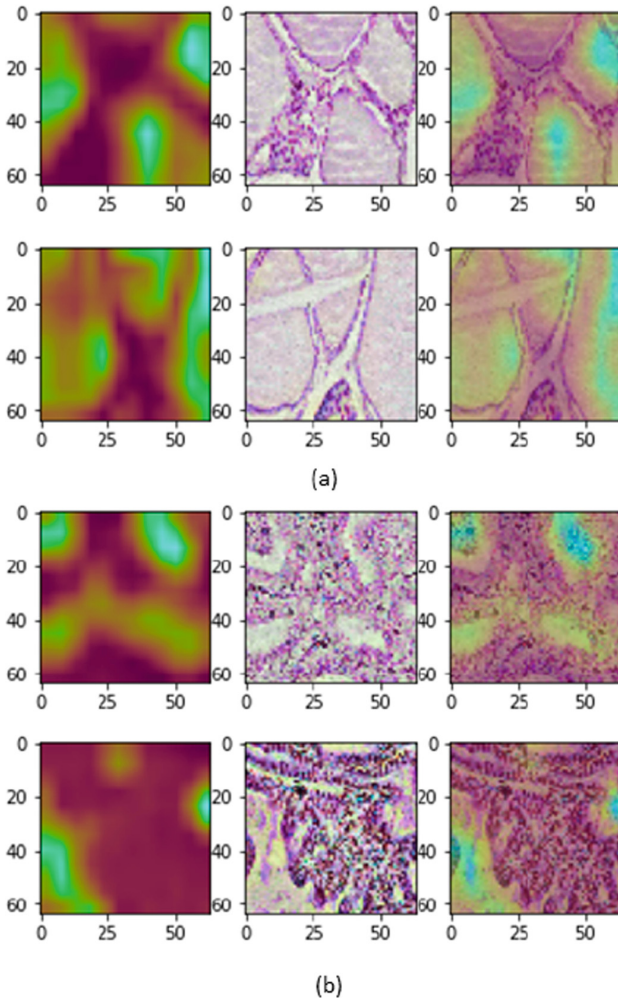


**Fig. 7.** CNN patterns found over four different thyroid histological samples. In (a) the pattern over normal thyroid tissue and (b) the pattern over thyroid cancer tissue.

# 5    Conclusion

Thyroid Cancer is the uncontrollable growth of abnormal cells in the gland, capable of evading other tissues. Fine-needle aspiration biopsies of thyroid nodules are the best screening technique to confirm the diagnosis and determine its group type and the stage of the disease. In this work, we proposed a Computer Vision-based method to assist pathologists with thyroid histopathological analysis in order to improve the diagnosis precision and efficiency. The proposed method based on a custom CNN architecture achieved an excellent result for such a complex task in differentiating the three classes. For future works, we intend to investigate new possibilities to represent the data in order to improve the differentiability between the normal and goiter classes.

# References

1. National Health Service: Thyroid cancer (2019). https://www.nhs.uk/conditions/thyroid-cancer/
2. National Cancer Institute: PDQ thyroid cancer treatment (adult) (2022). https://www.cancer.gov/types/thyroid/hp/thyroid-treatment-pdq
3. Alves, M.L.D., et al.: Valor preditivo do exame clínico, cintilografia, ultra-sonografia, citologia aspirativa e tiroglobulina sérica no nódulo tiroideano único atóxico: estudo prospectivo de 110 pacientes tratados cirurgicamente. [prediction of malignancy in the solitary thyroid nodule by physical examination, thyroid scintigraphy, thyroid ultrasound, fine-needle biopsy and serum thyroglobulin: a prospective study of 110 operated patients. in portuguese]. Arquivos Brasileiros de Endocrinologia & Metabologia **46**, 648–653 (2002)
4. Caturegli, P., Kimura, H., Rocchi, R., Rose, N.R.: Autoimmune thyroid diseases. Curr. Opin. Rheumatol. **19**(1), 44–48 (2007)
5. Araújo, T., et al.: Classification of breast cancer histology images using convolutional neural networks. PLoS ONE **12**(6), e0177544 (2017)
6. Siqueira, L.F.S., Araújo-Júnior, R.F., Araújo, A.A., Morais, C.L.M., Lima, K.M.G.: LDA vs. QDA for FT-MIR prostate cancer tissue classification. Chemometr. Intell. Lab. Syst. **162**, 123–129 (2017)
7. Conci, A., Azevedo, E., Leta, F.: Computação Gráfica: Teoria e Prática. [Computer Graphics: Theory and Practice. in Portuguese]. Volume 2. Elsevier Editora (2008)
8. Guan, Q., et al.: Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. J. Cancer **10**(20), 4876 (2019)
9. Halicek, M., et al.: Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks. Sci. Rep. **9**(1), 1–11 (2019)
10. Chen, P., Shi, X., Liang, Y., Li, Y., Yang, L., Gader, P.D.: Interactive thyroid whole slide image diagnostic system using deep representation. Comput. Methods Programs Biomed. **195**, 105630 (2020)

11. Han, B., Zhang, M., Gao, X., Wang, Z., You, F., Li, H.: Automatic classification method of thyroid pathological images using multiple magnification factors. Neurocomputing **460**, 231–242 (2021)
12. Duc, N.T., Lee, Y., Park, J.H., Lee, B.: An ensemble deep learning for automatic prediction of papillary thyroid carcinoma using fine needle aspiration cytology. Expert Syst. Appl. **188**, 115927 (2022)
13. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016). http://www.deeplearningbook.org
14. Ramprasaath, R.S., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)

# ECG Signal Extraction from Intensive Care Unit Monitor Videos

I. G. M. e Silva[1] , R. C. Coelho[1] , I. A. Zibordi Jr[2] , S. S. Camargo[1] ,
and C. M. G. de Godoy[1(✉)]

[1] Science and Technology Institute, Federal University of São Paulo, São José dos Campos,
SP, Brazil
gurjao.godoy@unifesp.br
[2] São Francisco Hospital and Maternity, Mogi Guaçu, SP, Brazil

**Abstract.** Computer Vision (CV) application benefits the health area, notably in its applications for assistive technologies and objective and in-depth analysis of biomedical images. However, there are currently no CV resources that innovate by collecting patients' vital signs directly from the ICU (Intensive Care Unit) medical equipment panel. Thus, the present work goal was to extract the electrocardiogram (ECG) signal from ICU monitors. The approach consisted of transforming ECG monitor signals into one-dimensional digital signals by segmenting them into frames, then extracting the segmentation's upper contour. We used nine heart monitor screen recordings (videos) available on YouTube as a database. The segmentation results validate using the Dice coefficient. Two frames of every recording were validated, generating 18 validations and an average Dice of 90.02 $\pm$ 5.74. We concluded that the approach proposed can extract ECG images from videos of Intensive Care Unit monitors and transform them into a signal in the time domain. It can help future ECG assessments, via computation vision, regarding the changes in heart rhythm (arrhythmias). It can also help circumvent limitations to access the ECG in Intensive Care Units by using, for example, a simple video camera, such as those of cell phones, close to the monitor. Such an innovative approach, in turn, would allow obtaining and transmitting the signals to the computer that will be responsible for its analysis.

**Keywords:** ECG Signal · Segmentation · Intensive Care Unit Monitor

## 1 Introduction

Computer vision (CV) is a technology that seeks to emulate human vision to automate image analysis, obtain high precision in this process and save time [1]. CV application benefits the health area, notably in its applications for assistive technologies and objective and in-depth analysis of biomedical images [2]. CV is also applicable to intensive care units [3], Homecare, and telemedicine [4], avoiding using storage devices, such as hard disks or memory cards.

Some of these demands may involve programming with a high level of complexity and high technology for the use of CV, which makes its implementation in the daily

routine of a hospital or clinic unfeasible. Developing systems that demand processing in personal computers is highly indicated for these cases.

Currently, no CV resources innovate by allowing, for example, the visual inspection, reading, and interpretation of the patient's vital signs from the medical equipment panel itself. That would help, for example, collect information from the electrocardiogram (ECG), the pressure curve, and the oxygen saturation value of patients on Intensive Care Unit (ICU) monitors.

Concerning precisely the ECG signal, the first challenge in seeking its interpretation is to be able to extract it correctly from the ICU monitor screen. That is precisely the objective of this work. Specifically, as a first step in using CV for ECG analysis in medical equipment panels, this work intends to develop a computational tool based on the segmentation technique that extracts the ECG image and its amplitude and time scale from videos recorded from ICU monitor screens.

## 2 Methods

The general goal is to extract ECG images from ICU monitors to be turned into a signal (over time). The idea is that the turned ECG could be understood and treated as the ECG signal itself.

The database used contains nine heart monitor screen recordings available on YouTube. The following criteria were applied to choose the videos with the help of a physician:

1) The ECG signal should be shown in green color (ICU's monitors normally use this color to show the ECG signal);
2) The ECG signal should not contain grids, that is, vertical and horizontal reference marks;
3) The camera used for filming should remain practically static or move little during recording;
4) The camera used for filming should be close to the monitor to allow clear visualization of the ECG signal, but at least 1 m away to avoid electromagnetic interference [5].

Figure 1 presents some examples of videos rejected for this project as they did not meet at least one of those criteria.

After choosing the videos to compose the database, the frame processing to extract the ECG signal started. All processing performs in MATLAB version 2018a. The first step consisted of transforming each video into a set of images to separate their frames. In the first frame of each video, a window containing only the ECG signal is selected manually, as shown in Fig. 2.

The selected window was applied to cut all video frames, as the processing is applied considering only the select region in all the videos. Therefore, just the ECG signal is segmented. The processing consisted of segmenting the image recorded on the ECG monitor to transform it into an ECG signal (time-domain). As for considering the RGB system color, all frames were split into R (red), G (green), and B (blue) components, and just the G component was used in the processing, as the ECG depicts in green color in all videos used. Thus, the green component pixels were selected in the range

100–255, generating a binary mask, where pixels in this range are white, and the others are black. The next step was to fill small holes that remained at the segmentation using a morphological closing operation with a circular structuring element of radius 1. The result of this process is approximately the ECG signal image segmented. To obtain the ECG signal amplitude, we extracted the upper contour of the segmentation. All columns of the segmented image were scanned from top to bottom, and the first white pixel found was selected. The result is the ECG signal, as exemplified in Fig. 3.
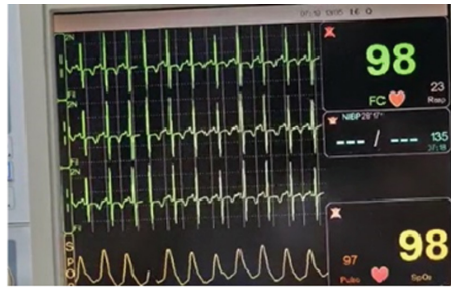


(a)



(b)



(c)

**Fig. 1.** Examples of videos rejected for the study. (a) The screen contains grids; (b) and (c) the monitor is far from the camera.

The ECG amplitude scale (1 mV) stands directly in a vertical bar on the screen, and the ECG horizontal length conventionally represents 6 s (Fig. 4). Accordingly, the number of pixels of those elements determines the ECG scale in millivolts and seconds.
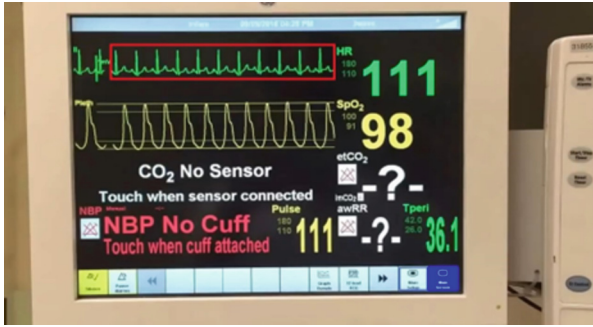
**Fig. 2.** Example of window selection (red outline) containing only the ECG signal at the first frame of each video. (Color figure online)
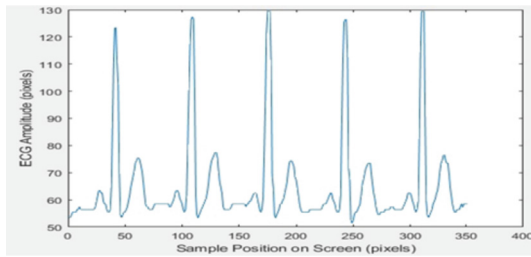


**Fig. 3.** ECG signal extracted from the segmented image.



**Fig. 4.** Elements in the ECG image used to determine the amplitude and time scale of the signal.

The validation considered the ECG signal segmentation results. For this, we use the Dice coefficient, which compares them to Gold Standards by Eq. 1.

$$Dice(A, B) = 2 * \frac{|A \cap B|}{|A| + |B|} \qquad (1)$$

The dice express the similarity between two regions (A and B). It varies from 0 to 1, with values close to 1 indicating the highest similarity between the regions. This work presented the Dice results on a scale from 0 to 100. Two Gold Standards were generated for each video, referring to the first and last frames. Thus, 18 images were validated. The Gold Standards were generated using the Gimp software by manually contouring the ECG signal by an ECG specialist, as shown in Fig. 5.
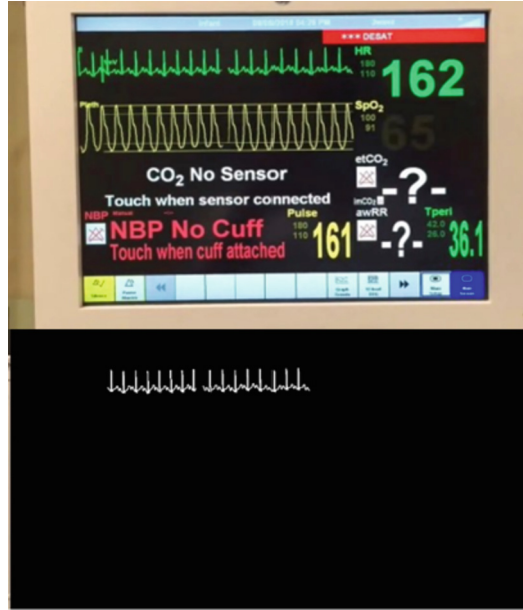
**Fig. 5.** Example of frame and Gold Standard manually generated for segmentation validation.

## 3 Results and Discussion

In this work, the approach chosen to transform ECG signals displayed on ICU monitors into one-dimensional digital signals consisted of segmenting them into frames and extracting the segmentation's upper contour. Such an approach helps provide a remote ECG assessment via computer vision.

After segmenting the ECG signals, the results were validated using the Dice coefficient. The database contained nine videos, and two frames of each recording were validated, generating 18 validations and an average Dice of $90.02 \pm 5.74$. The Dice results appear in the boxplot of Fig. 6.

The average Dice coefficient of 90.02 reveals high similarity between the segmented region and Gold Standard and good performance of the proposed method in segmenting ECG signals depicted on cardiac monitor screens. The minimum Dice obtained was 76.27, but the segmentation result was excellent, as shown in Fig. 7.

As shown in Fig. 7, the segmented region was thinner than the Gold Standard, which impacted the Dice result. However, it is possible to verify that the ECG segmentation was correct. As the Gold Standards were drawn manually over the frames, they may be slightly different from the signals shown on the monitor, which introduces an error in the quantitative validation. That is a physical limitation of the professional who carried out the tracing; thus, it cannot be circumvented. Visual inspection confirms that the segmentation was correct despite of a smaller Dice.

After finishing the segmentations, the one-dimensional digital signal was generated by extracting the upper contour of the segmented ECG. That was mirrored on the horizontal axis. The results were plotted on the original frame for visual inspection,

as exemplified in Fig. 8. It is possible to verify that the proposed method accurately extracts the ECG signal from the frames. Therefore, that result indicates that computer vision can apply to transforming ECG signals displayed on the ICU monitor screen into one-dimensional digital signals (time-domain).
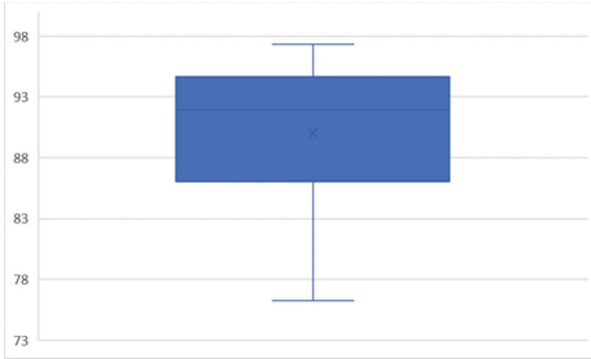


**Fig. 6.** Boxplot Dice coefficients result from the validation of the ECG signal segmentation.



**Fig. 7.** Gold Standard (white) and segmented signal (red). Scale not shown. (Color figure online)

The ECG temporal signals resulted from digital signal processing as the signal amplitude could convert from pixels to millivolts and the time scale convert from pixels to seconds. Additionally, the validation of the ECG signal relied on the Dice test and visual inspection.

One limitation of the method proposed in this work is that the segmentation of the ECG signal is impaired if there are other elements beyond the ECG in the selected window, such as grids, signal markings, and written warnings, as shown in Fig. 9. Thus, the window chosen must contain only the ECG signal. The green tones of the ECG signals can vary greatly depending on various factors, such as the heart monitor settings, quality of the device used to perform the recording, reflections on the monitor screen, and the ambient illumination where the monitor poses. Thus, to ensure adequate ECG signal segmentation from different videos, the method proposed in this work selects a wide range of colors in the G component (green). Due to that feature of the method, other elements in the window that are not green can be selected incorrectly, interfering with the final result.
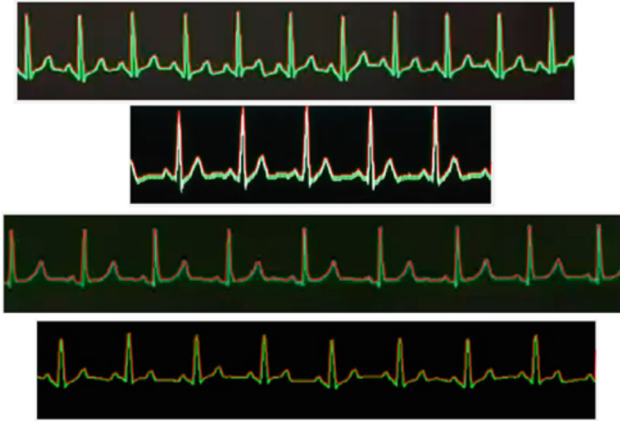
**Fig. 8.** ECG signals from cardiac monitors (green) and one-dimensional digital signals resulting from processing (red). Scale not shown. (Color figure online)



**Fig. 9.** Elements (green or not) present in the selected window (highlighted in red) that impair the ECG signal's segmentation. (Color figure online)

Even though the tool still needs improvements, it may help ECG monitoring in ICUs as patients' signals can be visually monitored and sent as an ECG signal to a single computer to be processed. Thus, if an intercurrence detects in a patient, help can be much faster, as the program would warn out any possible abnormalities.

Monitoring and assessing ICU patients' vital signals, such as ECG, is crucial as it can provide continuous and noninvasive follow-up [6, 7]. As one would expect, many authors have proposed diversified approaches for it [8–10].

However, to the best of our knowledge, using ICU monitor videos as a direct source of ECG signal is first described in the present work.

Other authors proposed extracting ECG signals from images [11, 12]. However, they proposed extracting ECG signals from ECG recordings' images on paper. Therefore, that approach is not applicable in ICU monitor videos, as proposed in the present work.

## 4   Conclusions

The present tool can extract ECG from videos of Intensive Care Unit (ICU) monitors and transform it into a time-domain ECG signal. It can help future ECG assessments, via computation vision, regarding pathological changes in the heart rhythm (arrhythmias).

It can also help circumvent limitations to access the ECG in ICU by using, for example, a simple video camera close to the monitor, such as those of cell phones. That approach, in turn, would allow obtaining and transmitting the signals to the computer responsible for ECG analysis.

**Conflict of Interest.**   The authors declare that they have no conflict of interest.

## References

1. Ramcharan, A., et al.: Assessing a mobile-based deep learning model for plant disease surveillance. arXiv, abs/1805.08692 (2018). https://doi.org/10.48550/arXiv.1805.08692
2. Esteva, A., et al.: Deep learning-enabled medical computer vision. npj Digit. Med. **4**(5) (2021). https://doi.org/10.1038/s41746-020-00376-2
3. Glancova, A., et al.: Are we ready for video recognition and computer vision in the intensive care unit? Surv. Appl. Clin. Inform. **12**(01), 120–132 (2021)
4. Leo, M., Medioni, G., Trivedi, M., Kanade, T., Farinella, G.M.: Computer vision for assistive technologies. Comput. Vis. Image Underst. **154**, 1–15 (2017)
5. Hans, N., Kapadia, F.N.: Effects of mobile phone use on specific intensive care unit devices. Indian J. Crit. Care Med. **12**(4), 170–173 (2008)
6. Khanna, A.K., Hoppe, P., Saugel, B.: Automated continuous noninvasive ward monitoring: future directions and challenges. Crit. Care **23**(194) (2019). https://doi.org/10.1186/s13054-019-2485-7
7. Gardner, R.M., Shabot, M.M.: Patient-monitoring systems. In: Shortliffe, E.H., Perreault, L.E. (eds.) Medical Informatics. Health Informatics, pp. 443–484. Springer, New York (2001). https://doi.org/10.1007/978-0-387-21721-5_13
8. Moorman, J.R.: The principles of whole-hospital predictive analytics monitoring for clinical medicine originated in the neonatal ICU. npj Digit. Med. **5**(41) (2022). https://doi.org/10.1038/s41746-022-00584-y
9. Davoudi, A., et al.: Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. Sci. Rep. **9**(1), 8020 (2019). https://doi.org/10.1038/s41598-019-44004-w
10. Martínez, P.M., Dafonte, C., Gómez, Á.: Traceability in patient healthcare through the integration of RFID technology in an ICU in a hospital. Sensors **18**(5), 1627 (2018). https://doi.org/10.3390/s18051627
11. Wang, S., Zhang, S., Li, Z., Huang, L., Wei, Z.: Automatic digital ECG signal extraction and normal QRS recognition from real scene ECG images. Comput. Methods Programs Biomed. **187**(105254) (2020). https://doi.org/10.1016/j.cmpb.2019.105254
12. Shen, T.W., Laio, T.F.: Image processing on ECG chart for ECG signal recovery. In: 2009 36th Annual Computers in Cardiology Conference (CinC), Park City, pp. 725–728. IEEE (2009)

# Unsupervised Clustering Methods for Lung Perfusion Data Segmentation in Electrical Impedance Tomography

Arthur S. Ribeiro[1(✉)] , Yu H. W. Xia[2] , Mônica M. S. Matsumoto[2] , and Marcus H. Victor Jr.[2]

[1] Institute of Science and Technology - ICT, Federal University of São Paulo - UNIFESP, São José dos Campos, Brazil
`arthur.souza@unifesp.br`
[2] Electronics Engineering Division, Aeronautics Institute of Technology - ITA, São José dos Campos, São Paulo, Brazil

**Abstract.** In this work, we evaluated unsupervised clustering methods in segmenting the electrical impedance tomography image during the assessment of pulmonary perfusion by injection of hypertonic saline solution. In clustering the image pixels, we assume the existence of purely lung pixels (solely due to lung perfusion without effects from other organs) and hybrid pixels (which contain heart and lung effects together). We used data from 5 pigs to generate truth masks and assess the quality of clustering. Among the methods tested, the k-means with the cosine metric proved to be the best, as it obtained the 95% sensitivity median and the 90% specificity median. We prioritized minimizing the false negative cases and false positive cases, as it would overestimate regional pulmonary perfusion.

**Keywords:** electrical impedance tomography · lung perfusion · clustering · k-means · hierarchical clustering

## 1 Introduction

Electrical impedance tomography (EIT) is a non-invasive, non-ionizing, and functional imaging modality. The EIT machine generates images by mapping biological tissues' electrical characteristics. The technique is primarily used to continuously monitor lung ventilation and perfusion in mechanically ventilated patients [1–3]. For patients with different ventilation conditions (undergoing surgical procedures or lung diseases), EIT allows adjusting the ventilator settings based on the individual needs [4,5].

The protocol to estimate lung perfusion begins with injecting a hypertonic saline solution into the right atrium that takes the blood to the lungs. This procedure modifies and decreases the regional impedance [6]. Usually, 16 or 32 electrodes are positioned around the thorax, and a high frequency and low amplitude electrical current is applied for tissue excitation. The electrical potentials are measured and used to calculate the thorax impedance distribution. A

reconstruction algorithm uses this data to create the lung perfusion distribution map [7].

The EIT has been verified and validated as a useful method in clinical practice [8,9], mainly in monitoring lung ventilation. The perfusion estimation using EIT has a hindrance caused by the problem of signal interference between lungs and heart where a given voxel captures partial behaviors of the contrast passage through the right heart, the lungs, and returning to the left heart [7,10]. Some approaches were developed trying to segment the lung and heart regions. These methods include comparing functional EIT images with anatomic images from a computed tomography (CT) slice in the plane of the electrodes. Often, the approaches use unsupervised methods combining statistical and spectral analysis with an image processing algorithm to define the heart and lung regions of interest (ROI) [10].

Another study evaluated whether the EIT could determine the redistribution of lung perfusion elicited by one-lung ventilation. To find the lung and heart ROIs, they applied a Fourier transform to the pixels' time courses of relative impedance change to examine the frequency components of the EIT data. The heart ROI (negative slope) and lung ROI (positive slope) were acquired by calculating the slope of the linear regression fit between the local pixel and global EIT data [11].

This paper proposes a method to segment the EIT image during a hypertonic saline injection to evaluate lung perfusion by clustering the pixels with only lung behavior (solely lung) and pixels with hybrid behavior (heart and lung at the same time). The approaches rely on clustering tools to detect hybrid and lung pixels in the time-series data of five pigs.

## 2  Methods

### 2.1  Animal Experiment

The study relies on landrace swine's experimental data without previous or induced lung injury. The project was approved by the Ethics Committee on the Use of Animals (CEUA) of the Faculty of Medicine of the University of São Paulo (FMUSP) under number 1242/2019. The study was carried out in the animal ICU of the Laboratory of Medical Investigation in Experimental Pulmonology (LIM-09), located on the 4th floor of the FMUSP, and in the Tomography room of the Department of Pathology, located in the basement of the same building.

A belt containing 32 EIT electrodes was positioned in the plane corresponding to the 4th-5th intercostal space on the previously shaved skin. We used the EIT device (Enlight-1800, Timpel, São Paulo, Brazil). For evaluating perfusion by EIT, data were acquired during an apnea period of 30 s: 10 s pre-injection, followed by a rapid injection of 10 mL of 7.5% NaCl through a central catheter located in the right atrium of the animal. The images obtained by EIT have a sampling rate of 50 Hz and spatial resolution of 32-by-32 pixels in just one slice, representing about 15–20 cm of the lung.

## 2.2   Clustering Algorithms

We have evaluated two clustering algorithms: k-means and hierarchical clustering. We used the complete dataset to test and assess both methods.

The k-means is an iterative, data-partitioning algorithm that assigns n observations to precisely one of the k clusters. Where k is chosen before the algorithm starts. K-means treats each observation in the data as an object with a space location. The method finds a partition in which objects within each cluster, using a metric distance, are as close to each other as possible. Our experiment used two (k = 2) clusters: hybrid and lung pixels. We chose the correlation and cosine metric distance based on the pig data signal. An assessment with a metric distance from an origin (cosine and correlation) gives a clustering considering the phase and magnitude between two time series, it is more useful than use the Euclidean metric distance that only find the shortest distance between two time series.

The cosine metric, also called cosine similarity, calculates the cosine of the angle between two points (treated as vectors). The cosine metric distance is shown in Eq. 1,

$$d_{cos}(a, b) = 1 - \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \sqrt{\sum_{i=1}^{n} b_i^2}} \tag{1}$$

where $a_i$ and $b_i$ are the samples of vectors $a$ and $b$, respectively, $n$ is the number of samples, with the $d_{cos}(a, b)$ as the cosine metric distance between $a$ and $b$. The cosine metric distance is one minus the cosine similarity and has a range of values between 0 and 2.

The correlation metric distance is shown according to Eqs. 2, 3 and 4,

$$d_{corr}(a, b) = 1 - \frac{\sum_{i=1}^{n} (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{n} (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^{n} (b_i - \bar{b})^2}} \tag{2}$$

$$\bar{a} = \frac{\sum_{i=1}^{n} a_i}{n} \tag{3}$$

$$\bar{b} = \frac{\sum_{i=1}^{n} b_i}{n} \tag{4}$$

where $a_i$ and $b_i$ are samples of $a$ and $b$ respectively, $n$ is the number of samples, with $d_{corr}(a, b)$ as the correlation metric distance between $a$ and $b$. The correlation metric distance is one minus the correlation and take a range of values from 0 to 2.

The hierarchical clustering method includes grouping data over various scales by creating a cluster tree or dendrogram. The tree is not a single set of clusters but a multilevel hierarchy, where clusters at one level are joined as clusters at the next level. To perform the clustering, the purpose is to find the similarity or dissimilarity between every pair of objects in the data set, calculating the distance between objects. We also used two clusters (hybrid and lung pixels).

Based on the pig data signal and to compare k-means and the hierarchical clustering methods, we used the correlation and cosine metric distance in clustering routines [12].

## 2.3   Implementation

The dataset consists of each pixel's impedance waveforms per time (samples). The waveforms are shown in Fig. 1(a) and correspond to those pixels that varied during the acquisition. The silent pixels were excluded from the analysis.

During a hypertonic saline injection, the time series of the impedance per time (samples) can be associated with the blood volume, while the time derivative of the impedance per time (samples) is associated with the blood flow. The blood volume does not give quantitative information about how much and what conditions (slow or fast flow) the blood is reaching the lung. Therefore, we decided to analyze only the time derivative waveforms (Fig. 1(b)).
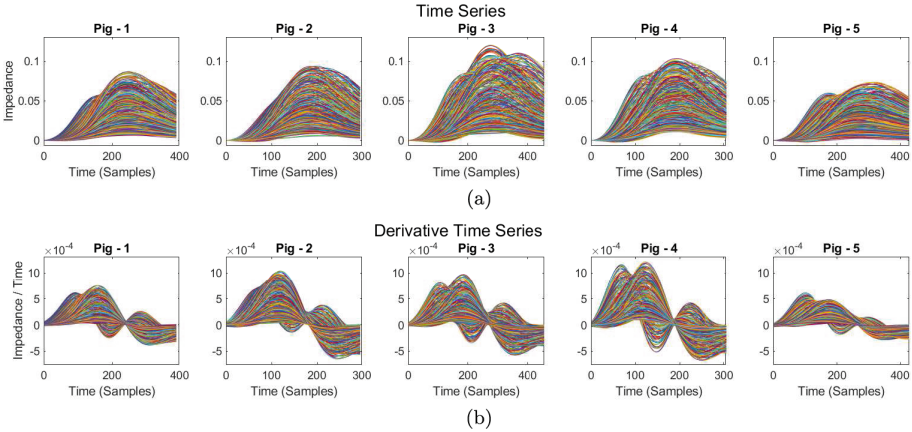


**Fig. 1.** Time series (a) and derivative (b) for the electrical impedance signal of each pixel.

The derivative signals are clustered using k-means and hierarchical clustering with the correlation and cosine metrics. After finishing the clustering step, we reorganized the data in hybrid masks with $32 \times 32$ pixels for each derivative result and we apply a morphological operation to dilate and expands the hybrid region. The kernel shape used was "square" with size "2" to increase the true positive pixels and reduce the pixels incorrectly classified.
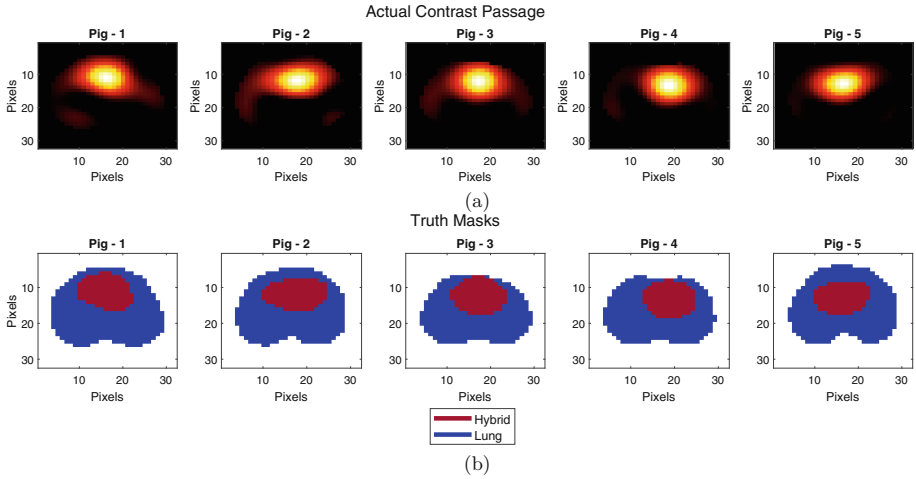
**Fig. 2.** Actual contrast passage (a) used to create the truth mask (b) with binary colors representing the hybrid and lung regions manually segmented.

The result masks were compared with the truth masks (manually created). The method to create the truth masks was observing the EIT derivative signal ($32 \times 32$ pixels) in time (samples) looking frame by frame and deciding to use the frame that do not presented lung perfusion pixels. The derivative perfusion signal was set with a sequential color scale as shown in Fig. 2(a). The pixels with higher impedance/time present a lighter color. Thus, we selected only the pixels with color $\geq 25\%$ of the chosen scale to the hybrid region (hybrid pixels with considerable signal amplitude), and the others were considered as lung region of the truth mask. The results of the truth masks are visualized in Fig. 2(b).

All hybrid masks comparison results are evaluated quantitatively using the following parameters:

– True Positive (hybrid pixels considered as hybrid pixels);
– True Negative (lung pixels considered as lung pixels);
– False Positive (lung pixels considered as hybrid pixels);
– False Negative (hybrid pixels considered as lung pixels).

The parameters were used to calculate the sensitivity, specificity, and accuracy of each clustering algorithms method.

## 3   Results

The analysis focus was on the masks comparisons. Each mask was compared with the truth mask (Fig. 2(b)). The k-means and hierarchical clustering results using the cosine metric are visualized in Tables 1 and 2 respectively. The results using the correlation metric are shown in Tables 3 and 4, respectively.

(a)



(b)
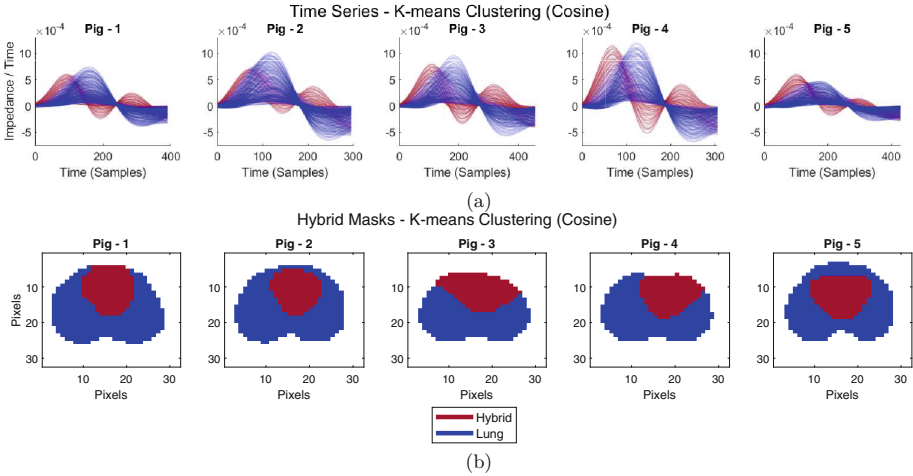
**Fig. 3.** Time series k-means clustering (Cosine) electrical impedance per samples by each pixel and its Hybrid Mask with K-means Clustering (Cosine); (a) Time series K - means Clustering (Cosine); (b) Hybrid Mask - K - means Clustering (Cosine).

**Table 1.** k-means clustering (cosine metric). Sensitivity (TPR) stands for true positive rate; Accuracy (ACC) stands for the accuracy of the method; and Specificity (TNR) stands for true negative rate. All values are in percentage.

| Pigs | Sensitivity % | Specificity % | Accuracy % |
|---|---|---|---|
| 1 | 95 | 90 | 91 |
| 2 | 83 | 92 | 90 |
| 3 | 93 | 85 | 87 |
| 4 | 96 | 90 | 91 |
| 5 | 98 | 88 | 90 |
| Median | 95 | 90 | 90 |

The tables were used to evaluate which method using each metric quantitatively has shown the best performance. The objective was to examine the higher sensitivity percentage median. The high sensitivity median represents that the method found the higher number of true positive cases (hybrid pixels as actual hybrid pixels) with the lower number of false negative cases (hybrid pixels considered as lung pixels). Evaluating the false negative instead of the false positive is damage control. For the proposed problem, taking a lung pixel as a hybrid pixel is less troublesome than taking a hybrid pixel as a lung pixel.

**Table 2.** Hierarchical clustering (cosine metric). Sensitivity (TPR) stands for true positive rate; Accuracy (ACC) stands for the accuracy of the method; and Specificity (TNR) stands for true negative rate. All values are in percentage.

| Pigs | Sensitivity % | Specificity % | Accuracy % |
|---|---|---|---|
| 1 | 92 | 91 | 91 |
| 2 | 73 | 95 | 89 |
| 3 | 89 | 90 | 90 |
| 4 | 95 | 94 | 94 |
| 5 | 98 | 89 | 91 |
| Median | 92 | 91 | 91 |

**Table 3.** k-means clustering (correlation metric). Sensitivity (TPR) stands for true positive rate; Accuracy (ACC) stands for the accuracy of the method; and Specificity (TNR) stands for true negative rate. All values are in percentage.

| Pigs | Sensitivity % | Specificity % | Accuracy % |
|---|---|---|---|
| 1 | 94 | 89 | 90 |
| 2 | 73 | 93 | 88 |
| 3 | 82 | 93 | 90 |
| 4 | 96 | 89 | 91 |
| 5 | 98 | 90 | 92 |
| Median | 94 | 90 | 90 |

**Table 4.** Hierarchical clustering (correlation metric). Sensitivity (TPR) stands for true positive rate; Accuracy (ACC) stands for the accuracy of the method; and Specificity (TNR) stands for true negative rate. All values are in percentage.

| Pigs | Sensitivity % | Specificity % | Accuracy % |
|---|---|---|---|
| 1 | 91 | 92 | 92 |
| 2 | 76 | 94 | 90 |
| 3 | 88 | 90 | 89 |
| 4 | 96 | 91 | 92 |
| 5 | 92 | 94 | 94 |
| Median | 91 | 92 | 92 |

Searching for a method with the higher sensitivity median, we found that the best method was the k-means clustering using the cosine metric, as shown in Table 1. This method presents its clustered time series and masks shown in Fig. 3(a) and 3(b), respectively. Furthermore, the comparison between the method masks and the truth masks is shown in Fig. 4.
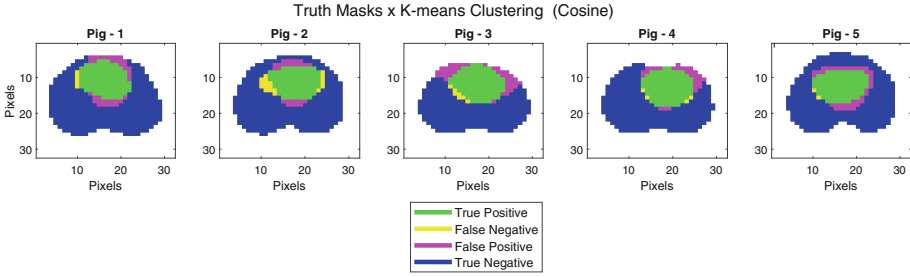
**Fig. 4.** Comparisons among the truth masks versus k-means clustering (cosine metric) masks.

## 4    Discussions

The k-means clustering (cosine metric) has gotten a 95% of sensitivity and 90% of specificity, representing a considerable probability of detection and a high selectivity of the method.

The understanding of the false pixels cases are that the FN (yellow pixels) means that the hybrid pixels in hybrid mask, has true positive almost all hybrid pixels of the truth mask. However, the FP (pink pixels) reveals some hybrid pixels in hybrid masks should be classified as lung pixels. This implies that the k-means clustering (cosine) finds the true positive with the truth masks hybrid pixels, but it needs to be investigated why some lung pixels are classified as hybrid pixels. Reduce the FP cases is the best way to increases the specificity median.

One approach to investigate what could be interfering in false pixels cases was to plot the clustered time series of this method (Fig. 3(a)), with the waveforms of FP and FN parameters taken in Fig. 4. We show such a plot for this method in Fig. 5.

Figure 5 shows essential information about the main errors in our method. We verify that the hybrid pixels and the lung pixels have very different and defined waveforms for each one. Comparing the hybrid and lung pixels peak to the peak of incorrectly classified pixels - FN (yellow pixels) and FP (pink pixels). We observed that the yellow and pink pixels are in an intermediary peak. This information suggests that the intermediary peak influences a wrong hybrid classification by the k-means clustering (cosine).

Furthermore, comparing the pink and yellow waveforms with the lung waveforms, we notice that the similarity (phase and magnitude) between then are higher than the hybrid waveforms. The hybrid waveforms have defined negative region, while the pixels incorrectly classified have not. This information suggests that the cosine metric distance used to create the hybrid masks, has some problems when the waveforms presents a lower similarity.

In synthesis, the method using the k-means clustering (cosine) to create the hybrid masks was the best approach giving a high probability of detection and selectivity, and promising results to be a useful segmentation method to find the

hybrid and lung region in a perfusion data set. The main problem is the wrong classification of the lung pixels as hybrid pixels (pink pixels) and to understand the errors, more tests need to be performed.
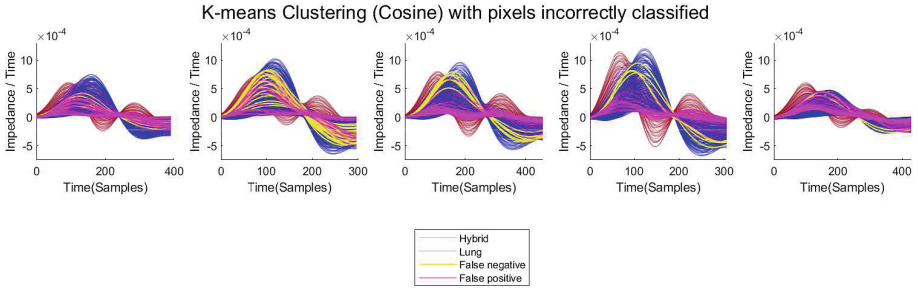


**Fig. 5.** Time series of k-means clustering (cosine metric) with pixels incorrectly classified (false negative and false positive waveforms) from the comparisons of Fig. 4.

## 5    Conclusion

This work has shown and compared different methods for waveform clustering in EIT data during lung perfusion assessment. The approaches included using k-means and hierarchical clustering methods using the cosine and correlation metrics. The best result was using the time-derivative dataset waveforms with k-means clustering (cosine metric).

This method presents a 95% sensitivity median with a 90% specificity median, highest within the compared methods. The considerable accuracy represents promising results, but more work should be performed to understand the errors between the clustering method and the truth masks.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Frerichs, I., Amato, M.B.P., van Kaam, A.H., et al.: Chest electrical impedance tomography examination, data analysis, terminology, clinical use and recommendations: consensus statement of the translational EIT development study group. Thorax **72**, 83–93 (2016)
2. Perier, F., Tuffet, S., Maraffi, T., et al.: Effect of positive end-expiratory pressure and proning on ventilation and perfusion in COVID-19 acute respiratory distress syndrome. Am. J. Respir. Crit. Care Med. **202**(12), 1713–1717 (2020). PMID: 33075235

3. Perier, F., Tuffet, S., Maraffi, T., et al.: Electrical impedance tomography to titrate positive end-expiratory pressure in COVID-19 acute respiratory distress syndrome. Crit. Care **24**(1), 678 (2020)
4. Leonhardt, S., Pikkemaat, R., Stenqvist, O., et al.: Electrical impedance tomography for hemodynamic monitoring. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 122–125. IEEE (2012)
5. Roldán, R., Rodriguez, S., Barriga, F., et al.: Sequential lateral positioning as a new lung recruitment maneuver: an exploratory study in early mechanically ventilated COVID-19 ARDS patients. Ann. Intensive Care **12**(1), 13 (2022)
6. Brown, B.H., Leathard, A., Sinton, A., et al.: Blood flow imaging using electrical impedance tomography. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 13, pp. 307–308 (1991)
7. Costa, E.L., Lima, R.G., Amato, M.B.: Electrical impedance tomography. Curr. Opin. Crit. Care **15**, 18–24 (2009)
8. Frerichs, I., Hinz, J., Herrmannv, P.: Regional lung perfusion as determined by electrical impedance tomography in comparison with electron beam CT imaging. IEEE Trans. Med. Imaging **21**, 646–652 (2002)
9. Richard, J.C., Pouzot, C., Gros, A., et al.: Electrical impedance tomography compared to positron emission tomography for the measurement of regional lung ventilation: an experimental study. Crit. Care **13**, R82 (2009)
10. Ferrario, D., Grychtol, B., Adler, A., et al.: Toward morphological thoracic EIT: major signal sources correspond to respective organ locations in CT. IEEE Trans. Biomed. Eng. **59**, 3000–3008 (2012)
11. Frerichs, I., Pulletz, S., Elke, G., et al.: Assessment of changes in distribution of lung perfusion by electrical impedance tomography. Respiration **77**, 282–291 (2009)
12. The MathWorks Inc.: Matlab® and statistics and machine learning toolbox $^{TM}$ release (r2020a). Natick, Massachusetts, USA (2020)

# New Strategy to Hyperspectral Image Segmentation Using Principal Components Analysis

B. A. Augusto[1] , L. Bachman[2] , M. C. Moraes[1] , F. A. Lima[2],
and T. M. Pereira[1(✉)]

[1] Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo, São José dos Campos,
Brazil
t.pereira@unifesp.br
[2] Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade de São Paulo,
Ribeirão Preto, Brazil

**Abstract.** Imaging exams are fundamental in the treatment of various diseases. Through them, it is possible to obtain faster, more accurate and safer diagnoses. Infrared spectroscopy exams help to differentiate healthy tissue from pathological tissue, identifying their characteristics, but this generates a large amount of data that requires a lot of time and computational capacity to process. Thus, this work intends to optimize the pre-processing of micro-FTIR images for cancer diagnosis, using k-means clustering and Principal Component Analysis (PCA).

**Keywords:** Fourier-Transform Infrared Spectroscopy · FTIR · Micro-FTIR · clustering · Image pre-processing · PCA · k-means · Histological Sections · H&E

## 1 Introduction

According to the world health organization, cancer is a leading cause of death in the world and represents around 10 million deaths only in 2020 [1]. Cancer is a generic term that describes more than 50 diseases and is characterized by uncontrolled growth of cells that can be spread over the body in a process named as metastasis.

There are several treatments for cancer that can be divided into 3 subtypes: surgery, radiotherapy, and Chemotherapy [2]. Therefore, the corrected diagnosis at the early stages is necessary for a better prognosis and survival rate [3]. In this way, several methods have been applied for diagnoses and the histopathology exam is one of the most important for correct classification as well as cancer staging. However, this exam has several limitations, mainly, due to cancer detection at early stages [4].

To overcome this situation, one of the possibilities is to collect biochemical markers from the suspicious biopsy. Several studies [4–6] in the literature reveal that infrared spectroscopy imaging (ISI) can be used to extract biochemical information on biopsy slides. The spectroscopy technique is label-free and demands simple sample preparation [7, 8].

The ISI collects a big amount of chemical information that comes from the tissue by analyzing the interaction between infrared light and molecules. Therefore the signal that comes from the samples can be related to the tissue as well as unwanted contaminants such as water vapor and paraffin [9–11]. The paraffin is used during the sample preparation for biopsy preservation [9].

Therefore, it is necessary to use appropriate image processing techniques to remove the chemical information from contaminants and access the biochemical markers correctly [11]. The image processing is divided into 2 steps the first is named pre-processing. The main goal of this step is to remove the artifacts related to contaminants and light scattering [10].

The second step is related to image segmentation. The main method used in the ISI for image segmentation is k-means clustering. This method is well established in the FTIR literature [12–15], and it can segment the histological structures into ISI images. However, the step is time-consuming and increases exponentially with image size. The main reason is related to the number of features (wavenumbers points) in each pixel which can be up to 1000 [12].

Another important point in data processing is to improve the signal to noise (RSN) by mathematical steps. The main method used to improve the RSN is the reconstruction of the spectrum using only first n PCs obtained from a PCA transform [16–18]. The methodology is well established in the FTIR literature [7, 19, 20].

The time-consuming processing to access the segmentations is still a bottleneck that has not been overcome by literature. There are no articles in the literature showing this strategy for optimizing the time in micro FTIR images.

Therefore, the authors in the present study focus on overcoming this problem by performing PC transformation before k-means clustering. Accordingly, it is necessary to develop new strategies to perform this step faster. There are several statistical approaches to reduce the number of features to preserve almost all information in the IR spectrum. One of the most common is the principal component analysis (PCA).

The goal of the present study is to evaluate the performance of the use of PCA transform before the segmentation step by k means clustering.

## 2 Materials and Methods

### 2.1 Samples

The samples used in the present studies came from Clinics Hospital of Ribeirão Preto. The hospital ethics committee approved the use of the sample for this study. The ethics committee number are 3285/2002, 1902/2002.

### 2.2 Data Collection

All colon biopsies were preserved in paraffin blocks. Two consecutive cuts of 5 mm thickness were performed. The first cut was deposited on CaF2 windows that were transparent to Infrared radiation. The second slide was deposited on a glass slide and stained with hematoxylin and eosin (H&E). Both cuts were not de-waxed chemically data collection.

We collected 71 hyperspectral images from 55 human colon tissues. The images were collected using a spotlight 400 FTIR (PerkinElmer, Massachusetts, USA). The hyperspectral images were recorded in transmission mode with 6.25 µm pixel resolution and each spectrum was obtained by coaddition of 16 interferograms. Before each collection, an appropriate background spectrum was collected in the clean region of the caf2 window.

The images were collected using microscope, in transmission mode, with an abroad IR source that shedding light on the sample with wavelengths ranges from $1,3\,(750\,cm^{-1})$ to $0,25$ µm $(4000\,cm^{-1})$. This region is used by the literature to classify histological structures because there are several biomolecules such as proteins, collagen, and nucleic acids that absorb light in these wavelength regions.

## 2.3 Pre-processing

The images collected were processed in google Colab using a toolbox written in python 3 and developed by the authors. After the dataset was imported, an area containing only paraffin was selected for the built model for removing the variance related to paraffin. The model used was linear and named Extended multiplicative signal correction (EMSC). The EMSC model was constructed using the first's 5 principal components (PCs) of the paraffin dataset and a 4th polynomial order was used to remove the differences in the baselines among the spectra of the image. The parameters used in the EMSC in the present study were optimized and published in the previous study [9].

After applying the EMSC, improvements were performed in each spectrum of the image, and the spectra with poor signal-to-noise ratio and/or without tissue were removed.

The processed image was transformed into 4 distinct images: the original images where each pixel has a spectrum between $900$–$1800\,cm^{-1}$ (SPC-image). The other three images were created by PC analysis. In these 3 images, each pixel has the n first PCs that represent 99% (PC99-image), 95% (PC95-image) and 90% (PC90-image) of the accumulated variance from the dataset.

## 2.4 K-Means Segmentation

The k-means clustering is a common method used for FTIR image segmentation. The k-means algorithm splits the pixel into the k group. Each group is represented by a centroid that is the mean of all n spectra labeled by the algorithm. The algorithm initialization starts to select k centroids (pixels) randomly. The Euclidian distance (ed) between all pixels of the image (spectra) and each k centroid are calculated and all pixels in the image are labeled according to minima ed. At the end of this step a new centroid is calculated by mean n labeled spectra. This process is repeated until some stop criteria are reached.

The 4 images (SPC, PC99, PC95, PC90) from each sample were submitted to k-means clustering, using as a similarity measure the Euclidean distance. The k-means on each image were performed 15 times to avoid problems related to cluster initialization [9]. This is because the results obtained from the k-means clustering method depend on the initialization points that are selected randomly. Therefore, the results from k-means are not always connected to the best clustering. One of the methods to overcome this

situation is to perform the k-means several times and choose the results that minimize the variance inside of a group and maximize the variance inter-group. The variance is calculated based on the similarity metric.

## 3   Results and Discussions

### 3.1   Time and Quantity of Components

One of the ways to measure the advantages of the PCA technique, in the segmentation step, is by observing the processing time. Table 1 shows the average number of components used in each situation and the mean of processing time required to k-means segmentation of all images considering the time spent in the PCA (when used). The number of features is defined as the number of absorbance point when considering the SPC-image and PCs-image the features is the number of PC variables that describes 99, 95 or 90% of the accumulate variance.

**Table 1.**  Time and quantity of components

| Data | Features (n) | Time (s) |
|------|------|------|
| SPC-image | 451 | 135.9 |
| PC99-image | 68 | 26.7 |
| PC95-image | 33 | 17.7 |
| PC90-image | 21 | 14.11 |

We can observe that there is a big difference in the number of components analyzed in each situation. When we applied only k-means clustering, the software used all the 451 points of the spectrum (SPC-Images). In the following situations, with 99% (PC99-image), 95% (PC95-image), and 90% (PC90-image) of accumulated variance, even using the same number of spectra, the average value is, respectively, 68, 33, and 21 features (PCs).

We could also notice a big difference between the processing time of the original data and the time of the data treated with PCA. The k-means segmentation time with PC99-images took only 17% compared to SPC-images, and the analysis with PC90-images took only 10.9% of the time. It represents a very significant time reduction, mainly taking into account that several works in the literature report that one of the biggest obstacles [13, 14] to the large-scale use of the FTIR imaging technique within the area of histopathology is due to the long time taken to acquire the images.

## 3.2   Similarity Between the Images

To measure how similar the images obtained by the k-means clustering method are, the coefficients of mutual information between the images were calculated [15].

The results of this analysis can be seen in Table 2, in which the average coefficients of mutual information between all 71 images obtained from the segmentations of the SPC-images and the PC95-image and PC90-image are shown, compared to the values of the PC99-image, of each number of clusters.

Note that this value varies between 0 (totally different images) and 1 (images that carry the same information).

**Table 2.**  Mutual information between the images.

| Clusters | SPC-img vs PC99-img | PC95-img vs PC99-img | PC90-img vs PC99-img |
|---|---|---|---|
| 2 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 |
| 4 | 0.49 | 0.99 | 0.97 |
| 5 | 0.49 | 0.99 | 0.97 |
| 6 | 0.53 | 0.98 | 0.96 |
| 7 | 0.52 | 0.98 | 0.96 |
| 8 | 0.54 | 0.97 | 0.94 |
| 9 | 0.52 | 0.97 | 0.94 |
| 10 | 0.52 | 0.95 | 0.93 |

As we can see when the process divides the images into 2 or 3 clusters, the images are practically identical regardless of whether we are comparing the images with or without PCA. After 4 clusters, a significant difference begins to exist between the SPC-image and its PCs-images. However, among the PCA images, their mutual information from all images is relatively high, always above 0.9, this raised the hypothesis that this difference with the SPC image is the removal of errors, so we opted for more punctual and qualitative images concerning their respective H&E image.

One of the advantages of the k-means technique over others for segmentation in ISI images is related to the fact that it is very sensitive to differences in the dataset [16], especially in images with a high number of clusters, so the biochemical differences in the tissue are well defined, and even with very different colors, the images obtained are very similar to the H&E, as can be seen in the comparison of Figs. 1, 2, 3, 4 and 5.
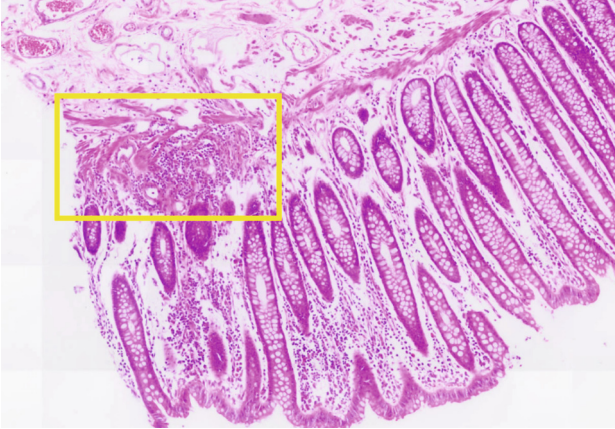
**Fig. 1.** H&E image from a normal colon tissue
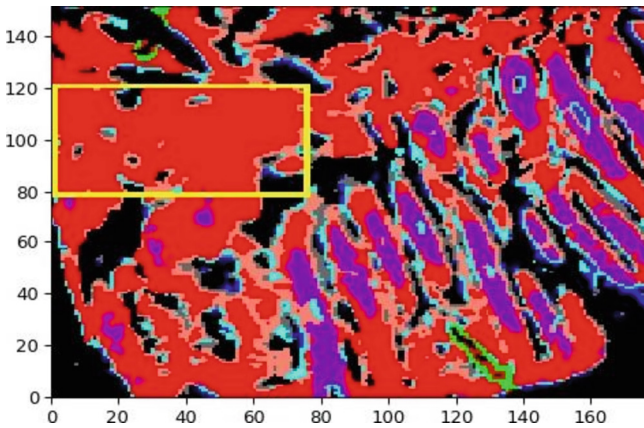


**Fig. 2.** The 10 clusters k-means SPC-image.

We can also notice that, in agreement with the table of Fig. 2, the images obtained through the PCA, Figs. 3, 4 and 5, are very similar, and it is even quite difficult to find differences without a very detailed analysis, while all of them have a great difference concerning image obtained only with k-means, Fig. 2.

So for a better understanding of these small differences, Figs. 6, 7, 8, 9 and 10 show a specific part of the tissue where only PC-images can segment the regions with a high amount of cells ( purple cluster) from regions with stroma ( red cluster).

Now with the more detailed images, we can see the small differences between the PCA images, Figs. 8, 9 and 10. We can also compare that the presence of purple subdivisions makes the PCA images closer to the morphological division of H&E, which contain several subdivisions in the tissue that are not shown by the image obtained with k-means only.
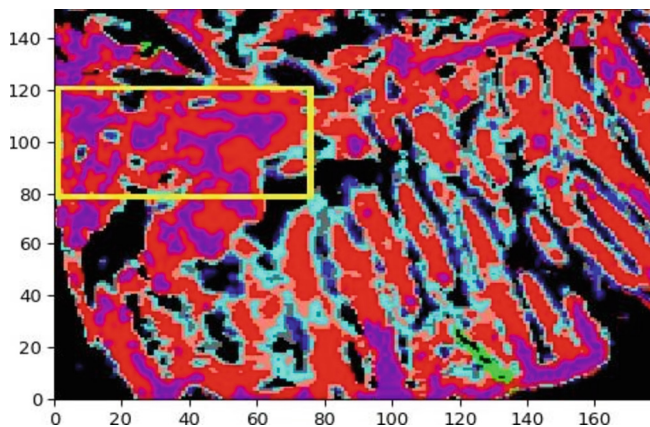
**Fig. 3.** The 10-cluster k-means from a PC99-image
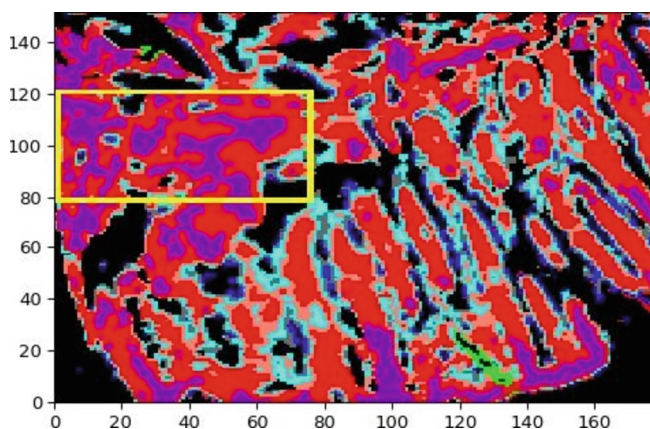


**Fig. 4.** The 10 cluster k-means from a PC95-image

This difference in image segmentation may be due to the interaction of PCA used to exclude less relevant data for segmentation, with the low signal-to-noise ratio of hyperspectral images, thus causing better noise removal and rendering the images obtained more faithful to the delimitation of the structures.

The quality of k-means segmentation using PC variables over spectral data is related to the fact that this segmentation method is highly sensitive to differences among the data. These differences can be related to contaminants not related to biochemical of the sample but vibrational contribution of water vapor, paraffin and/or noise. When we used only firsts k PCs, we also performing a noise removal [16].
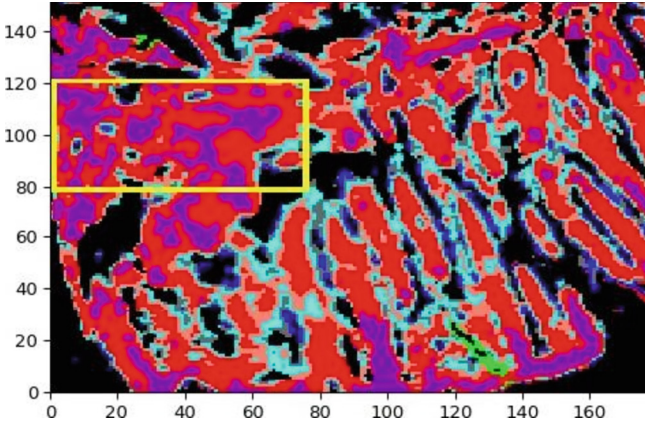
**Fig. 5.** 10 cluster k-means from a PC90-image.
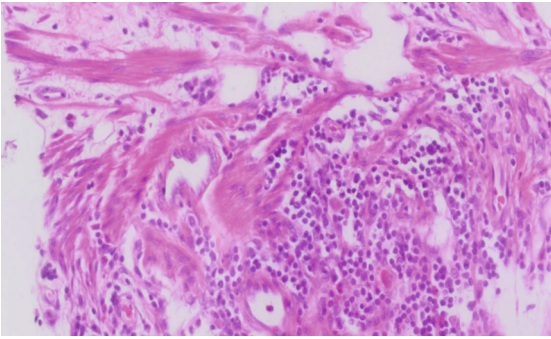


**Fig. 6.** Small part of Fig. 1
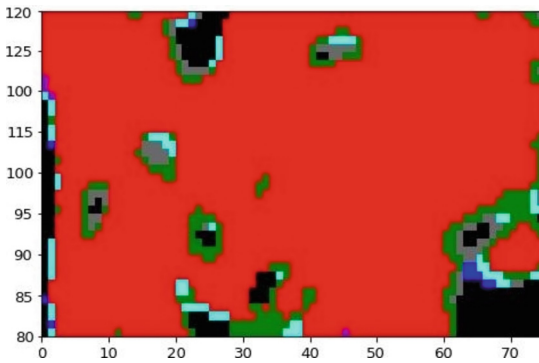


**Fig. 7.** Small part of Fig. 4 (yellow rectangle).
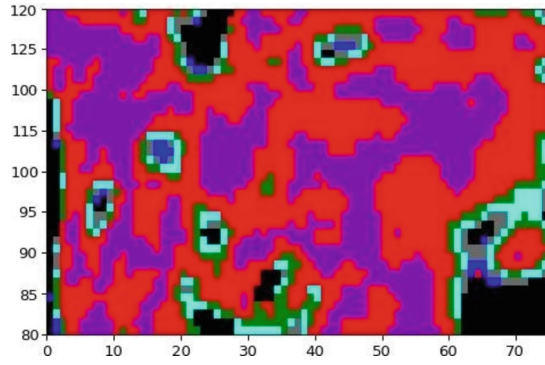
**Fig. 8.** Small part of Fig. 3 (yellow rectangle).
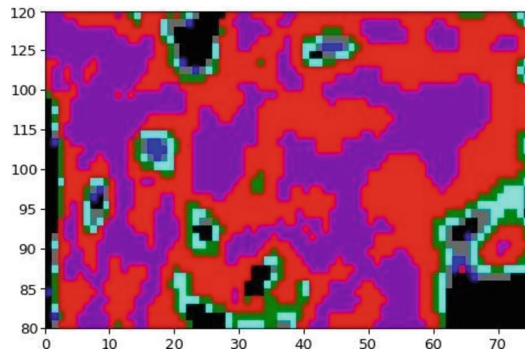


**Fig. 9.** Small part of Fig. 4 (yellow rectangle).
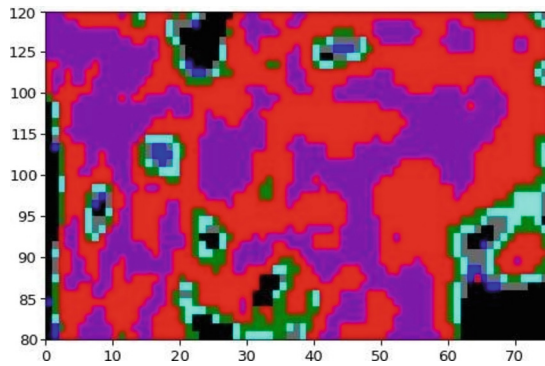


**Fig. 10.** Small part of Fig. 5 (yellow rectangle).

## 4 Conclusion

This work allowed us to evaluate the efficiency of the PCA technique in the preprocessing of hyperspectral images, analyzing the differences in the amount of data, the time to process them and between the images obtained.

The results showed that images involving PCA have a more defined segmentation process compared to traditional k-means. In addition, the technique allows a clearer delineation of small internal structures and promotes an improvement in the signal-to-noise ratio. Direct comparison between results obtained in the present study and literature is not possible so far, by not having same dataset. Nonetheless, future works may focus on implementing works from literature or share datasets to allow direct comparisons.

## References

1. Ferlay, J., et al.: Int. J. Cancer **149**, 778–789 (2021)
2. Treatment of Breast Cancer - American Family Physician. https://www.aafp.org/afp/2010/0601/p1339.html. Accessed 27 May 2022
3. Parsons, A., Daley, A., Begh, R., Aveyard, P.: BMJ **340**, 251 (2010)
4. Bird, B., Miljkovic, M.S., Remiszewski, S., Akalin, A., Kon, M., Diem, M.: Lab. Invest. **92**, 1358–1373 (2012)
5. Teixeira, C.S., et al.: Analyst **134**, 2361–2370 (2009)
6. Aguiar, J.C., Mittmann, J., Ferreira, I., Ferreira-Strixino, J., Raniero, L.: Spectrochim. Acta A Mol. Biomol. Spectrosc. **142**, 80–85 (2015)
7. Hughes, C., et al.: Analyst **140**, 2080–2085 (2015)
8. Pilling, M.J., Gardner, P., Henderson, A., Brown, M.D., Bird, B., Clarke, N.W.: Faraday Discuss
9. de Lima, F.A., et al.: Analyst. https://doi.org/10.1039/c6an01975g
10. Afseth, N.K., Kohler, A.: Chemom. Intell. Lab. Syst. **117**, 92–99 (2012)
11. Pereira, T.M., Diem, M., Bachmann, L., Bird, B., Miljković, M., Zezell, D.M.: Analyst **145**, 7907–7915 (2021)
12. Travo, A., et al.: Histopathology **56**, 921–931 (2010)
13. Nallala, J., Lloyd, G.R., Shepherd, N., Stone, N.: Analyst **141**, 630–639 (2016)
14. Bassan, P., Weida, M.J., Rowlette, J., Gardner, P.: Analyst **139**, 3856–3859 (2014)
15. Perez-Guaita, D., et al.: Faraday Discuss. **187**, 341–352 (2016)
16. Diem, M., et al.: J. Biophotonics **6**, 855–886 (2013)
17. Iem, M.A.X.D., Apamarkakis, K.O.P., Chubert, J.E.S., Ird, B.E.B., Oston, B.: 307–318
18. Bonnier, F., Byrne, H.J.
19. Schubert, J.M., et al.: Lab. Invest. **90**, 1068–1077 (2010)
20. Bassan, P., et al.: Analyst **135**, 268–277 (2010)

# Application of Digital Image Processing Techniques for Segmentation of Meningiomas in Magnetic Resonance Images

Beatriz Vasconcelos de Oliveira$^{(\boxtimes)}$ ⓘ, William Marcos Dantas ⓘ, Caroline Dantas Vilar ⓘ, and Heliana Bezerra Soares ⓘ

Biomedical Engineering Department, Federal University of Rio Grande do Norte, Natal, RN, Brazil

`beatriz.vasconcelos.106@ufrn.edu.br`

**Abstract.** Brain tumor is of the main diseases that affect the brain, meningioma being the most common among all types of brain tumors. This pathology has a benign character in most situations. In order to aid in a rapid and non-invasive prognosis, magnetic resonance imaging is used as the means of diagnosis. Therefore, with the objective of developing a tool to assist in the analysis of the lesion and medical decisions, an algorithmic strategy of automatic segmentation was developed for magnetic resonance images of meningiomas, in order to delimit the tumor region from the application of processing techniques digital images. The results obtained in this analysis corresponded to about 66% of correct answers in the segmentation of the region of interest.

**Keywords:** Brain tumor · meningioma · digital image processing

## 1 Introduction

Brain tumor is a disease that affects the nervous system and can be classified into two types: primary and secondary. Primary brain tumors are those that originate in the central nervous system itself, while secondary tumors originate from another region of the body and reach the brain through metastases [1].

Among the most frequent primary brain tumors is the meningioma, characterized by its common location in the meninges region and classified as a benign type of tumor for most cases, it also presents slow growth, with exposure to ionizing radiation as one of the main risk factors [1].

The World Health Organization has specified a classification that encompasses four grades of lesions to determine the severity of each tumor, subdivided into two categories: low grade (types I and II) and high grade (types III and IV). In relation to meningiomas, this classification is characterized by only three levels: grade I – lesions with low infiltration potential, considered typical; grade II – lesions with high infiltration potential, called atypical; grade III – less common type, but with anaplastic or malignant characteristics [2].

The diagnostic process of this pathology includes the analysis of neuroimages as a crucial step, the most common exams being: computed tomography and magnetic resonance imaging.

Magnetic resonance imaging is an image acquisition process based on nuclear spins and their interactions with the magnetic field generated by the equipment. These interactions are more intense in the hydrogen atom, since its nucleus has only one proton, so it presents angular momentum when subjected to the magnetic field. Therefore, as there is a diversity in the amount of hydrogen present in the tissues, for each sequence of electromagnetic pulses the tissues will present a different recovery time, which translated into the final image, represents a different intensity for each type of tissue [3].

Although this type of image is excellent for identifying the existence of the tumor and delimiting its region, there are some points that cannot be met by this modality, such as: definition of the type, degree of malignancy and approximation of the impregnation margins. However, the limitations mentioned above can already be overcome from the association of magnetic resonance images with more recent techniques [4].

The evaluation of imaging tests to identify tumor regions still takes place manually by specialist physicians, which requires a great deal of professional experience and a well-trained eye to identify these pathologies. This specificity can influence the final segmentation area, since this procedure is subject to natural human errors and is difficult to visualize, due to the nature of the image that has many similar levels of intensity [5]. Therefore, the computer-aided diagnosis system – CAD shows itself as an important alternative to aid in the detection and segmentation of brain tumors. In addition to contributing to an early diagnosis, which is a crucial factor in the efficiency of the treatment response and consequently in the patient's chances of recovery.

According to data from Brazilian National Cancer Institute – INCA (2022), among the total percentage of malignant tumors, cancer of the central nervous system corresponds to 1.8%. In addition, as estimate carried out in the year 2020 points to the emergence of 11.090 new cases, with a mortality rate that can reach the amount of 9.355 of the total cases [6].

Based on this problem and the data presented in the previous paragraph, the objective of this work is to develop a computational algorithm that, applied to magnetic resonance images, is able to identify and segment the region referring to meningioma-type brain tumors, through the application of digital image processing techniques.

## 2 Materials and Methods

### 2.1 Database

The database was produced and made available by Nanfang General Hospital and Tianjing Medical University General Hospital, both in China, and is publicly available for research use. The data are composed of T1-weighted MRI images of the brain region and have three distinct types of brain tumors: meningiomas, gliomas and pituitary tumors. The dataset refers to 233 patients, totaling 3.064 slices, of which 708 are meningiomas [7].

In a very generic way, we can explain that during an MRI exam, the images are generated through a kind of scan, that is, at the end of the analysis there are several

slices of the region of interest of a patient, better known as slices. Therefore, the same individual may contain more than one image referring to their lesion.

In this database, the information of each slice is distributed within a struct with the extension.mat, and includes the identification of the tumor type, the patient's identity code, the original image (see Fig. 1), a vector of coordinates and a tumor mask [7].
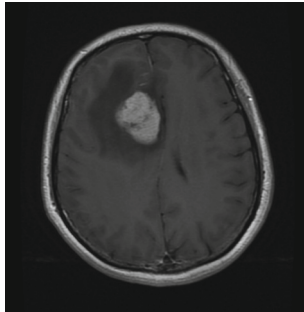


**Fig. 1.** Original image extracted from.mat file.

In addition to the original images of each patient, the author has a binary mask of the tumor region (see Fig. 2), which was segmented by three experienced professionals in the neurological area. From the analysis of these images in comparison to the originals, we noticed a discrepancy in the information, since the manual segmentation brought a visually larger region than the original. Therefore, for the purpose of evaluating the results, this study did not use these images as a parameter for comparison.
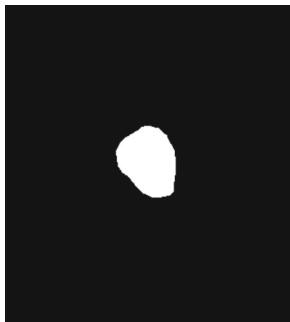


**Fig. 2.** Specialist segmented tumor mask, extracted from .mat file.

## 2.2   Digital Image Processing

For the development of the algorithm we used a sequence of digital image processing strategies, as described in Fig. 3, with the objective of segmenting the characteristic region of the tumor from the magnetic resonance image.
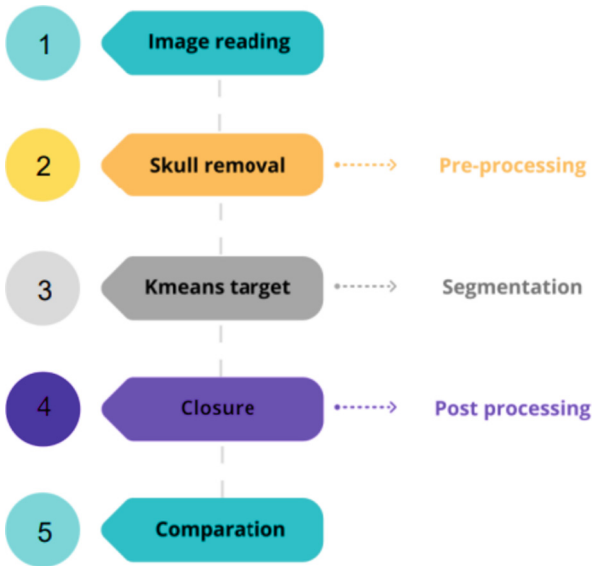


**Fig. 3.**  Block diagram with the techniques applied in the system.

Executing the first step of the algorithm, reading the image, it was necessary to extract the original image that was inside a struct and consequent conversion of this figure to the.png format. In this way, we were able to work in the other stages with the original image.

We have divided the processes into three main phases: pre-processing, segmentation and post-processing. In the pre-processing, we followed the strategy of removing the cranial contour through mathematical morphology. To perform the segmentation we adopted the k-means method. In post-processing, we reverted to using morphology to recover tumor edge features that might have been lost during the segmentation process. Finally, we compared the results obtained visually by superimposing the mask obtained on the original image, thus providing the verification of the tumor region in the two images.

**Removal of the Cranial Region**

When analyzing the original images, we noticed that the intensity of the pixels in the skull region are very similar to the corresponding pixels of the tumor. Therefore, this information interferes in the tumor segmentation process, as the algorithm identifies these two intensities as being the same object. Thus, for the step of removing the skull region, Fig. 4, we use the artifice of mathematical morphology as a technique to remove the interference region, applying the morphological erosion operation.
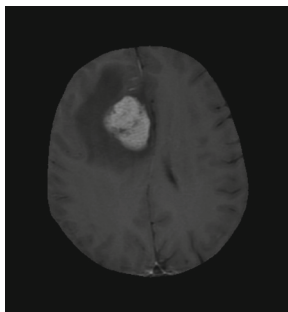
**Fig. 4.** Image after removal of the cranial region.

**Segmentation by K-Means**

Segmentation is one of the most complex processes in digital image processing, becoming the determining factor for the success of the final system. This tool aims to separate image regions that have similar characteristics, by grouping pixels based on the analysis of basic image attributes. We can delimit this technique into two subdivisions, automatic or semi-automatic segmentations. For this work, we will use automatic segmentation techniques, that is, without external interference in the determination of the regions of interest [8].

The process used for segmentation was k-means, a clustering technique well known for its easy implementation. The selection of objects is based on metrics that analyze each pixel in relation to the distance from a centroid, which in turn are recalculated at each iteration until reaching the convergence criterion. Thus, at the end of the execution we will have the definition of subgroups of the original image, Fig. 5 [9].
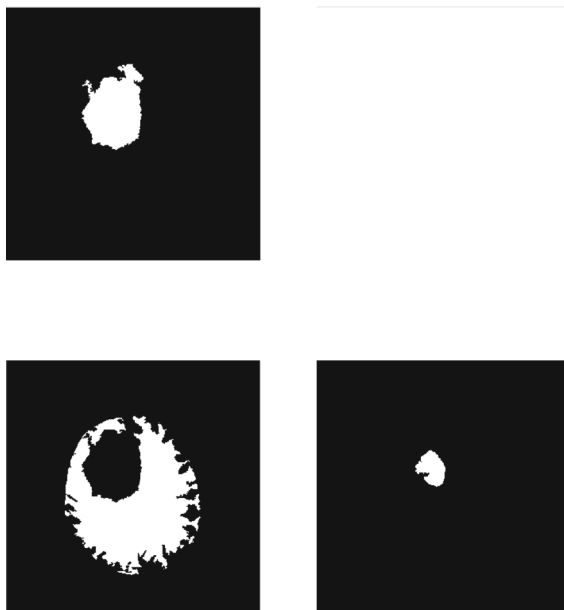


**Fig. 5.** Clusters obtained from the segmentation by k-means.

**Morphological Operation**

Morphological operations can be used in several context within image processing: both as information extraction tools to describe a given region, and applied in the pre- and post-processing stages, as is the case with our approach.

In the pre-processing, we use erosion, a morphological filtering technique that results in the reduction or thinning of objects based on the predetermined structuring element, in this case, we take a 30 radius disk [10] as the structuring element.

In post-processing, we apply the closure, Fig. 6. This method tends to smooth the image by filling the discontinuities, consequently closing the holes present in the segmentation result. As a structuring element, a disk with radius 5 was used so that the corrections were as minimal as possible, without adding information on the tumor and changing the correct dimensions of the tumor region [10].



**Fig. 6.** Mask of the segmented image after morphological closing operation.

**Results Comparison Process**

After applying the morphological corrections to the mask of the segmented image, we performed the superposition of this region with the original image, obtaining the result shown in Fig. 7.
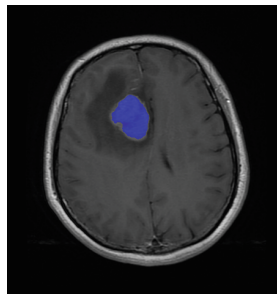


**Fig. 7.** Final result: superposition of the segmented tumor region over the original image.

# 3   Results

To evaluate the results, 300 images were analyzed out of a total of 708 that presented a meningioma-type tumor. We then classified the results into three subgroups: segmentation of the correct region with a complete and well-defined tumor (see Fig. 8); segmentation of the correct region, but with some artifact or lack of information at the end of the tumor (see Fig. 9); and segmentation of the incorrect region (see Fig. 10).

For completely correct results, the responses characterized the entire tumor and presented well-defined borders.



(a)                              (b)
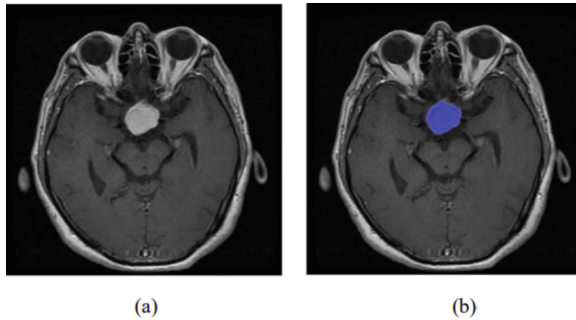
**Fig. 8.** (a) Original images; (b) Segmentation of the correct region, with a complete and well-defined tumor.

In some results, we get a partially correct segmentation. In these cases, the correct region was included in the response, but this region was bypassed, usually to an adjacent cranial region that was not completely removed in the process of removing the cranial part.



(a)                              (b)
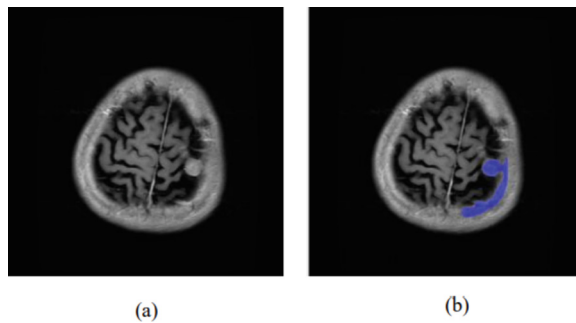
**Fig. 9.** (a) Original image; (b) Segmentation of the correct region, but with the presence of an artifact.

In some situations, the results were wrong, targeting a region different from the one of interest. This problem happened in most cases for slices that were in the eye region.

(a)                                    (b)
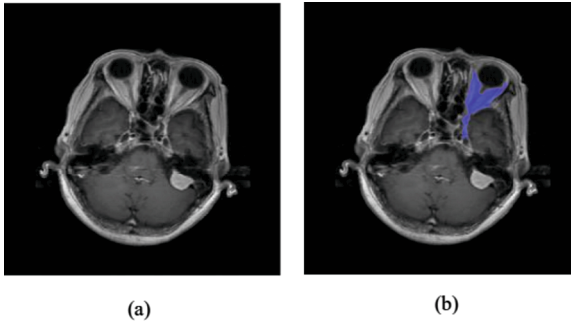
**Fig. 10.** (a) Original image; (b) Incorrect segmentation.

The results obtained numerically, Table 1, indicate that 66% of the images were segmented in the characteristic region of the tumor, with 49.67% referring to the complete and well-defined tumor, while 16.33% with the presence of artifacts or missing information on the tumor ends. Consequently, only 34% of the images presented errors in the segmentation region, that is, they did not correspond to the exact location of the pathology.

**Table 1.** Analysis of algorithm results in magnetic resonance imaging of meningiomas.

| Segmentation results | Amount | Percentage |
|---|---|---|
| Correct region | 149 | 49.67% |
| Correct region, but with some kind of flaw | 49 | 16.33% |
| Incorrect region | 102 | 34% |
| Total | 300 | 100% |

## 4   Discussion

Analyzing the results obtained, it was observed that the algorithmic strategy developed presented important results for the two types of anatomical cuts present in the image exams in the database. However, the responses were more satisfactory for the transverse sections when compared to the coronal ones, this may be due to the amount of information from the two images, since the coronal section presents many artifacts related to the neck region and cranial base.

Other conclusions associated with our tests are precisely in the evaluation of slices that are located very close to the eye region or that have a very thick braincase, that is, those that the first stage of skull removal were not as effective due to their thickness. These flaws must be circumvented, since both additional information and the lack of precision in the size of the lesion can lead to false diagnoses and impair the patient's treatment.

In the cases analyzed by the authors in [11] and [12] with a similar strategy, the techniques of filtering, k-means, thresholding and CNN were used, which obtained the correct percentages of 89.23% and 90.03%, respectively. In the work developed in [5] for computed tomography images, the authors used the techniques of edge recognition filters, morphological operations and segmentation, reaching an accuracy of 94.23%.

In the previous situations, when comparing the results of each algorithm with the 66% obtained in this work, they seem to be significantly better results, however, we must consider the difference between the processing techniques and the amount of images applied for testing. Another important factor to be evaluated was the diversity of slice planes for the resonance images used in this study, which directly interfered with our results and suggest future improvements in the implemented algorithm, in order to encompass the greatest variability of possible cases.

However, in general, the results were interesting, since the algorithm was able to accurately pinpoint the location of the tumor in more than half of the images, needing only a few adjustments to completely hit the 16.33% that included some artifact or ignored some pixels from the edge of the tumor.

## 5    Conclusion

The results obtained demonstrate that the system presents satisfactory results in the diversity of images to which it is submitted, correctly segmenting the region of interest and with a well-defined delimitation of the ends for most cases.

However, in situations with great diversity of information, mainly in the region of the eyes and cuts that cover the neck and cranial base, the system does not present a good performance. In these cases, the segmentation occurs in the correct positions, however, they suffer interference with the addition of artifacts or loss of pixels referring to the edges of the tumor. In these circumstances, the system needs to undergo a review to refine the segmentation, due to the need for precision in the results in order to avoid errors in the diagnosis or inconsistency in the regions presented.

As future work, the algorithm can be improved in order to improve accuracy and reduce error cases. Use of the deep learning technique, as well as the qualitative evaluation of the results by experts in the area.

**Conflict of Interests.**   The authors declare no conflict of interest.

## References

1. Mendes, G.A., Ongaratti, B.R., Pereira-Lima, J.F.S.: Epidemiologia de uma série de tumores primários do Sistema nervosa central [Epidemiology of a series of primary tumors of the central nervous system]. Arq. Bras. Neurocir. **33**(4), 279–83 (2014). (in Portuguese)

2. Brasil: Ministério da Saúde. Secretaria de Atenção à Saúde. Consulta pública n° 30, de 19 de Agosto de 2010 [Public consultation n° 30, of August 19, 2010]. (in Portuguese). https://bvsms.saude.gov.br/bvs/saudelegis/sas/2010/cop0030_19_08_2010.html. Acesso em 17 de junho de 2022

3. Hage, M.C.F.N.S., Iwasaki, M.: Imagem por ressonância magnética: princípios básicos [Magnetic resonance imaging: basic principles]. Ciência Rural **39**(4), 1287–1295 (2009). (in Portuguese)

4. de Fátima Viana Vasco Aragão, M.: A ressonância magnética em tumores astrocitários: avaliação da associação de padrão existente com a graduação histopatológica [Magnetic resonance imaging in astrocytic tumors: assessment of the association of existing pattern with histopathological grading]. 2010. 140 f. Tese (doutorado) – Universidade Federal de Pernambuco. CCS. Neuropsiquiatria e Ciências do Comportamento (2010). (in Portuguese)

5. Soni, A., Rai, A.: CT scan based brain tumor recognition and extraction using Prewitt and morphological dilation. In: International Conference on Computer Communication and Informatics (ICCCI – 2021), Coimbatore, India (2021)

6. Instituto Nacional de Câncer (INCA): Câncer do Sistema Nervoso Central [Cancer of the Central Nervous System]. 25 de Abril de 2022 (in Portuguese). https://www.inca.gov.br/tipos-de-cancer/cancer-do-sitema-nervoso-central. Acesso em 17 de junho de 2022

7. Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., et al.: Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLoS ONE **10**(10), e0140381 (2015). https://doi.org/10.1371/journal.pone.0140381

8. da Silva, A.M.M., Patrocínio, A.C., Schiabel, H.: Processamento e Análise de Imagens Médicas [Processing and analysis of medical images]. Revista Brasileira de Física Médica **13**(1), 34–48 (2019). (in Portuguese)

9. Júnior, D.E.M.: Utilização do algoritmos k-means na segmentação e classificação de imagens de refeições [Use k-means algorithms in the segmentation and classification of images of meals]. 2015. 75 f. Monografia – Universidade Federal de Alfenas, Curso de Bacharelado em Ciências da Computação (2015). (in Portuguese)

10. Gonzales, R.C.: Processamento digital de imagens [Digital image processing. In: Gonzales, R.C., Woods, R.C. (eds.) Revisão técnica: Marcelo Vieira e Maurício Escarpinati; [tradução Cristina Yamagami e Leonardo Piamonte], 3rd edn. Pearson Prentice Hall, São Paulo (2010). ( in Portuguese)

11. Mascarenha, L.R., Júnior, A.d.S.R., Ramos, R.P.: Segmentação automática de tumores cerebrais em imagens de ressonância magnética [Automatic segmentation of brain tumors in MRI images]. Einstein (São Paulo) (2020). (in Portuguese). https://doi.org/10.31744/einstein_journal/2020AO4948

12. Khan, A.R., Khan, S., Harouni, M., Abbasi, R., Iqbal, S., Mehmood, Z.: Brain tumor segmentation using k-means clustering and deep learning with synthetic data augmentation for classification. Microsc. Res. Techn. (2021). https://doi.org/10.1002/jemt.23694

# Mobile App for Assessing Hemifacial Spasm Treatment Response Using Machine Learning

J. L. S. da Silva[1] , C. M. G. de Godoy[1] , T. H. Osaki[2] , M. H. Osaki[2] ,
C. Yabumoto[2] , and R. C. Coelho[1(✉)]

[1] Science and Technology Institute, Federal University of São Paulo, São José dos Campos,
SP 12231-280, Brazil
rccoelho@unifesp.br

[2] Department of Ophthalmology and Visual Sciences, Paulista School of Medicine, Federal
University of São Paulo, São Paulo, SP, Brazil

**Abstract.** It is challenging to assess hemifacial spasm (HFS) patients as they exhibit high-frequency and heterogeneous anomalous eyelid movements. This study aimed to develop an application for a smartphone to objectively determine eyelid movements frequency so that treatment responses in these patients can be assessed accurately. The smartphone application was developed mainly using Python, a prominent and broadly used programming language focused on machine learning and data science tasks. The application can precisely predict the movement of the patient's eyes using an SVM regressor and classifier. The results are plotted for better visual inspection by using data visualization techniques. Thus, the application enables a continuous study of each patient using an integrated database in Google spreadsheets, which could better track the results of each treatment response. The application showed to be an efficient method to identify and represent eyelid movement occurrences in patients, objectively measuring the eyelid movement frequency and, thus, assessing the treatment response in patients with hemifacial spasms. This system could enable customized and fine adjustments to botulinum toxin doses based on each patient's needs.

**Keywords:** Machine Learning · Eye Aspect Ratio · Hemifacial Spasm

## 1 Introduction

Hemifacial spasm (HFS) is a neurological condition characterized by involuntary, tonic, and clonic spasms of the muscles innervated by the ipsilateral facial nerve. This condition is usually secondary to facial nerve compression at the root exit zone caused by an aberrant artery [1, 2]. Spasms usually originate in the periocular region (orbicularis oculi muscle) before affecting the midface and lower third of the face over months to years. Most cases present with an affected and a non-affected side in this condition [1]. Botulinum toxin-A (BTX-A) injections in the affected muscles are considered the treatment of choice for this condition [1–4], reducing facial spasms. The efficacy of botulinum toxin administration is generally assessed using clinical rating scales, such

as the Jankovic rating scale [3, 4] Although grading systems facilitate the classification of clinical symptoms, these tools are not accurate in assessing treatment response [5]. Furthermore, in HFS, it is challenging to impartially assess the eyelid spasms due to high-frequency anomalous eyelid movements [6]. Previously employed objective approaches to assess this condition relied on less accurate systems based on manually reviewing videotapes or other indirect approaches [7, 8]. An accurate system has been described to assess this condition's eyelid movements fairly (spontaneous blink + anomalous eyelid spasms) [5]. However, that system is complex and cannot apply in clinical practice.

This study aimed to develop an application for a smartphone to objectively determine the eyelid movements frequency so that the therapeutic effect of botulinum toxin can be practically assessed in patients with HFS. The app improves existing methods as it will be handy for evaluating these patients' response rates to treatment. To the best of our knowledge, no previous smartphone application has been reported for this purpose.

## 2 Materials and Methods

### 2.1 Libraries and Data Sources

The data was collected in 2-min videos from subjects who agreed to participate in the study. Patients with HFS and normal subjects from the Division of Ophthalmic Plastic Surgery, Department of Ophthalmology and Visual Sciences, Federal University of Sâo Paulo, were recruited to participate. Python manipulated and extracted information from the videos. It is a commonly used programming language for machine learning and data science in general [9]. Python contains some libraries used to build the application: OpenCV, Imutils, Dlib, Matplotlib, Scipy, Streamlit, gspread, secrets, datetime, Pillow and Seaborn. Each library had a different use. OpenCV was responsible for interacting between the webcam and the script. It also applied to image transformation (thresholds and resizing) and frame manipulation. Imutils applied for manipulating the data points that corresponded to the landmarks. Dlib contained the model that detected the eye points based on the 68 landmarks (with HOG [10] and SVM [11]). Using Scipy, we calculated the area within the eye, and with Matplotlib the data was plotted to visualize the results of the tests.

The model was trained on the public online database Ibug 300W ICCV 2013 [12], which was part of a face identification challenge. It contains four different datasets: Helen, LFPW, AFW, and IBUG, corresponding to more than 11.500 images and their labels. These datasets were also manipulated to reduce the size of the XML files. It reduced from 68 to 12 points of data for each image (only the eye points were preserved). Using Streamlit library, we developed the web application that contains the machine learning model. That allows any person to use the application with a smartphone. Gspread library established a connection between the app and Google Sheets, where all the data collected was stored, as is an easy access platform enabled by Google to create tables that attend to the needs of this task. Secrets is a built-in library from Python used to create a unique ID for each patient to relate the registration database with the data collection database. Datetime library allowed to retrieve the day the data was collected. Pillow was used to opening the main image. Finally, Seaborn was used alongside Matplotlib to plot the line charts that describe the variation of the eye's area over time.

## 2.2  Machine Learning Script Structure

The script, built-in Python, is based on the algorithm of Rosebrock [13]. In his work, he utilizes a machine learning model which relies on the Histogram of Oriented Gradients (HOG). HOG is a descriptor that extracts features from an image based on a convolutional operation between kernels and the image. It models the resulting data in a concatenated vector of components of the normalized cell histograms from all the captured block regions. Such extractor is commonly used with Support Vector Machine (SVM), a supervised learning model that classifies and makes predictions based on regression analysis. Thus, the data points are located in the image after finding the features that result from the HOG extractor. The face detector is available by the *Dlib* library and is used to identify the 68 landmarks in videos or the webcam.

Based on that, the model was adapted to segment only the eye as a region of interest. As previously said, the XML data was manipulated so that only the 12 eye points were focused. Therefore, the data is collected in the form of lists, so it is reproducible as line plots to identify the patient eye aspect ratio over time.

This script structure is the main part of the developed app. It provides a friendly user interface, as it can be used on mobile and desktop devices without accessing any background code.

## 2.3  Eye Aspect Ratio Function and Measurement

The eye aspect ratio calculus was inspired by an article that focuses on the results of the SVM model in predicting facial landmarks [14]. It uses the Euclidean distance of the height and width of the eye to calculate its area, as shown in Eq. 1. This Eq. Refers to the points of the eyes from $p_1$ to $p_6$ in Fig. 1.

$$EAR = \frac{||p_2 - p_6|| + ||p_3 - p_5||}{2 * ||p_1 - p_4||} \tag{1}$$

The results are primarily constant with slight variations when the eyes are open. That is a very accurate method to classify between open and closed eyes, as the *SVM* model had an excellent performance in classifying the landmarks in many situations. This calculus is performed for both eyes as all types of blepharospasm disease can be measured.

The resulting values accumulate in a Python list so that it can be plotted into a line chart containing, for each eye, the eye aspect ratio over time. Figure 2 is an example of how the results are shown.

A threshold was based on the average size of the person's eye to measure the EAR as an eye blink. For most cases, EAR rounds about 0.25, or 0.3. However, it depends on the eye size; thus, it needs to be fine-tuned. As for the fine-tuning, it was defined a function that retrieves a modified EAR to keep track of each patient threshold. The modified EAR, based mainly on the actual size of the eye [14], is calculated as the sum between the open eye size and the closed eye size, as can be seen in Eqs. 2 to 4.

$$EAR_{closed} = \frac{||p_2 - p_6||_{min} + ||p_3 - p_5||_{min}}{2 * ||p_1 - p_4||_{max}} \tag{2}$$
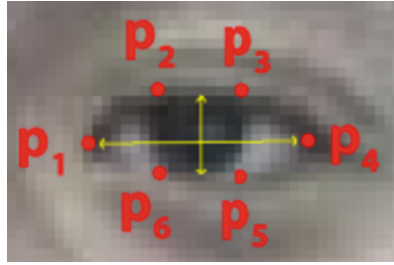
**Fig. 1.** Points of the eye used in Eq. 1.
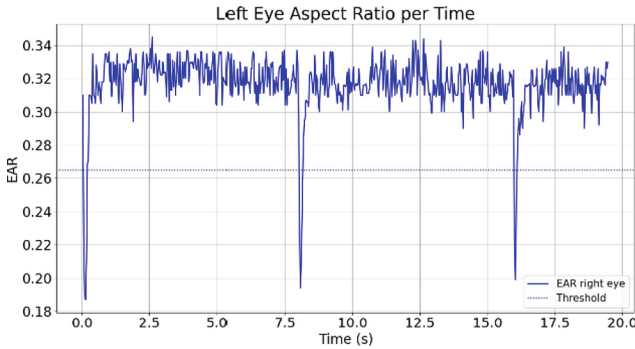


**Fig. 2.** Example of a line chart of an eye aspect ratio (EAR) over time.

$$EAR_{open} = \frac{||p_2 - p_6||_{max} + ||p_3 - p_5||_{max}}{2 * ||p_1 - p_4||_{min}} \tag{3}$$

$$Modified_{EAR_{Thresh}} = (EAR_{Open} + EAR_{Closed})/2 \tag{4}$$

By using those equations, we obtain the proper EAR of the patient, which is calculated before each further analysis. After that, another counter starts after the value of the EAR is below the threshold. If the EAR does not rise for one frame, it is considered only one short blink and is not counted. Therefore, it avoids the risks of getting false negative values.

## 2.4   Streamlit Application

The app was based on the Streamlit, a recently released library containing features to develop online web applications for mobile and desktop services. A user interface was created and divided into three sections: the home page, the analysis/login page, and the sign-up page. The opening page summarizes the project and how to use the application. Figure 3 illustrates the app's initial page.

The analysis/login page (Fig. 4) leads to the app's main functionality, which is the eye recognition and evaluation using the Dlib facial landmarks detector. It contains a separator between pre and post-treatment evaluation and a history section. In that section,
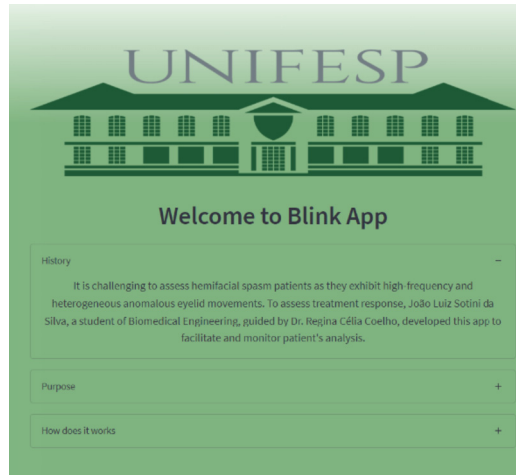
**Fig. 3.** The app interface.

it is possible to visualize the patient's records as a data frame. In the Charts section, the patient or physician can select and visualize specific patient records plots. Finally, the sign-up page is where the user, a patient, or a doctor can register in the app to use its resources and features.

## 2.5   Google Sheets Database

As the app relies on retrieving patients' data for further evaluation of the symptoms and severity of spasms, it is inherent to store and relate this data to each patient who uses its services. Therefore, a simple and efficient solution was to create a Python automated Google Sheet that is updated whenever the user wants to store his data records. The library Gspread was used to access the Sheet called 'Person', which contains the main working sheet called 'Registration' and the ID-related sheets, which will be explained later. The 'Registration' working sheet contains the personal data, including login name, surname, city of residence, age, password, and ID. The user defines the login name and the password in the sign-up steps. The ID is an automatically generated random unique key. When the patient registers an account, the script also generates a new working sheet related to the user's ID (its name is the user's ID), which is exclusive to every new user who registers. It contains information about the date that the data was collected, the eye aspect ratio collected from the video from each eye (left and right eyes), the number of blinks that the person had during the analysis, the stage of the analysis (pre or post-treatment) and the threshold that indicates whether the patient has blinked or not.

**Fig. 4.** The analysis page.

## 3   Validation

To validate the efficiency of the application in recognizing blinks, the control group had their total number of blinks annotated for 2 min. After this, the results were stored in a Google spreadsheet and compared to the application results using the root mean squared error (RMSE), as the results are integers derived from a regression task. The results were normalized using the functions StandardScaler, and the evaluation metric used was 'mean_squared_error' [15], both from Scikit-Learn [11]. The results are between 0.0 and 1.0. If the RMSE results are closer to 0.0, the model predicts effectively.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (y - \hat{y})^2} \tag{5}$$

where $\hat{y}$ is a vector with the predicted values, and y is the ground truth value for each value analyzed. The result is the root of the difference between the predicted and the ground truth values divided by the number of values (value represented by n).

## 4   Results and Discussion

Figure 5 illustrates a frame of a patient video and the predictions computed by the developed app. Figures 6 and 7 show the charts plotted inside the app concerning the same patient's eye aspect ratio to her pretreatment and post-treatment.
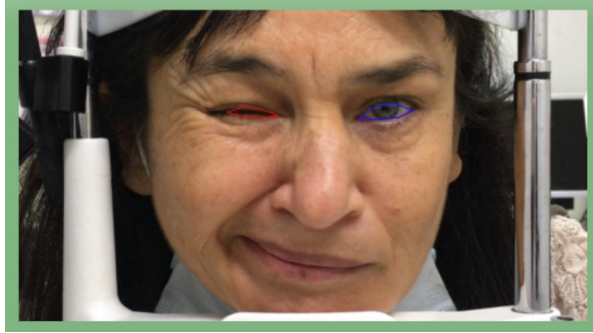
**Fig. 5.** A patient video frame used in the app. The red and blue lines around the eyes are the predictions being computed in real-time.
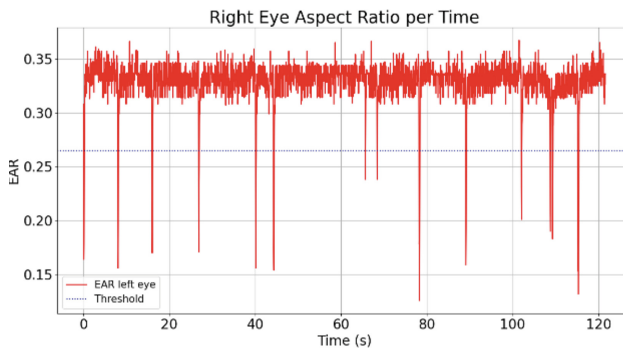


**Fig. 6.** Line plot of the EAR (Eye Aspect Ratio) over time and the threshold of a pretreatment patient's right eye (patient is depicted in Fig. 5).
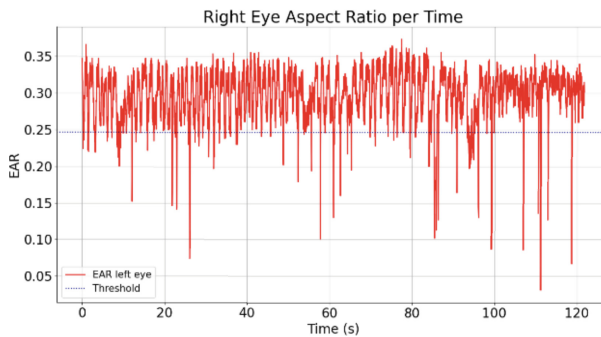


**Fig. 7.** Line plot of the EAR (Eye Aspect Ratio) over time of post-treatment patient's right eye (patient depicted in Fig. 5).

The results were compared to the control group's ground truth annotations (made by physicians). The results (Table 1) show that the mean RMSE of the normalized predictions compared to the normalized gold standard was, on an average, 0.0995. Concerning this mean RMSE outcome, it is possible to realize that the model succeeded in predicting the blinks of the control subjects.

**Table 1.** Annotations and Result

| Control Patient | Annotated Blinks (Dr. M. H. Osaki) | Annotated Blinks (Dr. T. H. Osaki) | Predicted Blinks |
|---|---|---|---|
| 1 | 129 | 127 | 127 |
| 2 | 58 | 58 | 58 |
| 3 | 18 | 18 | 14 |
| 4 | 97 | 97 | 97 |
| 5 | 69 | 68 | 70 |
| 6 | 39 | 40 | 36 |
| 7 | 69 | 70 | 65 |
| 8 | 65 | 63 | 56 |
| 9 | 58 | 58 | 58 |
| 10 | 33 | 32 | 23 |

The app can predict most blinks, although the EAR threshold is hard to define in every situation, considering that some patients had semi-blinks, which can configure as blinks in medical terms. However, it is not accounted for by the calculation, as it mainly measures the fully closed eyes. Nevertheless, that is not a problem, as the eyelid movement is accounted for in the chart, being possible to observe the half-blinks (see control patient number 10 in Table 1). Another advantage of the app developed in the present study is its high accuracy in detecting movement. This advantage is not possible with conventional human-based video review assessment. This app could enable more accurate and customized dose adjustments for each patient due to its accuracy in detecting eyelid movement. The drug duration is another outcome used to assess the effect of botulinum toxins. The objective duration associated with BTX-A treatment has been assessed indirectly in previous studies that analyzed changes in eyelid morphometric patterns in HFS patients over four months [8]. Patients were evaluated at baseline and at the 15-day and 2-, 3-, and 4-month time points. The return of the parameters to their pretreatment status at the 4-month time point implies that the studied period during which these changes took place represents the BTX-A duration. Similarly, the smartphone app presented herein could indirectly facilitate the unbiased assessment of the BTX-A duration by evaluating the time required for the eyelid movement frequency to return to its baseline levels.

## 5   Compliance with Ethical Requirements

Federal University of São Paulo Review Board approved this study (CAAE89528618.4.0000.5505), and all patients were treated according to the Declaration of Helsinki.

## 6   Conclusions

The app presented in this work is an efficient resource for identifying and representing eyelid movement occurrences in patients. The easy access and facilitated usability platform improve the physicians' visualization and control of hemifacial spasm disease. That allows objectively assessing treatment response, as it delivers high accuracy metrics and charts based on the uploaded videos. It also contributes to accurately monitoring patients over time as it allows doctors to observe the disease patterns more easily (currently, there is no practical method for this purpose). This system could enable customized and fine adjustments to botulinum toxin doses based on each patient's needs.

**Conflict of Interest.**   The authors declare that they have no conflict of interest.

## References

1. Ross, A.H., Elston, J.S., Marion, M.H., et al.: Review and update of involuntary facial movement disorders presenting in the ophthalmological setting. Surv. Ophthalmol. **56**, 54–67 (2011)
2. Green, K.E., Rastall, D., Eggenberger, E.: Treatment of blepharospasm/hemifacial spasm. Curr. Treat. Options Neurol. **19**(11), 41–55 (2017)
3. Kenney, C., Jankovic, J.: Botulinum toxin in the treatment of blepharospasm and hemifacial spasm. J. Neural Transm. **115**(4), 585–591 (2008)
4. Bilyk, J.R., Yen, M.T., Bradley, E.A., et al.: Chemodenervation for the treatment of facial dystonia: a report by the american academy of ophthalmology. Ophthalmology **125**(9), 1459–1467 (2018)
5. Osaki, M.H., et al.: An objective tool to measure the effect of botulinum toxin in blepharospasm and hemifacial spasm. Eur. J. Neurol. **27**(8), 1487–1492 (2020)
6. Osaki, M.H., et al.: Analysis of blink activity and anomalous eyelid movements in patients with hemifacial spasm. Graefes Arch. Clin. Exp. Ophthalmol. **258**(3), 669–674 (2020)
7. Yoshimura, D.M., Aminoff, M.J., Tami, T.A., et al.: Treatment of hemifacial spasm with botulinum toxin. Muscle Nerve **15**(9), 1045–1049 (1992)
8. Osaki, T., Osaki, M.H., Osaki, T.H., et al.: Influence of involuntary eyelid spasms on eyelid morphometric changes in patients with hemifacial spasm. Br. J. Ophthalmol. **100**(7), 963–970 (2016)
9. Van Rossum, G., Drake, F.L.: Python 3 Reference Manual. CreateSpace, Scotts Valley (2009)
10. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
11. Sagonas, C., Antonakos, E., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: 300 faces In-the-wild challenge: database and results. In: Image and Vision Computing (IMAVIS), Special Issue on Facial Landmark Localisation "In-The-Wild", vol. 27, pp. 3–18 (2016)

12. Rosebrock, A.: Facial landmarks with dlib, OpenCV, and Python (2017). https://www.pyimag esearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/
13. Soukupová, T., Cech, J.: Real-Time Eye Blink Detection using Facial Landmarks. Center for Machine Perception, Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University in Prague (2016). http://vision.fe.uni-lj.si/cvww2016/proceedings/pap ers/05.pdf
14. Dewi, C., Chen, R.-C., Jiang, X., Yu, H.: Adjusting eye aspect ratio for strong eye blink detection based on facial landmarks. PeerJ Comput. Sci. **8**, e943 (2022). https://doi.org/10. 7717/peerj-cs.943

# Machine Learning for the Classification of Surgical Patients in Orthodontics

Carlos Andrés Ferro-Sánchez[1] , Christian Orlando Díaz-Laverde[2] ,
Victor Romero-Cano[1] , Oscar Campo[1] ,
and Andrés Mauricio González-Vargas[1(✉)]

[1] Faculty of Engineering, Universidad Autónoma de Occidente, Cali, Colombia
amgonzalezv@uao.edu.co
[2] Health Faculty, Universidad del Valle, Cali, Colombia

**Abstract.** Dentofacial anomalies, also known as malocclusions, are alterations with a congenital, traumatic, or growth origin. These anomalies can generate functional and aesthetic problems in those who suffer from them and have been reported by the World Health Organization as the third most prevalent oral disease. The most commonly used methods for correcting these anomalies are orthodontics and orthognathic surgery. The diagnosis, and the correct selection of the treatment to be carried out, are part of an extensive process that involves collecting different cephalometric and clinical data, and depend on the clinician's experience. Therefore, no standardized process allows the classification or diagnosis among patients who achieve the best result with orthodontics, that is, non-surgical procedures or if surgical intervention is necessary. This study aims to propose a digital tool based on machine learning algorithms that may help the clinician to select an orthodontics or surgical treatment for patients who are about to start their treatment.

**Keywords:** Orthodontic · Machine learning · Malocclusion · Surgical · Cephalometric · Classification

## 1 Introduction

*Malocclusion* is an anomaly that is characterized by the alteration of craniofacial growth, or the presence of a poor relationship or misalignment between the upper and lower dental arches concerning the transverse or vertical antero-posterior planes [1], which can generate functional problems, aesthetic and psychosocial, and affect social development or emotional wellbeing in both children and adults [2]

Currently, the most widely used methods for correcting malocclusions are orthodontic treatments, sometimes combined with orthognathic surgery [3], depending on the severity of the malocclusion and its classification. Edward Angle proposed a classification of malocclusions based on the anteroposterior relationship of the upper and lower buccal segments [4]:

- Class I: A normal anteroposterior relationship prevails.
- Class II: characterized by mandibular retrognathism and maxillary prognathism. This classification has two subdivisions:
  - Excessive distance (overjet) between the upper and lower incisors.
  - Existence of retroclination of maxillary central incisors.
- Class III: Characterized by prognathism of the mandibular segment.

Since 1989 [5], the World Health Organization (WHO) has reported malocclusions as the third most prevalent oral disease, after caries and periodontal disease [6]; likewise, it affects about 50% of the world population [7]. Recent studies reported the global prevalence of malocclusion in permanent dentition in class I at 74.7%, class II at 19.56%, class III at 5.93%, deep bite at 21.98%, open bite at 3.97% and posterior crossbite at 9.39% [8]. In Latin America, the Pan American Health Organization reported a prevalence and incidence of malocclusions greater than 85% in the population [9].

Diagnosis or problem definition and treatment planning are the most important steps in the correction of malocclusions. Unfortunately, bite correction does not always lead to correction of facial esthetics, and sometimes, facial imbalances occur in the desire to correct the bite [10,11]. For this reason, if not done correctly, it can end up in extensive treatments that can generate repercussions such as root resorption and increased sensitivity to pain, in addition to affecting esthetics. Therefore, if the diagnostic results indicate that the patient's desired results are not achievable with orthodontic treatment alone, orthognathic surgery or a combination of both should be considered as a therapeutic method [12], as is the case in borderline patients for whom orthodontic treatment is chosen due to cultural, esthetic and financial conditions, even if they are surgical cases [13].

Decision making, or classification of patients according to the required treatment, is an exhaustive process that requires the organization of different diagnostic data, prior knowledge, and experience of the clinician. Therefore, there is no standardized way of doing it. If we could have a tool based on artificial intelligence that helps decision making, it could help the workflow of orthodontists with a high casuistry, as well as help those who still do not have a lot of experience. Currently, some expert algorithms have been developed, capable of reproducing an expert's classification or decision making capacity [12,14] with an accuracy of 96% [12]. In another study to create a layered system for classifying malocclusions, according to Angle's classification, authors used logistic regression, K-nearest neighbors, random forests, and Bayesian classifiers, and attained accuracies of 88.89%, 83.33%, 88.89%, and 55.66%, respectively [15]. However, no previous study currently allows the classification between surgical and non-surgical patients in the Colombian population.

This study evaluates the feasibility of obtaining the classification of surgical and non-surgical patients with a sample of Colombian patients who are in the process of diagnosis or undergoing orthodontic treatment. In addition, machine learning algorithms focused on binary classification are proposed. First, we present the methodology for acquiring and classifying patients. Second, the process of cleaning and analyzing the data used to train the machine learning

algorithms used in the study is presented. Finally, we discuss the results and present some conclusions about the work.

## 2    Materials and Methods

### 2.1    A. Data Acquisition

The cephalometric data necessary for this study were obtained from 104 cephalometries taken sequentially and chronologically in a dental office in Cali, Colombia. The location of the craniometric and cephalometric points was done manually in the NTN viewer software. These are calculated according to the bony structures of the skull of the patients. This procedure was performed with the help of 2 experienced orthodontists and the process was verified again 4 days later with the same operators until an adequate Cohen's Kappa coefficient was obtained, in this process the patients for whom there were discrepancies between the two specialists were eliminated, therefore the final sample was 86 patients. The cephalometric measurements consisted of 8 angular and 8 linear for a total of 16 measurements that provide sufficient information to determine the dental and skeletal characteristics that are the object of this study (Fig. 1). The measurements used for the characterization of the radiographs are described below.

- Linear
    - Overbite: Vertical overlap of teeth, measured between the upper and lower incisal edges.
    - Overjet: Horizontal overlapping of the teeth, measured from the incisal edge of the upper incisor to the buccal surface of the lower incisor.
    - Spee curve (depth): Occlusal curvature was observed in the sagittal view of the lower arch; for this study, the depth was measured with respect to the occlusal plane.
    - U1-NA: Relationship of the maxillary central incisor with the reference line N-A. Distance from the labial surface of the incisor anterior to the N-A line.
    - L1-NB: Relationship of the mandibular central incisor with the reference line N-B. Distance from the labial surface of the incisor anterior to the N-B line.
    - UL-EP: Distance from the upper lip to the E line traced between the E and pogonion (Pg) points of soft tissues.
    - LL-EP: Distance from the lower lip to the E line drawn between the E points and the soft tissue pogonion.
    - L1-APg: Relationship of the mandibular incisor concerning the line between point A and pogonion.

- Angular:
    - IMPA: Angle formed between the lower incisor and the mandibular plane.
    - Upper incisor to palatal plane (UIPP)

- FMIA: Frankfort to lower incisor.
- FMA: Angle formed between the Frankfort plane and the mandibular plane.
- SNA: Angle formed between the saddle-nasion points (N) and point A, which refers to the maxilla's horizontal position with respect to the skull's base.
- SNB: Angle formed between the saddle-nasion points and point B, which refers to the mandible's horizontal position with respect to the skull's base.
- U1-NA: Relationship of the maxillary central incisor with the reference line N-A. The inclination of the axis of the maxillary incisor.
- L1-NB: Relationship of the mandibular central incisor with the reference line N-B. Mandibular incisor axis inclination.

These samples went through an anonymization process where personal data was eliminated. Later, they were classified with the help of an orthodontist with more than ten years of experience, who labeled them as surgical and non-surgical, distributed in 49 samples for surgical and 37 for non-surgical.
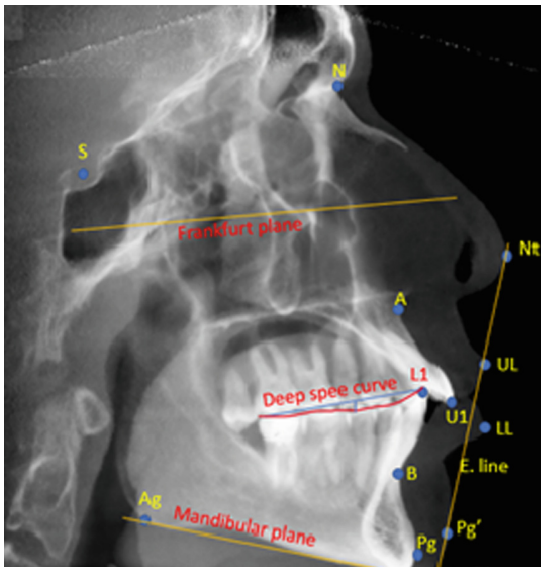


**Fig. 1.** Location of cephalometric marks.

## 2.2   Data Preprocessing

Once the data set was collected and labeled, it was analyzed. It was observed that there were missing data for some angular and linear measurements. We also

found an imbalance in the classes. Because it is a small database, eliminating the samples with missing data is not recommended since this may sacrifice the representativeness of the available data. Therefore, a data imputation process was carried out using the KNNImputter model from the *sklearn* library, which allows imputation to complete the missing values using a K-nearest neighbors-based methodology, avoiding altering the normal distribution of the data. For this step, it was necessary to divide the database into a training set (80%) used to train and validate using a 5-fold cross-validation, and a testing set (20%), using the train_test_split function of the sklearn library to prevent the model from knowing all the data during the test. It was identified that much of the missing data belonged to the surgical class, and there were only two samples of the non-surgical class with missing data in the characteristic overbite. The missing data were located in the characteristics UIPP, Overbite, Overjet, Spee curve, and L1-APG, all with 27 missing data except for overbite, which presented 29 missing data. Once the data samples were imputed, the first training set was generated.

The training dataset distribution presented and imbalance between surgical (n = 38) and non-surgical (n = 30), which could affect the training of machine learning algorithms. To balance the training set, and because it is a small training set, it was decided to carry out an oversampling, using two techniques applied to copies of the original set, from which two training test datasets were obtained:

– Dataset with random oversampling: this dataset was built using random over-sampling by resampling, which is based on the random selection of examples of the minority class, to which it makes a smooth replacement and adds them to the training set.
– Synthetic Random Minority Oversampling (SMOTE) dataset: This was obtained using a model that works by selecting a random point from the minority class and calculating the K nearest neighbors for the selected point. The newly added data is selected between the selected point and its neighbors.

Once this process was finished, three training sets were obtained, which were used to train and evaluate the different classifier machine learning algorithms.

## 2.3   Classification Models

In order to perform the classification task, different classifier algorithms were used, such as K-Nearest Neighbors (KNN), Support Vector Classifiers (SVC), Logistic Regression (LR), Decision Trees (DT), Random Forests with and without pre pruning (RF), Bayesian classifiers (NB), gradient boosting for classification (GB), and multilayer perceptron (MLP).

For the selection of features, we used the SelectPercentile and f_classif function of the sklearn library. This method uses a univariate statistical test to select the best features according to the requested percentile. The percentiles 5, 25, 50, 75 and 95, were tested for the different models and trained using the pipeline function and GridSearchCV to obtain the features that allowed the best performance for each independent model.

Because the samples of both classes present a high dispersion, the application of scaling was tested using the StandardScaler function, Robust scaler function and the MaxAbsScaler function. This process was configured within a pipeline in the input data processing stage, where the GridserachCV function selects the scaling that gives the best performance for the evaluated model.

For the construction of the models, a group of hyperparameters was tested (Table 1), using GridserchCV to find the values that would achieve the best performance for each model in accuracy, sensitivity, and f1_score metrics.

**Table 1.** Models and hyperparameters

| Model | Algorithm | Hyperparameters |
|---|---|---|
| KNN | KNeighbors Classifier | N_neighbors, weights |
| SVM | SVC | Kernel, gamma, C |
| LG | Logistic Regression | C, penalty, max_iter, solver |
| RF with prepruning | RandomForest Classifier | N_stimators, max_features, max_depth, criterion |
| RF | Random Forest Classifier | N_stimators, max_features, criterion |
| NB | GaussianNB | Var_smoothing |
| DT with prepruning | Decision Tree Classifier | Criterion, Max_depth |
| DT | Decision Tree Classifier | Criterion, min_samples-leaf |
| GB | Gradient Boosting Classifier | N_stimators, max_features, max_depth, criterion |
| MLP | MLP Classifier | Max_iter, activation, hidden_layer_size, solver |

### 2.4   Model Testing

To test the models, the metrics precision, recall, f1 score, and accuracy were used since they are widely common measures to evaluate the performance of binary classifiers [16] in supervised machine learning algorithms. In addition, the metric AUC allows us to evaluate the quality of the classification models and thus choose the best model to use [17].

# 3   Results

Eight classification models were trained with the three different training datasets, one of which is affected by class imbalance, another balanced using random oversampling, and finally, a balanced dataset using SMOTE oversampling. These datasets provide the best combination of the most important characteristics, the need for scaling or not of the data, and the selection of the best combination of hyperparameters for each model.

The test of these models showed that 75% of them presented better performance with the 16 initial features, which were SVM, LG, RF, DT, NB, and GB 25% used all except the linear measure U1-NA (Table 2).

Several of the algorithms provided acceptable performance (Table 3). However, those with the best performance were the decision trees and gradient boosting for classification, which correctly classified all 10 samples out of 11 belonging to the surgical class in the test set. However, in the case of the non-surgical class, all 7 samples were correctly classified (Table 4).

The other algorithms that presented performance of interest were RF, SVM, and MLP, which presented interesting results in most of the metrics, with a greater focus on the AUC metric.
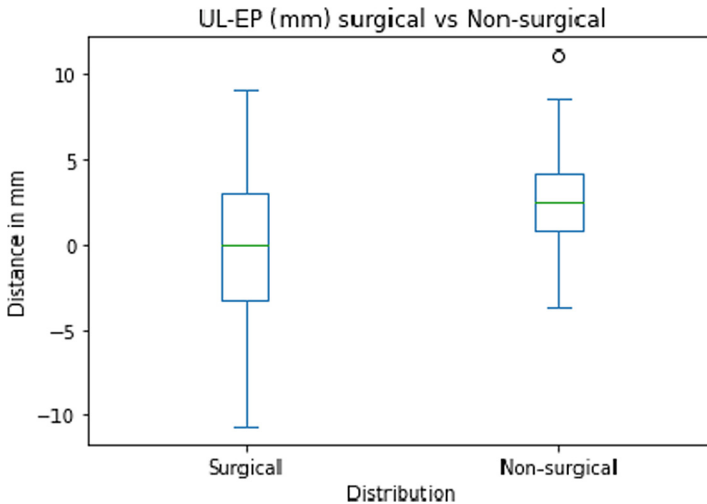


**Fig. 2.** Distribution of UL-EP measurement among classes

**Table 2.** Features and hyperparameters selected

| Model | Best hyperparameters | Percentile features | Features selected | Scaler | Train set |
|---|---|---|---|---|---|
| KNN | n_neighbors = 1, weights = 'uniform' | 95 | All, except lineal measurement U1-NA | Robust Scaler | Random oversampling |
| SVM | C=10, gamma = "Scale" | All features | All features | Robust Scaler | Random oversampling |
| LG | max_iter = 10000, penalty = 'l2', solver = 'lbfgs', C = 1.0 | All features | All features | None | Random oversampling |
| RF | max_depth = 5, max_features = 'None', n_estimators = 250, criterion: 'gini' | All features | All features | None | Random oversampling |
| NB | var_smoothing = 0.001232846 | All features | All features | MaxAbs Scaler | Random oversampling |
| DT | max_depth = 50, criterion = 'gini' | All features | All features | None | Random oversampling |
| GB | criterion = 'friedman_mse', max_depth = 10, max_features = 'auto', n_estimators = 10 | All features | All features | None | Random oversampling |
| MLP | hidden_layer_sizes = (75,), max_iter = 3000, activation = 'tanh', solver = 'adam' | 95 | All, except lineal measurement U1-NA | None | Random oversampling |

**Table 3.** Scores of the metrics of qualification for each model in the test set

| Model | Accuracy (%) | Recall (%) | F1_score (%) | Precision (%) | AUC |
|---|---|---|---|---|---|
| KNN | 66.7 | 85.7 | 66.7 | 54.5 | 0.72 |
| SVM | 88.9 | 85.7 | 85.7 | 85.7 | 0.97 |
| LG | 66.7 | 85.7 | 66.7 | 54.5 | 0.69 |
| RF | 88.9 | 100 | 87.5 | 77.8 | 0.94 |
| NB | 55.6 | 57.1 | 50 | 44.4 | 0.7 |
| DT | 94.4 | 100 | 93.3 | 87.5 | 0.955 |
| GB | 94.4 | 100 | 93.3 | 87.5 | 0.955 |
| MLP | 83.3 | 100 | 82.4 | 0.7 | 0.92 |

**Table 4.** Scores for DT qualification metrics

| | Precision | Recall | f1_score | Support |
|---|---|---|---|---|
| Surgical | 1 | 0.91 | 0.95 | 11 |
| Non-surgical | 0.88 | 1 | 0.93 | 7 |
| Accuracy | | | 0.94 | 18 |
| Macro avg | 0.94 | 0.95 | 0.94 | 18 |
| Weighted avg | 0.95 | 0.94 | 0.94 | 18 |

## 4    Discussion

Although there is currently a system that handles the classification of orthodontic surgical patients with an accuracy of 96% [12], this study had samples obtained from 316 patients of Korean nationality, and with exclusion criteria for missing teeth, malformed teeth, history of orthodontic treatment, skeletal asymmetries and maxillofacial deformities, which benefits the quality of the data. They also used 12 cephalometric measurements, in addition to six other clinical

indices. Our study was performed based on cephalometric measurements taken from Colombian patients, in which there is a wide racial variety [18] which are represented in variations in dental and bone relationships [19,20], which represented a great variation in the cephalometric measurements [21,22], despite, some of the algorithms presented good accuracy and sensitivity (Accu = 94.4% and Recall = 100%) when classifying the test samples. It was possible to observe the importance of measures that relate to the state of the soft tissues, such as the measurement of the upper lip with Ricketts' E-line (UL-EP), where a certain increase in the projection of the upper lip can be glimpsed in surgical patients, that is, the upper lip extends beyond the margin of the E-line, resulting in negative measurements (Fig. 2). To better characterize both classes, it is necessary to acquire a larger data set that allows a better representation of the problem, and in addition to this, the use of more cephalometric measurements that are indicative of the patient's soft tissue status. In future experiments, we plan to collect more samples, apply more complex algorithms, extract more important data in the diagnostic process and even explore the use of methods based on convolutional neural networks [18,21].

## 5 Conclusions

This study evaluated the possibility of generating a system that helps orthodontic clinicians to select a treatment among orthodontics or orthognathic surgery, using cephalometric measurements and a sample of patients of Colombian nationality. The results show good performance of the selected algorithms since they showed an acceptable sensitivity and generalization in the classification, which can be very useful for clinicians in their decision making.

The acquisition of a greater amount and type of data could open the way to use other types of more complex algorithms such as convolutional networks for the extraction of the features embedded in the data.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Zhou, Z., Liu, F., Shen, S., Shang, L., Shang, L., Wang, X.: Prevalence of and factors affecting malocclusion in primary dentition among children Xi'an, China. BMC Oral Health **16**(1), 1–11 (2016). https://doi.org/10.1186/S12903-016-0285-X
2. Almeida, A.B., Leite, I.C.G., Melgaço, C.A., Marques, L.S.: Dissatisfaction with dentofacial appearance and the normative need for orthodontic treatment: determinant factors. Dental Press J. Orthod. **19**(3), 120–126 (2014). https://doi.org/10.1590/2176-9451.19.3.120-126.OAR
3. Jawad, Z., Bates, C., Hodge, T.: Who needs orthodontic treatment? Who gets it? And who wants it? Br. Dent. J. **218**(3), 99–103 (2015). https://doi.org/10.1038/sj.bdj.2015.51

4. Gravely, J.F., Johnson, D.B.: Angle's classification of malocclusion: an assessment of reliability. Br. J. Orthod. **1**(3), 79–86 (1974). https://doi.org/10.1179/BJO.1.3.79

5. Santos, R.R., Nayme, J.G., Garbin, A.J., Saliba, N., Garbin, C.A., Moimaz, S.A.: Prevalence of malocclusion and related oral habits in 5- to 6-year-old children. Oral Health Prev. Dent. **10**(4), 311–8 (2012). https://doi.org/10.3290/J.OHPD.A28901

6. Cenzato, N., Nobili, A., Maspero, C.: Prevalence of dental malocclusions in different geographical areas: scoping review. Dent. J. (Basel) **9**, 117 (2021)

7. Myriad Edition. El desafío de las enfermedades bucodentales - una llamada a la acción global. https://www.fdiworlddental.org/sites/default/files/2021-03/book_spreads_oh2_spanish.pdf. Accessed 21 May 2022

8. Alhammadi, M.S., Halboub, E., Fayed, M.S., Labib, A., El-Saaidi, C.: Global distribution of malocclusion traits: a systematic review. Dental Press J. Orthod. **23**(6), 40.e1–40.e10 (2018). https://doi.org/10.1590/2177-6709.23.6.40.E1-10.ONL

9. Kiep, P.: Grado de maloclusiones según el índicede estéti-cadental en pacientes que acudieron a la universidad del pacífico. Rev. cient. cienc. salud **3**(1), 56–62 (2021). https://doi.org/10.53732/rccsalud/03.01.2021.56

10. Arnett, G.W., Bergman, R.T.: Facial keys to orthodontic diagnosis and treatment planning. Part I. Am. J. Orthod. Dentofacial Orthop. **103**(4), 299–312 (1993). https://doi.org/10.1016/0889-5406(93)70010-L

11. Paiva, J.B., Attizzani, M.F., Junior, H.M., Neto, J.R.: Facial harmony in orthodontic diagnosis and planning. Braz. Oral Res. **24**(1), 52–57 (2010). https://doi.org/10.1590/s1806-83242010000100009

12. Choi, H.: Artificial intelligent model with neural network machine learning for the diagnosis of orthognathic surgery. J. Craniofac. Surg. **30**(7), 1986–1989 (2019). https://doi.org/10.1097/SCS.0000000000005650

13. Incorvati, C., Gulotta, C., Maria, F., Mirabile, C., Badiali, G., Marchetti, C.: Current trends in skeletal borderline patients: surgical versus orthodontic treatment decisions–what is the evidence? Appl. Sci. **12**(9), 4636 (2022). https://doi.org/10.3390/APP12094636

14. Hung, K.F., Ai, Q.Y.H., Leung, Y.Y., Yeung, A.W.K.: Potential and impact of artificial intelligence algorithms in dento-maxillofacial radiology. Clin. Oral Invest. **26**(9), 5535–5555 (2022). https://doi.org/10.1007/S00784-022-04477-Y

15. Jayathilake, A.M.I.C.K., Nawarathna, L.S., Nagarathne, P.N.P.S.: Prediction of malocclusion pattern of the orthodontic patients using a classification model. In: Shakya, S., Balas, V.E., Haoxiang, W., Baig, Z. (eds.) Proceedings of International Conference on Sustainable Expert Systems. LNNS, vol. 176, pp. 279–287. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4355-9_22

16. Chicco, D., Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics **21**(1), 1–13 (2020). https://doi.org/10.1186/S12864-019-6413-7/TABLES/5

17. Hajian-Tilaki, K.: Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J. Intern. Med. **4**(2), 627 (2013). /pmc/articles/PMC3755824/. Accessed 24 May 2022

18. Schwartz-Marín, E., Wade, P.: Explaining the visible and the invisible: public knowledge of genetics, ancestry, physical appearance and race in Colombia. Soc. Stud. Sci. **45**(6), 886–906 (2015). https://doi.org/10.1177/0306312715621182

19. Jiménez, I., Villegas, L., Salazar-Uribe, J.C., Álvarez, L.G.: Facial growth changes in a Colombian Mestizo population: an 18-year follow-up longitudinal study using

linear mixed models. Am. J. Orthod. Dentofac. Orthop. **157**(3), 365–376 (2020). https://doi.org/10.1016/J.AJODO.2019.04.032

20. Aguirre, L., Castillo, D., Solarte, D., Moreno, F.: Frequency and variability of five non-metric dental crown traits in the primary and permanent dentitions of a racially mixed population from Cali, Colombia. Dental Anthropol. J. **19**(2), 39–48 (2006). https://doi.org/10.26575/DAJ.V19I2.119

21. Behbehani, F., Hicks, E.P., Beeman, C., Kluemper, G.T., Rayens, M.K.: Racial variations in cephalometric analysis between Whites and Kuwaitis. Angle Orthod. **76**(3), 406–411 (2006). ISSN 0003-3219, https://doi.org/10.1043/0003-3219(2006)076[0406:RVICAB]2.0.CO;2

22. Lee, J.J., Ramirez, S.G., Will, M.J.: Gender and racial variations in cephalometric analysis. Otolaryngol.–Head Neck Surg. Official J. Am. Acad. **117**(4), 326–329 (1997). ISSN 0194-5998, https://doi.org/10.1016/S0194-5998(97)70121-9

# A MATLAB-Based Graphical User Interface to Assess Conventional and Chirp-Coded Ultrasonic Excitation

Rojelio de Bairro[1(✉)] , Fábio Henrique Almeida Fernandes[1,2] ,
Ednilson de Souza Contieri[1] , Cristhiane Gonçalves[4] ,
Gilson Maekawa Kanashiro[1,3] , Amauri Amorin Assef[1,2] ,
Joaquim Miguel Maia[1,5,6] , and Eduardo Tavares Costa[7]

[1] Graduate Program in Electrical and Computer Engineering (CPGEI),
Federal University of Technology - Paraná (UTFPR), Curitiba, Brazil
rojeliodebairro@yahoo.com.br
[2] Academic Department of Electrical Engineering (DAELT), UTFPR, Curitiba,
Brazil
[3] Federal Institute of Education, Science and Technology of Paraná (IFPR),
Paranavaí, Brazil
[4] Academic Department of Electronics (DAELE), UTFPR, Ponta Grossa, Brazil
[5] Academic Department of Electronics (DAELN), UTFPR, Curitiba, Brazil
[6] Graduate Program in Biomedical Engineering (PPGEB), UTFPR, Curitiba, Brazil
[7] DEEB-FEEC & CEB, State University of Campinas (UNICAMP), Campinas,
Brazil

**Abstract.** Innovative coded excitation techniques have been proposed
to increase the signal-to-noise ratio (SNR) of ultrasound signals, which
are significantly attenuated by scattering and absorption. Among the
methods applied, the linear-frequency modulation signal, commonly
defined as chirp signal, has been studied to provide images with greater
depth, even in high attenuation media, maintaining the spatial resolution
found in conventional excitation systems. This article presents a graph-
ical user interface (GUI) based on Matlab to simulate short-duration
conventional excitation (CE) pulses and long-duration chirp-coded exci-
tation (CCE) pulses. The GUI allows the selection of apodization win-
dow, center frequency, and pulse duration parameters. In addition, it is
possible to configure the bandwidth of the chirp signal. Pulse evaluations
were performed with a central frequency of 1.6 MHz, using three cycles
for CE and a duration of 5, 10, and 20 µs for CCE with a bandwidth
of ±200 kHz, ±400 kHz, and ±1 MHz in a phantom simulated with ten
targets. The echo signals for the CCE were processed using a matched fil-
ter to evaluate the spatial resolution and attenuation. Simulation results
demonstrate the flexibility and performance of the proposed GUI for
ultrasound excitation studies. The evaluation of CCE with a frequency
of 1.6 MHz ± 1 MHz and matched filter improved spatial resolution by
86%. In contrast, a maximum increase in attenuation of the processed
signal of 33% was observed.

**Keywords:** Ultrasound · conventional excitation · chirp-coded excitation · match filter · signal processing

## 1 Introduction

Ultrasound imaging is one of the most widespread modalities in various applications and clinical examinations in the medical field. Among the advantages of ultrasound, it can be underlined the ability to generate images in real-time, its non-ionizing nature, being a non-invasive method, and having a relatively low cost compared to other medical imaging diagnoses [3,13,15].

The spatial resolution and penetration depth are the main parameters to determine the quality of the ultrasound image [4]. There is an improvement in the spatial resolution due to the high frequencies of ultrasonic emission with conventional excitation (CE) of short duration and Gaussian profile, typically up to 20 MHz. However, the attenuation is strongly frequency-dependent and, consequently, imposes limitations on the penetration capacity of the acoustic wave [14]. An alternative to overcome this obstacle is to increase the power of acoustic emission, increasing the amplitude or duration of the excitation pulse. However, increasing the excitation signal amplitude increases the instantaneous acoustic power, a parameter that is limited to pre-established values by regulatory agencies [9]. An ultrasound-encoded excitation method, which causes an increase in the transmitted pulse, was proposed by O'Donnell to overcome that limitation in 1992 [10].

The use of coded excitation techniques makes it possible to lengthen the transmitted pulse with modulations temporarily, subsequently compressing the output (reception), usually through a matched or mismatched filter. Coded chirp excitation (CCE) ultrasound signals are used for this function. The advantages of using coded signals are an increase in penetration depth and an increase in the signal-to-noise ratio (SNR). A higher SNR allows images of structures located deeper within the human body to be captured with better resolution [7,8].

For the development and evaluation of innovative transmission techniques, it is necessary to have access to excitation control parameters that, usually, are not fully available in commercial ultrasound equipment. Thus, the development of computational simulation tools, and open and flexible ultrasound platforms for research, with the ability to generate coded ultrasound pulses, have been proposed [5,7,8,11,12]. As an example, Medeiros et al. [6] presented a flexible Matlab-based interface to simulate complex analog waveforms with fixed pulse duration.

This article presents a new graphical user interface (GUI) with the App Designer program, which is part of Matlab, to simulate CE and CCE pulses of short and long duration. The methods and results of generated signals in a simulated phantom are presented. In the case of CCE, applying a matched filter for signal optimization was evaluated.

## 2  Materials and Methods

The computational tool used in this work to develop the GUI was the software Matlab ® Release 2022a (MathWorks Inc, USA). Figure 1 shows the interface developed for simulating CE and CCE pulses. Pulses' interaction with the medium - in this case, a computational phantom - was also evaluated using the GUI.
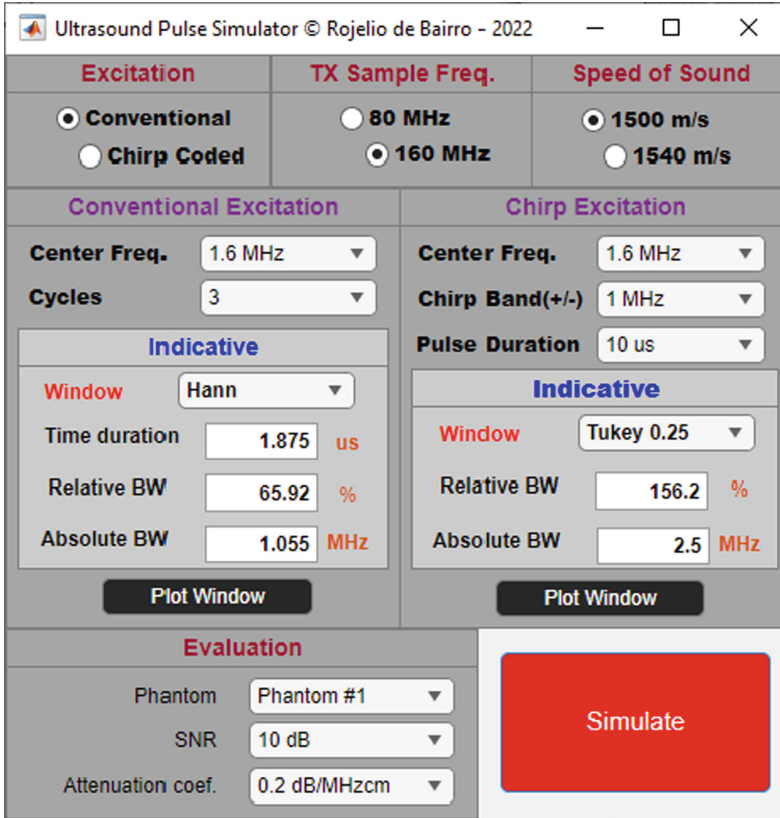


**Fig. 1.** Graphical user interface proposed for ultrasonic excitation research

For CE evaluation, it is possible to configure the central frequency ($f_c$), the number of cycles, and the apodization window. The CCE configuration has the parameters for bandwidth and pulse duration. The linear frequency chirp is calculated by using (1), where $f_0$ and $\phi_0$ are the starting frequency and the initial phase, respectively, at time $t = 0$, $f_1$ is the final frequency, and $T$ is the time it takes to sweep from $f_0$ to $f_1$.

$$x(t) = sin\left[\phi_0 + 2\pi\left(\frac{f_1 - f_0}{2T} \cdot t^2 + f_0 t\right)\right] \tag{1}$$

$$BW_{rel} = \frac{BW_{abs}}{f_c} \cdot 100 \qquad (2)$$

The spectrum's absolute ($BW_{abs}$) and relative ($BW_{rel}$) bandwidths in the generated signal frequency are calculated in both cases. The $BW_{abs}$ is obtained through the Fast Fourier Transform (FFT) at $-6$ dB, and the $BW_{rel}$ is calculated according to (2).
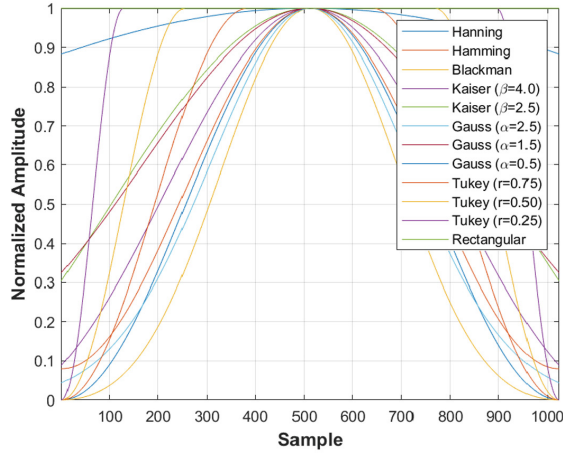


(a)
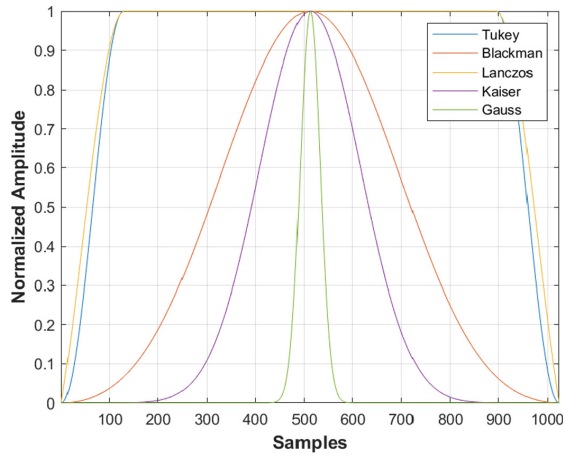


(b)

**Fig. 2.** Windowing options for (a) CE and (b) CCE

For pulse shape definition, windowing functions were included for both excitation modes. Figure 2a and Fig. 2b show the window options implemented in the

GUI for CE and CCE, respectively. The x-scale indicates the number of samples used in both simulations for illustrative purposes. This number depends on the parameters selected in the GUI.

The available parameter configuration options are shown in Table 1. However, other options can be easily implemented.

**Table 1.** GUI Configuration Parameters

| Item | Options |
|---|---|
| Speed of sound [m/s] | 1500 and 1540 |
| Sampling frequency [MHz] | 80 and 160 |
| Central frequency [MHz] | 0.5, 1.6 and 5.0 |
| Chirp signal bandwidth [MHz] | $\pm 0.2$, $\pm 0.4$, $\pm 1.0$ and $\pm 2.0$ |
| Chirp pulse duration [μs] | 5, 10 and 20 |
| Number of cycles | Between 3 and 10 |
| *Phantom selection for simulation | Pre-computed phantoms |
| Gaussian white noise [dB] | 0 (max.) to 30 (min.) |
| Medium attenuation coefficient | 0.1 to 2.0 dB/MHz.cm |

*The phantom must be generated by script before the simulation.

After configuring all the excitation parameters, the simulation is performed by clicking the Simulate button. Then, the impulse response, its frequency response, the convolution of the signal with the simulated phantom, and the logarithmic compression of the final signal, presented in the next section, are generated. The phantom shown in Fig. 3 was modeled in a one-dimensional structure with ten scatters (A1 to A10) spaced 10 mm apart, as described in [2].

For evaluation of the GUI, CE tests were performed with a central frequency of 1.6 MHz, duration of 3 pulses, and Hanning windowing. The CCE was evaluated with a central frequency of 1.6 MHz, durations of 5, 10, and 20 μs, bandwidth of $\pm 200$ kHz, $\pm 400$ kHz, and $\pm 1$ MHz, and Tukey windowing (25%). The sampling frequency was set at 160 MHz, and the sound propagation speed was set at 1500 m/s.

This work results in a quality optimization of the ultrasound signal with the CCE. Waveforms were processed using a matched filter. This compression technique allows maximizing the SNR in the presence of white Gaussian noise to improve the quality of the ultrasonic image. After acquiring the radiofrequency (RF) echo, the time inversion of the impulse response of the CCE pulse is performed to generate the matched filter. This signal is convolved with the backscattered signal to generate the compressed signal. White noise with attenuation of 10 dB was added to the convolved signal to verify the attenuation and axial resolution behavior.
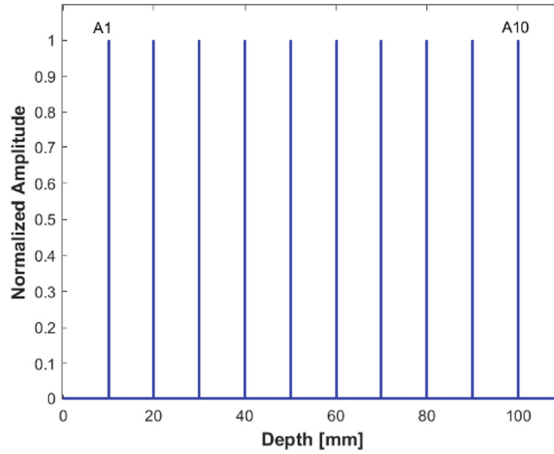
**Fig. 3.** Computer phantom with ten targets

## 3  Results

This section evaluates excitation/compression mechanisms' effects on ultrasonic signals. Furthermore, a comparison of the magnitude of the main lobes and axial resolution of the ten targets at $-6$ dB of the generated signals by CE and CCE with matched filter is presented. The tests were performed considering the attenuation of the medium of 0.2 dB/MHz.cm.

### 3.1  Conventional Excitation

The CE pulse is shown in Fig. 4a. For this signal, the $BW_{abs}$ and the $BW_{rel}$ were equal to 1.055 MHz and 65.92%, respectively, obtained through the frequency spectrum shown in Fig. 4b. Figure 4c shows the result of the convolution of the CE pulse with the computational phantom to simulate the backscattered signals. In addition to the resulting echo, the red dashed curve is the signal envelope obtained by the absolute value of the Hilbert Transform [1]. Logarithmic compression is applied to the envelope to better visualize the signal with a dynamic range of $-60$ dB, as shown in Fig. 4d. This signal has a significant amount of noise that prevents the definition of the deepest targets, precisely, A9 and A10.

### 3.2  Coded Chirp Excitation

Figure 5a shows the CCE signal with a frequency of 1.6 MHz $\pm$ 1 MHz and a duration of 10 µs due to the applied windowing. $BW_{abs}$ and $BW_{rel}$, equal to 2.5 MHz and 156.3%, respectively, were obtained from the frequency spectrum of the chirp signal in Fig. 5b. Figure 5c shows the result of the convolution of the CCE pulse with the computational phantom and the respective envelope.
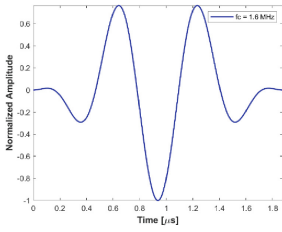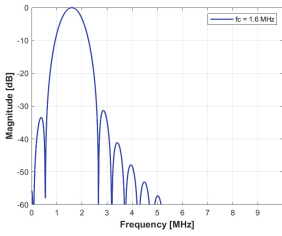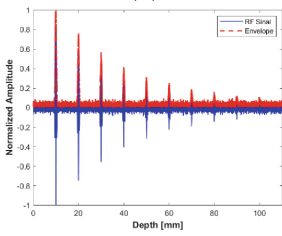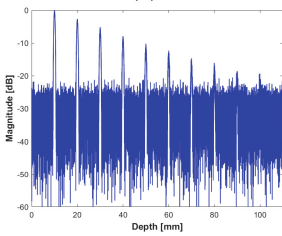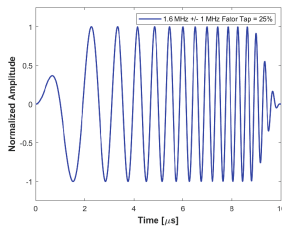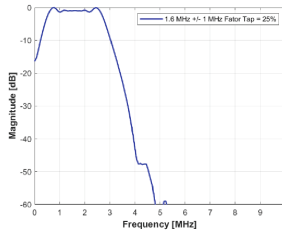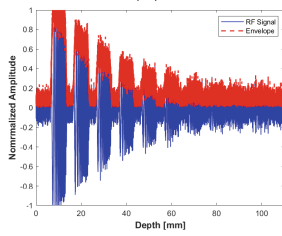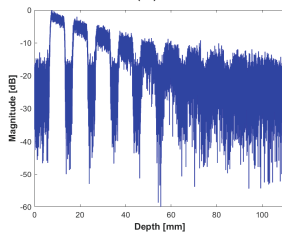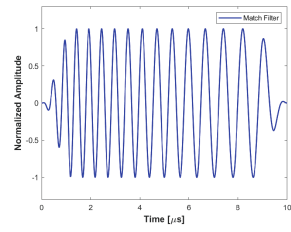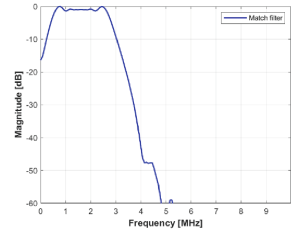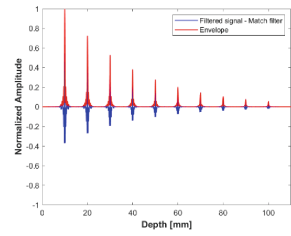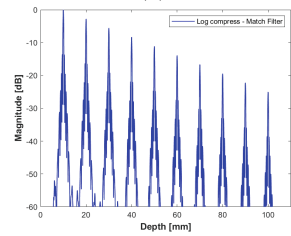
**Fig. 4. (a)** CE pulse. **(b)** Pulse frequency spectrum. **(c)** Simulation of backscattered signals. **(d)** Logarithmic compression of the echo signal

**Fig. 5. (a)** CCE pulse. **(b)** Pulse frequency spectrum. **(c)** Simulation of backscattered signals. **(d)** Logarithmic compression of the echo signal

**Fig. 6. (a)** Matched filter. **(b)** Filter frequency spectrum. **(c)** CCE pulse compressed by the matched filter. **(d)** Logarithmic compression of the resulting signal

Figure 5d shows the sign after logarithmic compression. In this case, the target axial resolution is compromised as a function of the width of the long-duration CCE pulses.

### 3.3  Coded Chirp Excitation with Pulse Compression

The time inversion of the CCE impulse response, corresponding to the matched filter, and its frequency spectrum are shown in Fig. 6a and Fig. 6b, respectively. The convolution result with the backscattered signal shown in Fig. 5c with the matched filter of Fig. 6a is shown in 6c. Compared to Fig. 4d and Fig. 6d, applying the matched filter improves the SNR and axial resolution of the ten targets. However, that compression caused secondary lobes that could indicate false artifacts. These lobes can be minimized in future works by applying mismatched filters [7,8].
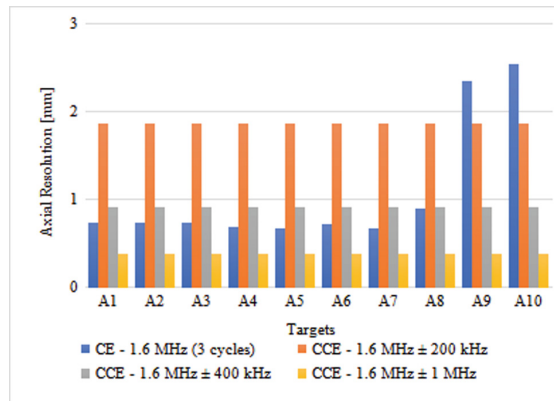


**Fig. 7.** Comparison of the axial resolution of CE and CCE signals in the phantom

In order to quantitatively compare CE and CCE with matched filter, Fig. 7 shows the results of the evaluation of the axial resolution in the phantom. As expected, the axial resolution of the CE gets worse with depth, especially on the A9 and A10 targets. This behavior did not occur in the CCE, which remained stable. The mean value of the axial resolution of the CCE for the frequency of 1.6 MHz and bandwidths of $\pm 200$ kHz, $\pm 400$ kHz, and $\pm 1$ MHz was equal to 1.842, 0.900, and 0.361 mm, respectively, with zero standard deviation in all the cases. Considering targets A9 and A10, the best axial resolution of CCE with matched filter occurred with $\pm 1$ MHz band and was equal to 85 and 86%, respectively.

Figure 8 presents the evaluation of the signal attenuation, and the target A1 has an attenuation of 0 dB. CE presented better results for all targets, with a maximum difference of 33% in target A10. On the other hand, the attenuation of the CCE was the same for each target in the different bandwidths. Thus, there
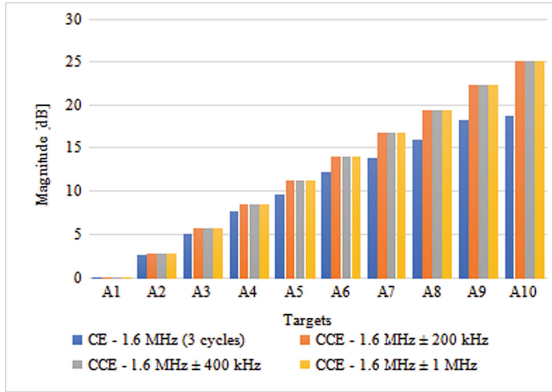
**Fig. 8.** Comparison of CE and CCE signal attenuation in the phantom

is a cost-benefit relationship between the application of the techniques since processing the encoded chirp signal requires a more significant computational effort.

## 4    Conclusion

This work presented a flexible GUI for simulating short and long excitation pulses. This computational tool can be used in the research of new techniques for optimizing the quality of images generated by ultrasound using CE and CCE. The GUI is easy to use, and customizations can be performed according to scientific research needs. The presented study shows that the optimization of excitation parameters plays a relevant role in improving the SNR using chirp-coded pulses in deeper regions. This work is a continuation of the research started in [6], in which the GUIDE tool from Matlab was used. In this work, the APP Designer software enabled a better use of new layout applications and integration with digital signal processing tools that will be evaluated in future research.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. Assef, A.A., de Oliveira, J., Maia, J.M., Costa, E.T.: FPGA implementation and evaluation of an approximate Hilbert transform-based envelope detector for ultrasound imaging using the DSP builder development tool. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2813–2816. IEEE (2019)
2. Garcia, V.B.: Development and evaluation of a multi-channel transmission system with arbitrary waveform with chirp coded excitation. Master's thesis, State University of Campinas, Campinas (2020)
3. Hartley, C.J.: Ultrasonic Bioinstrumentation: Da Christensen. Wiley, New York (1989). 235 p
4. Jensen, J.A.: Medical ultrasound imaging. Prog. Biophys. Mol. Biol. **93**(1–3), 153–165 (2007)
5. Jensen, J.A., et al.: Ultrasound research scanner for real-time synthetic aperture data acquisition. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**(5), 881–891 (2005)
6. Medeiros, R.A.C., Assef, A.A., Maia, J.M., da Silva Ruzyk, M.J., de Souza Kohler, M.A.F., Costa, E.T.: A Matlab GUI interface for multi-level pulse amplitude modulation (PAM) generation in medical ultrasound research. In: 2021 IEEE UFFC Latin America Ultrasonics Symposium (LAUS), pp. 1–4. IEEE (2021)
7. Misaridis, T., Jensen, J.A.: Use of modulated excitation signals in medical ultrasound. Part I: basic concepts and expected benefits. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**(2), 177–191 (2005)
8. Misaridis, T., Jensen, J.A.: Use of modulated excitation signals in medical ultrasound. Part II: design and performance for medical imaging applications. IEEE Trans. Ultrason. Ferroelectri. Freq. Control **52**(2), 192–207 (2005)
9. Ng, K.H.: International guidelines and regulations for the safe use of diagnostic ultrasound in medicine. J. Med. Ultrasound **10**(1), 5–9 (2002)
10. O'Donnell, M.: Coded excitation system for improving the penetration of real-time phased-array imaging systems. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**(3), 341–351 (1992)
11. Ozum, H.E., Yetik, H., Kilinc, M.S., Ergun, A.S.: An open source, modular and scalable HIFU driver system. In: 2017 IEEE International Ultrasonics Symposium (IUS), pp. 1–4. IEEE (2017)
12. Qiu, W., Yu, Y., Tsang, F.K., Sun, L.: A multifunctional, reconfigurable pulse generator for high-frequency ultrasound imaging. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **59**(7), 1558–1567 (2012)
13. Smith, J., Jones, M., Jr., Houghton, L., et al.: Future of health insurance. N. Engl. J. Med. **965**, 325–329 (1999)
14. Stein, J.H., et al.: Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American society of echocardiography carotid intima-media thickness task force endorsed by the society for vascular medicine. J. Am. Soc. Echocardiogr. **21**(2), 93–111 (2008)
15. York, G., Kim, Y.: Ultrasound processing and computing: review and future directions. Annu. Rev. Biomed. Eng. **1**(1), 559–588 (1999)

# A Comparison of Classifiers for Epileptic Seizure Prediction Based on Heart Rate Variability

Rafael Sanchotene Silva[1(✉)] , Cesar Ramos Rodrigues[1] , Roger Walz[2] , and Jefferson Luiz Brum Marques[1]

[1] Department of Electrical and Electronics Engineering, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil
`r.sanchotene@gmail.com`

[2] Department of Clinical Medicine, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil

**Abstract.** Epilepsy is a neurological disorder characterized by recurrent seizures due to abnormal discharges in cortical networks of the brain. A seizure prediction method with a low false-positive rate in a high confidence interval and without side effects may improve patients' quality of life. Heart rate variability (HRV) analysis is among the most promising approaches for seizure prediction. This method indirectly assesses the behavior of the autonomic nervous system (ANS) through cardiac rhythm activity. Artificial intelligence (AI) classifiers may predict seizures and distinguish the different phases in ECG signals. This work evaluated several classifiers for seizure prediction and studied them in terms of computational cost for training, sensitivity, accuracy, false-positive rate (FPR), and their suitability for wearable applications using the HRV approach. Relied on the results, the Support Vector Classifier (SVC) obtained the best set of scores, including the highest accuracy, 97.57%, as well as the second-highest Sen, Spe, and NPV scores, 97.70%, 97.51%, and 98.83%, respectively for preictal periods, considering an evaluation of 14.08 h from six patients' ECG data.

**Keywords:** Epilepsy · HRV · Seizure Prediction · Machine Learning

## 1 Introduction

Epilepsy is one of the most common chronic neurological disorders, affecting approximately 50 million people worldwide [1]. This abnormal brain behavior can cause some clinical manifestations in the patients that include involuntary movements, distorted perceptions, loss of consciousness, permanent brain damage, and in the most severe cases, sudden unexpected death (SUDEP) [2–4]. Despite anti-epileptic drugs controlling many cases of epilepsy, 30% of patients still suffer from uncontrolled seizures [4]. Other alternatives to epilepsy treatment include surgical intervention and neurostimulation, but some patients are not eligible for these [5].

Epileptic seizures prediction methods are based on two signal approaches, the electroencephalogram (EEG) and the electrocardiogram (ECG) [2,5,6]. The most common methods based on EEG signals involve coupling electrodes to the patient's scalp [7], which is impractical for wearable device applications. Other methods rely on surgical procedures to map the epileptic focus, and the potential treatments carry the risk of brain damage and infection [5].

Given the highlighted limitations and drawbacks of the EEG-based method and with the aim of the feasibility for portable devices available to patients, the methodology of HRV analysis emerges as a promising alternative [8]. The choice of ECG signal is immediate because it offers many advantages, such as easier acquisition, processing, and lower cost, making it suitable for detection and prediction systems and patient comfort. Further, according to the results of a recent study [9], based on an analysis that included preictal periods, there was no evidence for significant differences in HRV metrics according to epilepsy type. Therefore, this fact makes this method particularly useful when compared to EEG approaches. Figure 1 depicts the concept of a wearable device split into two parts. The first part includes a dedicated circuit designed for signal conditioning, summarized here as an analog front-end (AFE), and a heartbeat detector. Then, identified heartbeats are sent to a microcontroller such as ESP32®, which contains the second part performed by the AI algorithms for seizure prediction.
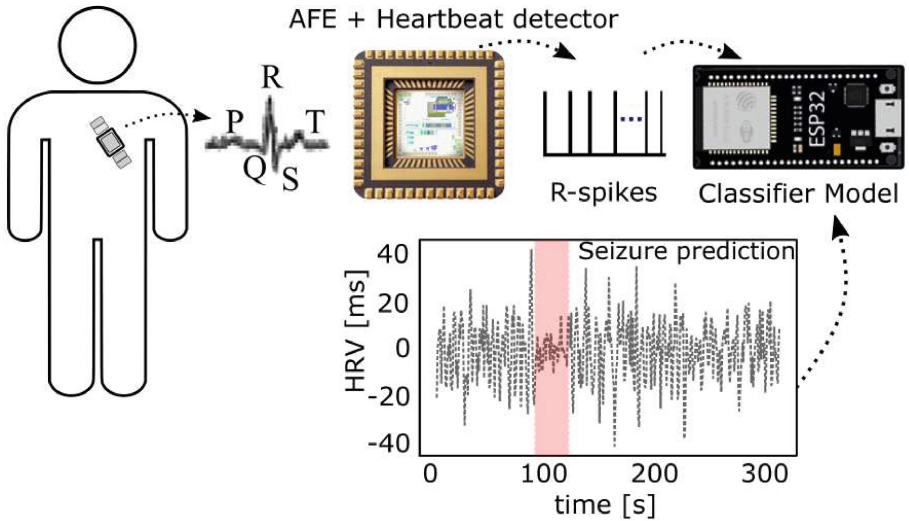


**Fig. 1.** Application.

The state-of-art related EEG approach is replete with research on AI comparison to find the best algorithm for seizure prediction, and examples include [6,10–13]. To the best of the authors' knowledge, although there are several EEG studies using machine learning models, there is still a lack of comprehensive

evaluation of AI models for seizure prediction based on ECG. Among the best examples in this area are [5,14], both studies with Support Vector Machine classifiers (SVC), and that do not consider the characteristics required for real-time operation.

This work is the sequence of the former which consisted of the design and fabrication of the integrated circuit for ECG acquisition and extraction of heartbeats for HRV applications in [15]. The main objective of this study is to show the comparison and selection of the best trained and validated classifier model suited for wearable applications relying on HRV for seizure prediction.

## 2   Material and Methods

Although research on EEG-based prediction is extensive, there is no consensus on the duration of the prediction phase. Figure 2 shows the classification of ECG recordings into four main intervals, as done in [16]: interictal - the signal baseline and at least 30 min before seizure onset; preictal - periods from 5–15 min before seizure onset; ictal - identified seizure onset in the dataset; and the postictal - period immediately after the seizure ends. As with EEG-based approaches, seizure prediction with ECG is based on signal phase discrimination. Thus, true alarms are the preictal periods detected within a predetermined interval [17].
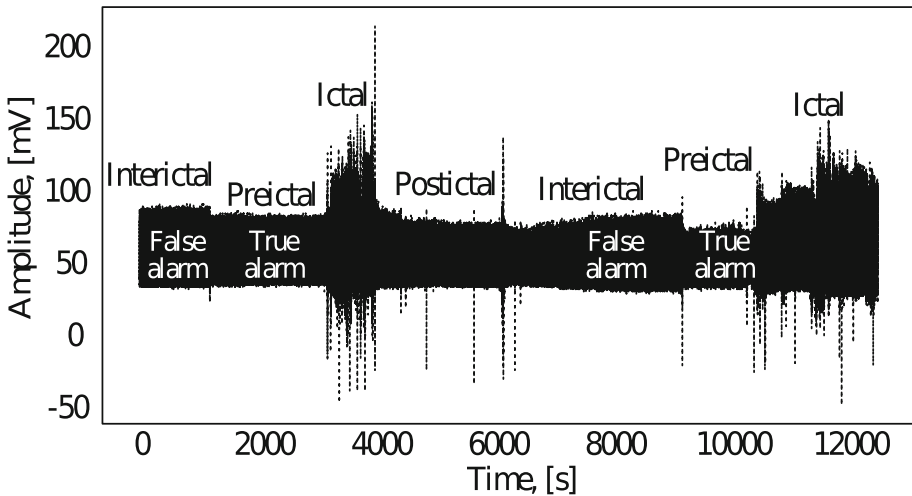


**Fig. 2.** Periods in an epilepsy ECG recording. In the interictal period, an alarm triggered is considered a false alarm, whereas, in the preictal period, it is considered a true alarm.

To evaluate the performance of the supervised machine learning models for seizure prediction, each one it is have been analyzed in terms of its accuracy

(ACC), sensitivity (Se), specificity (Spe), positive predictive value (PPV), negative predictive value (NPV), and false-positive rates (FPRs). For timing computing, it was used the 20-run average time for training and validation with the randomized search cross-validation technique using twenty iterations and five folds. In addition, for the sake of simplicity and since the postictal phases are not significantly different from the interictal phases, we only consider the interictal, preictal, and ictal phases in this work.

## 2.1  ECG Databases

The AI algorithms were evaluated with PhysioNet databases [18]. The first dataset consists of seven ECG recordings with ten identified seizures. Signals come from a heterogeneous group of patients with partial epilepsy who exhibited postictal heart rate oscillations. These recordings, sampled at 200 Hz, contained remarkable low-frequency transient signals in the 10–100 mHz. The second data set included recordings from fourteen patients admitted to the Department of Neurology and Neurophysiology at the College of Siena [19]. The patients, nine men (age 25–71) and five women (age 20–58), were monitored with video-EEG containing recordings of 1 or 2 ECG channels with a sampling rate of 512 Hz.

## 2.2  HRV Features

For predictions based on the HRV approach, several metrics have been extracted and supported by the state of the art. They can be divided into two groups: linear and non-linear. The main linear metrics derived from beat-to-beat intervals (a.k.a. RR-intervals) are the average of NN intervals, the root means squared difference between consecutive normal heartbeats (RMSSD), and the standard deviation of NN intervals (SDNN). Based on the Poincare plot, nonlinear metrics can be derived to quantify self-similarity in a time series, such as the NN intervals formed from identified beats. Cardiac-Sympathetic Index (CSI), Cardiac Vagal Index (CVI), modified CSI [2], and the ratio of CSI/CVI are the most commonly used nonlinear metrics. Besides the aforementioned nonlinear metrics, it was examined one of the Hjorth parameters commonly used to analyze EEG signals, the Hjorth activity (AT) [20], which reflects a time function's variance.

## 2.3  Hypothesis

Following the EEG approaches to seizure prediction, the basic assumption is that the seizure does not occur suddenly, i.e., the process begins several minutes to hours before clinical manifestations (i.e., preictal phase) [12]. Based on this assumption, previous research [11,12,16,17] has extracted various features from brain-behavior analysis to understand the significant and useful preictal changes in EEG signals [17]. The same approach is possible for HRV-based techniques [5, 21] by extracting the relevant metrics for seizure prediction, as addressed in the following.

– **Linear metrics**: In analyzing the linear metric, we made the following assumptions based on [22]: As the frequency of heartbeats increases, both NN intervals and HRV decrease in absolute value (shorter intervals). During ictal periods, heart rate increases, indicating a decrease in HRV in relation to the interictal phase (baseline), and also a dominance of the sympathetic branch (which is activated under stress conditions) relative to vagal activity (which is activated under rest conditions).

– **Nonlinear metrics**: Lower variability in ANS indicates a dominant role for the sympathetic nervous system (higher heart rate) over the parasympathetic nervous system [22]. It is reasonable to assume that CVI metrics are not significantly altered by parasympathetic activity during the preictal and ictal phases. On the other hand, CSI and modified CSI will increase dramatically at the onset of the seizure (preictal) and during the seizure (ictal) compared with baseline (interictal phase). It is expected that the Hjorth parameter will show a tendency toward the unbalancing between the sympathetic and parasympathetic branches, i.e., as the seizure onset approaches, the value of AT compared with the baseline increases.

## 3    Supervised Machine Learning Classifiers

To balance the estimator training step, it was used the stratification technique, which keeps the same percentage of samples for each labeled period [23]. Following the flow shown in Fig. 3, the first step begins with the acquisition of RR intervals from a low-power analog front-end (see Fig. 1), then the RR intervals are evaluated to remove ectopics and outliers beats, and then HRV features are extracted and segmented. Following the sequence of steps, the segmented feature is individually scaled based on its minimum and maximum values into zero to one range. In the last step, the models are trained/validated and tested with totally new data, and the best model in terms of computational costs, sensitivity, and false-positive rate is embedded into a low-power microcontroller.

Of the trained/validated and tested models, this work considered as criteria only those that have achieved reasonable sensitivity, i.e., the overall summed average of the three ECG periods, above 85%. This consideration allows the model to be used in real-time applications with reliably estimated reach. The AI models used in this work include Decision Tree (DT), Extra Tree (ET), Random Forest (RF), k-nearest neighbors (KNN), Adaptive Boost (ADA), Support Vector Machine (SVC), and the Artificial Neural Network (ANN).
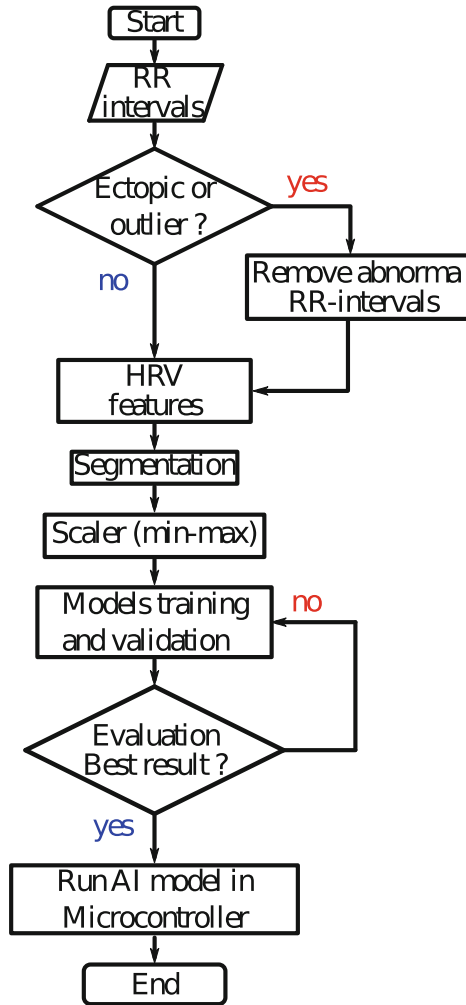
**Fig. 3.** Flowchart.

## 4   Results and Discussion

The time evolution of a patient's CSI and CVI metrics, who had a seizure at 8.13 min after the start of recording, is depicted in Fig. 4. In Fig. 4-a), the contribution of parasympathetic branch activity to cardiac regulation is higher than sympathetic branch activity around 5 min before seizure onset. From this period, it can be seen in Fig. 4-b) that there is a trend toward sympathetic predominance, which can be useful for prediction alarm systems.

Using principal component analysis (PCA) [24], it reduced the dimensionality of seventeen features to three: the average of RR-intervals, Hjorth activity (AT),

and the modified CSI. The features extracted from an ECG signal from the Siena dataset corresponding to a 27-year-old female with three focal onsets of impaired awareness (IAS) are shown in Fig. 5.

The following measurements were collected by the preprocessing algorithm over a 5-min acquisition window with a sliding step of 1 s. According to the findings, as the seizure onset approaches, the heartbeat frequency tends to increase, resulting in a decrease in HRV and, as a result, RR intervals. Hjorth activity (AT) and altered cardio-sympathetic parameters, on the other hand, significantly increase as shown by the green-highlighted markers. Besides, it is possible to observe that Fig. 5-b) presents two red-highlighted markers that correspond to the modified CSI parameter non-changing, which means miss-classification by the models or to the false-positive detections.
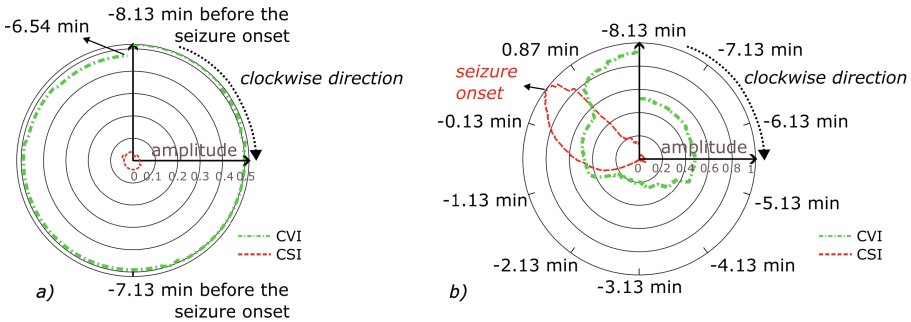


**Fig. 4.** CVI and CSI time evolution in two cases: a) instants before seizure onset, and b) when the seizure onset. In the first case, the sympathetic activity amplitude is lower than the parasympathetic; in the second case, the sympathetic branch becomes dominant 5 min before the seizure onset.

Postictal and Siena datasets containing interictal, preictal, and ictal ECG phases were used in training and validation, totaling 97.47 h. Besides, a total of 14.08 h of new data from six patients were used to test the models. In the last test evaluation, false-positive rates were calculated using 31.17 h of interictal data from the MIT/BIH ECG dataset. Using Python Scikit Learn library [23], the following model configurations were used to achieve the results shown in Table 1.

– **Decision Tree (DT):** class weight: "balanced", criterion: "entropy", maximum depth: 64, minimum samples leaf: 4, minimum samples split: 16;
– **Extra Tree (ET):** criterion: "entropy", maximum depth: 30, minimum samples leaf: 2, minimum samples split: 16;
– **Random Forest (RF):** bootstrap: "false", maximum depth: 20, minimum samples leaf: 2;
– **K-Nearest Neighbours (KNN):** algorithm: "kd_tree", leaf size: 40;
– **Adaptive Boost (ADA):** learning rate: 0.8, number of estimators: 32, base estimator: Decision Tree;

– **Support Vector Machine Classifier (SVC):** C: 100, class weight: "balanced", coef0: 1, decision function shape: "ovo", degree: 10, kernel: "poly";
– **Artificial Neural Network:** weight constraint: 5, loss: "categorical_crossentropy", epochs: 200, dropout rate: 0.0, batch size: 100, activation: "softmax". Built with two hidden layers of 128, and 64 perceptrons by using "relu" activation function.
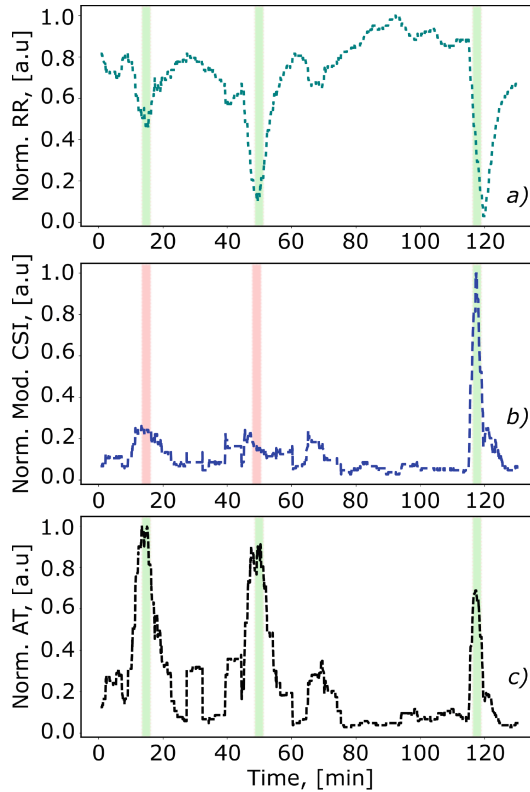


**Fig. 5.** Normalized features extracted from a patient with two IAS. a) RR-intervals average, b) Modified CSI, and c) Hjorth activity (AT).

According to Table 1, it is possible to draw the following conclusions: ANN was the most accurate classifier able to recognize preictal periods with 100% sensitivity, as well as negative predictive value (NPV), which means the likelihood of 100% of a patient not being in the preictal period. These results are especially interesting since can be applied to wearable alarm systems, which may alert the patient in the event of a predicted seizure. Otherwise, i.e., if no alarm arises, the patient can keep living normally.

In terms of specificity, i.e., the proportion of data classified as non-preictal among those which are not, the best model was ADA with 98.66%. ADA also

achieved the lowest FPR, just 0.013, and the highest positive predictive value (PPV) 97.06% for preictal periods among all models when considering both datasets. Despite the interesting results, ADA has missed nearly 11.6% of the preictal periods compared to ANN which may be a problem for patients who has several seizures in short periods. Still analyzing the results, the SVC model reached the best accuracy of 97.57% for preictal period identification. Despite the second-worst time cost, SVC achieved the second-highest Sen, Spe, and NPV scores, 97.70%, 97.51%, and 98.83%, respectively.

**Table 1.** Validation results for different predictor AI models.

| Model | Data | ACC (%) | Sen (%) | Spe (%) | PPV (%) | NPV (%) | [a]FPR | [b]FPR |
|-------|------|---------|---------|---------|---------|---------|--------|--------|
| DT | Interictal | 97.70 | 100.0 | 96.55 | 93.55 | 100.0 | 0.03448 | – |
|    | Preictal | 84.29 | 93.10 | 79.89 | 69.83 | 95.86 | 0.20115 | 0.03568 |
|    | Ictal | 86.59 | 59.77 | 100.0 | 100.0 | 83.25 | 0.00000 | 0.01001 |
| ET | Interictal | 92.59 | 100.0 | 88.89 | 81.82 | 100.0 | 0.11111 | – |
|    | Preictal | 89.27 | 77.39 | 95.21 | 88.99 | 89.39 | 0.04789 | 0.02913 |
|    | Ictal | 96.68 | 90.42 | 99.81 | 99.58 | 95.42 | 0.00192 | 0.00860 |
| RF | Interictal | 95.27 | 100.0 | 92.91 | 87.58 | 100.0 | 0.07088 | – |
|    | Preictal | 89.78 | 85.82 | 91.76 | 83.90 | 92.83 | 0.08238 | 0.03311 |
|    | Ictal | 94.51 | 83.52 | 100.0 | 100.0 | 92.39 | 0.00000 | 0.00862 |
| KNN | Interictal | 89.40 | 100.0 | 84.10 | 75.87 | 100.0 | 0.15900 | – |
|    | Preictal | 83.14 | 73.18 | 88.12 | 75.49 | 86.79 | 0.11877 | 0.02933 |
|    | Ictal | 89.91 | 70.50 | 99.62 | 98.92 | 87.10 | 0.00383 | 0.00950 |
| ADA | Interictal | 96.17 | 100.0 | 94.25 | 89.69 | 100.0 | 0.05747 | – |
|    | Preictal | 95.27 | 88.51 | 98.66 | 97.06 | 94.50 | 0.01341 | 0.03810 |
|    | Ictal | 99.11 | 97.32 | 100.0 | 100.0 | 98.68 | 0.00000 | 0.00816 |
| SVC | Interictal | 96.93 | 100.0 | 95.40 | 91.58 | 100.0 | 0.04598 | – |
|    | Preictal | 97.57 | 97.70 | 97.51 | 95.15 | 98.83 | 0.02490 | 0.04245 |
|    | Ictal | 94.76 | 86.21 | 99.04 | 97.83 | 93.49 | 0.00958 | 0.01835 |
| ANN | Interictal | 94.13 | 84.67 | 98.85 | 97.36 | 92.81 | 0.01149 | – |
|    | Preictal | 94.25 | 100.0 | 91.38 | 85.29 | 100.0 | 0.08621 | 0.03100 |
|    | Ictal | 91.19 | 84.67 | 94.44 | 88.40 | 92.50 | 0.05556 | 0.00907 |

[a]False-positive rate from postictal and Siena databases which contains 31.17 h of interictal, preictal, and ictal phases; [b]False-positive rate (FPR) from 14.07 h of interictal data (MIT-BIH). Average time cost of 20 runs for training/validating of models. Setup: 8 GB memory, CPU i5-1.8 GHz. Average time: DT - 3.24 min; ET - 17.3 min; RF - 121.23 min; KNN - 155.26 min; ADA - 159.32 min; SVC - 1538.4 min; ANN - 6306.31 min.

This set of features turns the SVC into the best classifier for seizure prediction considering the datasets used. Further, in wearable applications, the SCV model can be trained in off-line mode or with cloud computing and then embedded into a microcontroller, thereby mitigating the problem of time costs.

# 5   Conclusion

This study compared supervised machine learning algorithms for predicting epileptic seizures. Three metrics extracted from ECG were used to train-validate, and test several models. Also, three databases were investigated for this purpose, yielding more than 100 h of data. According to the results, by considering only preictal periods, the best model was SVC which obtained the highest accuracy, 97.57%, the second-highest Sen, Spe, and NPV scores, 97.70%, 97.51%, and 98.83%, respectively. Based on the shown results, refractory patients to the conventional treatments may benefit, increasing their life's quality through wearable devices with embedded trained SVC models to predict seizures.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. World Health Organization: Epilepsy. https://www.who.int/en/news-room/fact-sheets/detail/epilepsy. Accessed 25 Apr 2022
2. Jeppesen, J., Beniczky, S., Johansen, P., Sidenius, P., Fuglsang-Frederiksen, A.: Detection of epileptic seizures with a modified heart rate variability algorithm based on Lorenz plot. Seizure **24**, 1–7 (2015). https://doi.org/10.1016/j.seizure.2014.11.004
3. Melo, H., et al.: Ultra-short heart rate variability reliability for cardiac autonomic tone assessment in mesial temporal lobe epilepsy. Epilepsy Res. **174**, 106662 (2021). https://doi.org/10.1016/j.eplepsyres.2021.106662
4. Truong, N.D., et al.: Seizure susceptibility prediction in uncontrolled epilepsy. Front. Neurol. **12** (2021). https://doi.org/10.3389/fneur.2021.721491. https://www.frontiersin.org/article/10.3389/fneur.2021.721491. ISSN 1664-2295
5. Pavei, J., et al.: Early seizure detection based on cardiac autonomic regulation dynamics. Front. Physiol. **8** (2017). https://doi.org/10.3389/fphys.2017.00765. https://www.frontiersin.org/article/10.3389/fphys.2017.00765. ISSN 1664-042X
6. Bhattacharya, A., Baweja, T., Karri, S.P.K.: Epileptic seizure prediction using deep transformer model. Int. J. Neural Syst. **32**(02), 2150058 (2022). https://doi.org/10.1142/S0129065721500581. PMID 34720065
7. Jacobs, D., Hilton, T., del Campo, M., Carlen, P.L., Bardakjian, B.L.: Classification of pre-clinical seizure states using scalp EEG cross-frequency coupling features. IEEE Trans. Biomed. Eng. **65**(11), 2440–2449 (2018). https://doi.org/10.1109/TBME.2018.2797919
8. Dono, F., et al.: Interictal heart rate variability analysis reveals lateralization of cardiac autonomic control in temporal lobe epilepsy. Front. Neurol. **11** (2020). https://doi.org/10.3389/fneur.2020.00842. https://www.frontiersin.org/article/10.3389/fneur.2020.00842. ISSN 1664-2295
9. Faria, T., et al.: Does the type of seizure influence heart rate variability changes? Epilepsy Behav. **126**, 108453 (2022). https://doi.org/10.1016/j.yebeh.2021.108453

10. Shasha, Z., Chen, D., Ranjan, R., Hengjin, K., Tang, Y., Zomaya, A.: A lightweight solution to epileptic seizure prediction based on EEG synchronization measurement. J. Supercomput. **77**, 1–19 (2021). https://doi.org/10.1007/s11227-020-03426-4

11. Abdelhameed, A., Bayoumi, M.: A deep learning approach for automatic seizure detection in children with epilepsy. Front. Comput. Neurosci. **15** (2021). https://doi.org/10.3389/fncom.2021.650050. https://www.frontiersin.org/article/10.3389/fncom.2021.650050. ISSN 1662-5188

12. Xu, Y., Yang, J., Zhao, S., Wu, H., Sawan, M.: An end-to-end deep learning approach for epileptic seizure prediction. In: 2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), pp. 266–270 (2020). https://doi.org/10.1109/AICAS48895.2020.9073988

13. Selim, S., Elhinamy, E., Othman, H., Abouelsaadat, W., Salem, M.A.-M.: A review of machine learning approaches for epileptic seizure prediction. In: 2019 14th International Conference on Computer Engineering and Systems (ICCES), pp. 239–244 (2019). https://doi.org/10.1109/ICCES48960.2019.9068190

14. Burguera, A.: Using machine learning and heart rate variability features to predict epileptic seizures, October 2019

15. Silva, R.S., Plantes Neto, A.R., Brum Marques, J.L., Kavehei, O., Rodrigues, C.R.: A compact QRS detection system based on 0.79 W analog CMOS energy-of-derivative circuit. Microelectron. J. **113**, 105097 (2021). https://doi.org/10.1016/j.mejo.2021.105097. https://www.sciencedirect.com/science/article/pii/S0026269221001087. ISSN 0026-2692

16. Gadhoumi, K., Lina, J.-M., Gotman, J.: Seizure prediction in patients with mesial temporal lobe epilepsy using EEG measures of state similarity. Clin. Neurophysiol. Off. J. Int. Feder. Clin. Neurophysiol. **124** (2013). https://doi.org/10.1016/j.clinph.2013.04.006

17. Chen, H.-H., Cherkassky, V.: Performance metrics for online seizure prediction. Neural Netw. **128**, 22–32 (2020). https://doi.org/10.1016/j.neunet.2020.04.022. https://www.sciencedirect.com/science/article/pii/S0893608020301428. ISSN 0893-6080

18. Goldberger, A.L., et al.: PhysioBank, PhysioToolkit, and PhysioNet. Circulation **101**(23), e215–e220 (2000). https://doi.org/10.1161/01.CIR.101.23.e215. https://www.ahajournals.org/doi/abs/10.1161/01.CIR.101.23.e215

19. Detti, P., Vatti, G., de Lara, G.Z.M.: EEG synchronization analysis for seizure prediction: a study on data of noninvasive recordings. Processes **8**(7) (2020). https://doi.org/10.3390/pr8070846. https://www.mdpi.com/2227-9717/8/7/846. ISSN 2227-9717

20. Höller, Y., Trinka, E., Kalss, G., Schiepek, G., Michaelis, R.: Correlation of EEG spectra, connectivity, and information theoretical biomarkers with psychological states in the epilepsy monitoring unit - a pilot study. Epilepsy Behav. **99** (2019). https://doi.org/10.1016/j.yebeh.2019.106485. https://www.epilepsybehavior.com/article/S1525-5050(19)30470-6/fulltext

21. Romigi, A., et al.: Heart rate variability parameters during psychogenic non-epileptic seizures: comparison between patients with pure PNES and comorbid epilepsy. Front. Neurol. **11** (2020). https://doi.org/10.3389/fneur.2020.00713. https://www.frontiersin.org/article/10.3389/fneur.2020.00713. ISSN 1664-2295

22. Shaffer, F., Ginsberg, J.P.: An overview of heart rate variability metrics and norms. Front. Public Health **5** (2017). https://doi.org/10.3389/fpubh.2017.00258. https://www.frontiersin.org/article/10.3389/fpubh.2017.00258. ISSN 2296-2565

23. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
24. Schölkopf, B., Smola, A., Müller, K.-R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. **10**(5), 1299–1319 (1998). https://doi.org/10.1162/089976698300017467

# Wavelet Coherence Corticomuscular Analysis During Tasks Involved in Object Manipulation

C. D. Guerrero-Mendez$^{(\boxtimes)}$ , C. F. Blanco-Diaz , and T. F. Bastos-Filho

Postgraduate Program in Electrical Engineering, Federal University of Espírito Santo (UFES), Vitória, Brazil
`crguerrero69@uan.edu.co`

**Abstract.** The assessment of corticomuscular connectivity allows the study of brain and muscle connections that exist in a motor movement when an action is executed. In addition, the incorporation of connectivity in the implementation of rehabilitation systems for people suffering from any disability entails performance-related advantages compared to conventional rehabilitation systems. Accordingly, this study aims to quantify the connectivity between the motor cortex through EEG (Electroencephalography) and muscle activation through sEMG (surface Electromyography) during different movements involved in the manipulation of an object (Rest, Reach, Grasp, Lift, Hold, Release, and Replace). For this, a public database (WAY-EEG-GAL) was used to simultaneously record the EEG and sEMG signals of twelve subjects during the manipulation of an object. Connectivity was quantified using wavelet coherence between 2 EEG channels ($C_3$ and $C_4$) and 5 muscles located on the right upper limb. The results show significant differences ($p < 0.05$) between Rest, Reach, Lift and Replace movements when the subject uses a 0.330 kg weight and sandpaper grip surface. The results of this study allow us to quantify the phases of movement in which there is discriminative corticomuscular coherence, which, applied to rehabilitation systems in real conditions, can bring an increase in the detection of individual intention, a greater number of commands, and a better data transfer rate.

**Keywords:** Time-Frequency Coherence · Corticomuscular Analysis · Upper Limb Tasks · Connectivity · Hybrid Brain-Computer Interface (hBCI)

## 1 Introduction

The movement rehabilitation involved mainly in Activities of Daily Living (ADLs) that require a timed and synchronized activation between the brain and muscles, as in the manipulation of an object, is of essential importance in the rehabilitation of people who have suffered some motor impairment in their upper limbs. One of the common activities that are performed in ADLs is object manipulation, as they are found in activities such as lifting, moving, eating,

among others, which require different phases of movements to execute it [1,2]. These include reaching, grasping, lifting, holding, releasing, and replacing the limb [1].

The identification and prediction of continuous motor tasks, such as those encountered in the manipulation of an object, increase the usability, comfort, and controllability of rehabilitation devices, such as neuroprostheses [3]. In addition, proper identification of the different stages of movements allows for lower latency in the response of rehabilitation systems. However, currently, rehabilitation systems such as Brain-Computer Interface (BCI) systems present several challenges in the identification of mental tasks because a single signal modality presents a limited percentage of effectiveness in identifying the user's intent [4,5]. For this reason, the literature has proposed the use of multimodal rehabilitation systems, such as hybrid Brain-Computer Interface (hBCI) systems due to these systems use more than one acquisition mode to identify the intention of movement, generating advantages related to increased performance [4,6].

One way to implement hBCI systems is by estimating the corticomuscular connectivity found between the electrical brain activation information regularly recorded by EEG and the electrical muscle activation recorded by sEMG [7,8]. Cortico-muscular coherence reflects the levels of connection through the auto-spectrum and cross-spectrum of EEG-sEMG signal, allowing to determine the intention of the subjects using the brain-muscle connection [8]. In the literature, studies have quantified connectivity in the manipulation of an object at two different weights [7]. Additionally, cortico-muscular connectivity has been reported during sustained motor contraction tasks [8,9], and continuous movements [9,10]. Other studies have studied corticomuscular connectivity in finger movement and holding movements with different forces [5,11].

To our knowledge, the differences between EEG-sEMG connectivity in each of the movements that occur individually in the development of a multi-articular activity that requires continuous and synchronized activations such as the manipulation of an object involving tasks of resting, reaching, grasping, lifting, holding, releasing, and replacing have not currently been quantified. Therefore, this work focuses on quantifying EEG-sEMG connectivity in each of the movements that are present in the manipulation of an object using wavelet coherence for time-frequency domain representations. These temporal and spectral representations are important in the scientific community because significant results in movement identification have been obtained by segmentation in these domains [12]. Additionally, determining the differences between the phases that occur in a movement, such as manipulation of an object, helps the implementation of rehabilitation systems in real conditions, as well as the decreased latency in the response of these systems. For this, a public database was used, which records the EEG-sEMG signals in the movement of the manipulation of an object. In addition, wavelet coherence was implemented, which estimates the connectivity between the two signals by making use of the wavelet auto-spectra and cross-spectra of the signal in time-frequency domain. As a result, a characterization of the coherence is obtained, where significant differences were found between the phases of the manipulation of an object ($p < 0.05$).
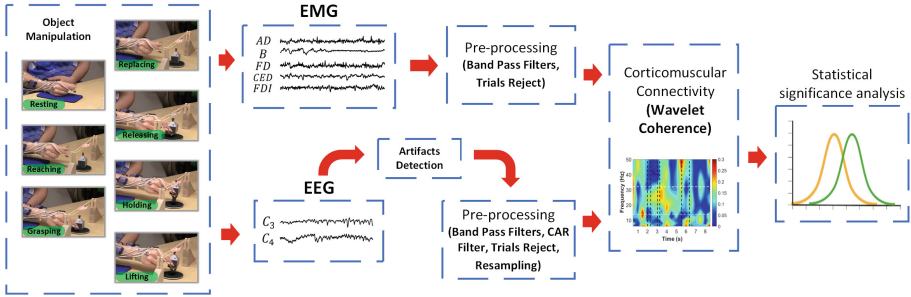
**Fig. 1.** Methodology implemented for the estimation of corticomuscular connectivity between EEG and sEMG signals.

## 2     Materials and Methods

### 2.1     Object Manipulation Dataset

WAY-EEG-GAL is a free and open dataset consisting of EEG and sEMG recordings as well as 3D hand and object position measurements [2]. The dataset records twelve healthy right-handed subjects (8 females and 4 males, aged 19–35 years) using 32 EEG channels located superficially on the cerebral cortex according to the 10–20 international system, and five sEMG channels in the following muscles: Anterior Deltoid (AD), Brachioradialis (B), Flexor Figitorum (FD), Common Extensor Digitorum (CED), and First Dorsal Interosseous (FDI). sEMG signals were recorded with five surface sensors at a sampling rate of 4 kHz, and EEG signals were recorded using the ActiCap device at a sampling rate of 500 Hz. Subjects in the experiment performed manipulation of an object, varying randomly in 3 different weights (0.165, 0.330, and 0.660 kg) and on three different contact surfaces (sandpaper, chamois, silk). Initially, subject remained seated with the right arm resting on a table. After 2 s, a visual cue was presented by an LED for subjects to start performing the movement. In this, the subjects performed a reaching movement, then the grasping of the object with the index and thumb fingers, and then the lifting and holding for 2 s at a height of 5 cm. After 2 s, the subject released the object and replaced the arm to its initial position. Finally, 10 series of approximately 32 trials were recorded, for a total of 328 trials per subject.

Five series of weights were used for this study, where each series included 11 trials, for a total of 55 trials per participant. Taking into account the lobe where most of the changes produced by upper limb motor tasks are recorded, EEG channels $C_3$ and $C_4$ were used together with 5 sEMG channels, because the multi-joint nature of the task generates changes in the variation of muscle activity according to each phase of the movement. For this reason, it is necessary to study the largest number of muscles available. Finally, the information recorded during the manipulation of the object at the weight of 0.330 kg while maintaining the contact surface on the sandpaper was used.

## 2.2   Methodology Overview

The methodology for connectivity estimation using wavelet coherence is presented in Fig. 1. First, each of the movements involved in the object manipulation is segmented, then, artifact detection is performed, and channels and subjects are selected. Subsequently, signal preprocessing is implemented, and by means of wavelet coherence, corticomuscular connectivity is estimated and significant values are evaluated. Finally, a statistical study is performed to evaluate the differences in connectivity between movements.

## 2.3   Pre-processing

To estimate corticomuscular connectivity, artifact detection was performed using threshold evaluation criteria. For this purpose, only the EEG signal in channels $C_3$ and $C_4$ was taken into account, where it was evaluated that at each time instant for each trial and subject the signal did not exceed the threshold of $\pm$ 350 $uV$ [13]. If the signal exceeded this threshold, the trial in the two EEG-sEMG signals was rejected to maintain synchronization in the connectivity study, i.e., if a trial was rejected for channel $C_3$, it was also rejected for channel $C_4$ and the sEMG channels. For the selection of subjects, the percentage of eliminated trials was taken into account, i.e., if in each subject the total of eliminated trials exceeds 10%, the subject is rejected [7]. In this study, no subject was eliminated.

Subsequently, an $8th$ order Butterworth band-pass filter was applied between the frequency range 6–50 Hz to perform the analysis considering the brain frequency bands alpha ($\alpha$, 8–13 Hz), beta ($\beta$, 14–32 Hz), and gamma ($\gamma$, 33–50 Hz) in the two EEG-sEMG signals, in each trial and subject. Additionally, a Common Average Reference (CAR) filter was applied to the EEG signal to eliminate the related noise in all electrodes located on the motor cortex. Then, a resampling was implemented on the EEG signal at 4 kHz to make the samples between the EEG-sEMG signals similar [14]. In this study, all movements involved in the manipulation of an object (Rest, Reach, Grasp, Lift, Hold, Release and Replace) were selected. The segmentation of these movements took into account the information provided by the database of the positions of the hand and the object, as well as the times in which the LED was turned on and off.

## 2.4   Wavelet Coherence

Coherence is a measure of frequency correlation between two signals, which is used to determine the strength of the connection between brain and muscle activity in the execution of a movement by generating correlation values in the range of 0–1 [8]. In this study, the connectivity estimation is performed using wavelet coherence, which uses the calculation of the wavelet auto-spectrum and cross-spectrum of EEG-sEMG signals. For this purpose, the wavelet transform is used, which allows extending the connectivity analysis thanks to the representation of the signal components in the time-frequency domain. Several studies

have introduced the wavelet transform and its use in connectivity estimation [14–16]. The calculation of the wavelet coherence is presented in Eq. 1.

$$WCoh_{x,y}(t,f) = \frac{|SW_{x,y}(t,f)|^2}{SW_{x,x}(t,f) \cdot SW_{y,y}(t,f)} \tag{1}$$

where $SW_{x,x}(t,f)$, and $SW_{y,y}(t,f)$ are the wavelet auto-spectrum of the EEG signal ($x$) and the sEMG signal ($y$) taking into account time ($t$) and frequency ($f$), respectively. On the other hand, $SW_{x,y}(t,f)$ is the cross-spectrum between the EEG-sEMG signals. The wavelet spectrum is calculated according to Eq. 2.

$$SW_{x,y}(t,f) = \int_{t-\delta/2}^{t+\delta/2} MW_x(a,b)MW_y(a,b)da \tag{2}$$

where $MW_x(a,b)$ and $MW_y(a,b)$ are the wavelet Morlet Transforms of the EEG (x) and sEMG (y) signals calculated by means of the Eqs. 3 and 4, respectively. $\delta$ is the size of the integration window, which is dependent on the frequency calculated, according to [16].

$$MW_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^* \left( \frac{t-b}{a} \right) dt \tag{3}$$

$$MW_y(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} y(t)\psi^* \left( \frac{t-b}{a} \right) dt \tag{4}$$

where $a$ denotes the wavelet scale that is inversely proportional to the frequency, $b$ is the smoothing parameter, $t$ is the local time origin of the wavelet analysis, $*$ is the conjugate, and $\psi(\frac{t-b}{a})$ represents the wavelet basis function [14].

In this study, the frequency spectrum used is from 6 to 50 Hz with a frequency bin resolution of 2 Hz. Signals were used from 0.5 s to the end of the task execution, approximately 9.5 s, with sliding Hanning time windows with time steps of 200 ms. Finally, 3 cycles were used for the wavelet estimation, according to [16].

**Coherence Threshold Estimation.** To extract significant coherence values, a significance threshold was calculated using Eq. 5, where $\varepsilon$ corresponds to the 95% confidence level and $L$ to the number of time windows. The evaluation of significant coherence is an important parameter because it determines from what level of connectivity there really is a strong connection between the brain and muscles [7,14].

$$Confidence\ Limit = 1 - (1-\varepsilon)^{\frac{1}{L-1}} \tag{5}$$

Therefore, if the coherence values are above the threshold, the coherence between EEG-sEMG signals is considered significant.

## 2.5   Statistical Significance Analysis

A statistical-significance study was performed to determine differences between the movements involved in the manipulation of an object in the estimation of connectivity using EEG-sEMG channels. For this, first the distribution and homogeneity of the connectivity data were evaluated using the Shapiro-Wilk and Levene tests. Then, the two-sample t-test were applied with a $p-value$ of 0.05 to compare the different phases of movement. The alternative hypothesis is that there are different phases of movement that have greater connectivity with respect to others and the null hypothesis is that there are not differences.
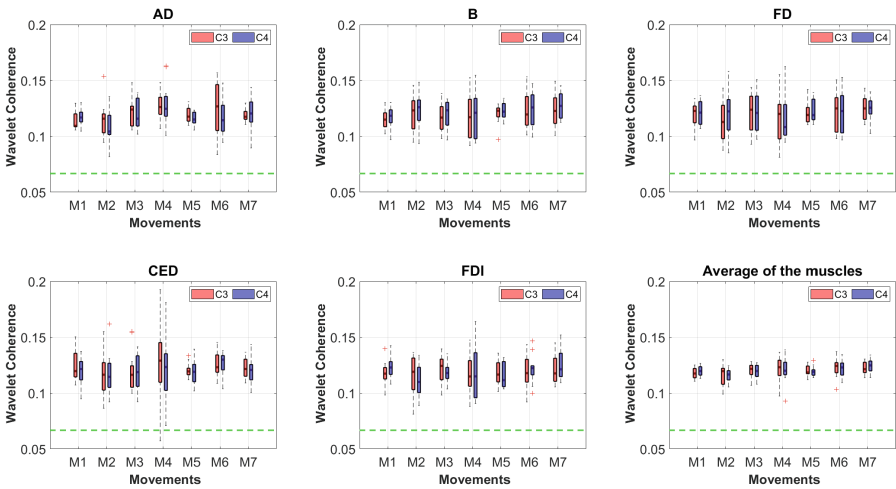


**Fig. 2.** Corticomuscular connectivity in all subjects and trials according to the 2 EEG channels ($C_3$ (red color), $C_4$ (blue color)), and 5 muscles recorded, segmenting the movements involved in manipulating an object; Rest (M1), Reach (M2), Grasp (M3), Lift (M4), Hold (M5), Release (M6), and Replace (M7). The green line corresponds to the threshold of significant connectivity. (Color figure online)

## 3   Results

The wavelet coherence estimation is presented in Fig. 2. For this, the 7 movements involved in object manipulation, which are Rest (M1), Reach (M2), Grasp (M3), Lift (M4), Hold (M5), Release (M6), and Replace (M7), were segmented taking into account that in this type of tasks the movements are executed continuously. In addition, the estimated connectivity of each EEG channel, $C_3$ (red) and $C_4$ (blue), in the 5 recorded muscles is presented, as well as the connectivity of each EEG channel averaged over the 5 sEMG channels for the 7 movements. Additionally, according to the evaluation of significant coherence, the threshold calculated for the average of trials and subjects is presented in green color.

Time-frequency representations of the estimated connectivity by wavelet coherence are presented in Fig. 3. In this figure, the distribution of the estimated connectivity for a subject in the connection of $C_3$ and $C_4$ with each of the recorded muscles is presented. In the figures, the $\alpha$, $\beta$, and $\gamma$ frequency bands are segmented into white lines, as well as the time at which each movement involved in the manipulation of the object starts or ends. For this, the line between 0.5 and 2 s represents the Rest, the line between 2 and 2.85 s represents the Reach, between 2.85 and 3.85 s is the Grip, between 3.85 and 4.35 s is the Lift, between 4.35 and 6.90 s the Hold, between 6.90 and 7.85 s the Release, and between 7.85 and 9.5 s the Replace of the limb. It is important to mention that these execution times change between subjects. Additionally, the times taken by the subject in the execution of each phase of the manipulation movement of an object are presented.
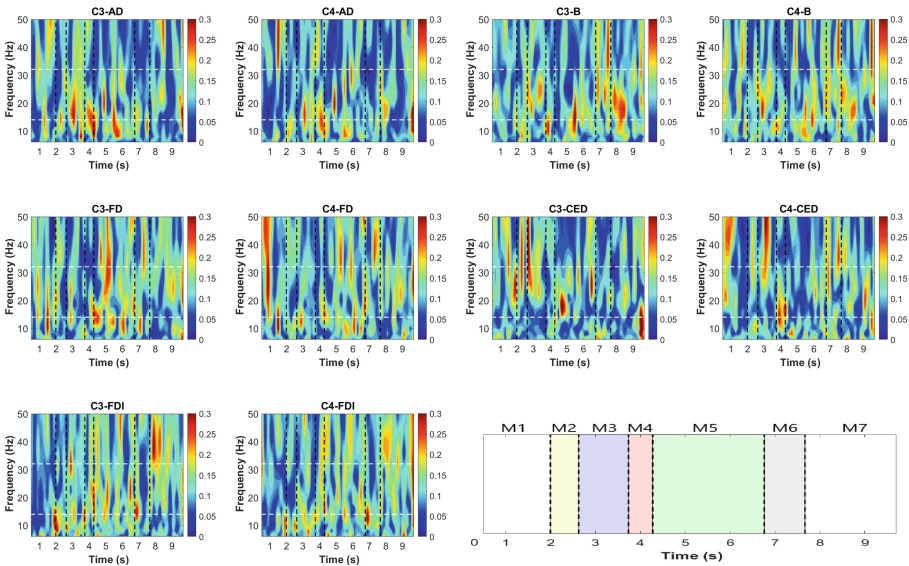


**Fig. 3.** Time-frequency representation of wavelet coherence for a single subject in the 2 EEG channels ($C_3$, $C_4$) and the 5 recorded muscles. Connectivity level from 0 to 0.3. White horizontal lines correspond to segmentation into ($\alpha$), ($\beta$), and ($\gamma$) frequency bands. Black vertical lines correspond to the beginning and end of each movement.

The results of the statistical analysis performed to determine significant differences in connectivity between movements are presented in Table 1. This table shows the movements where significant differences were found as well as in which connectivity estimation channels this difference was present with a value of $p < 0.05$.

**Table 1.** Combination of movements and connectivity channels where statistically significant differences were found with $p < 0.05$.

| Movements | Connectivity |
|---|---|
| M1-M2 | $C_4$-FDI |
| M1-M4 | $C_3$-AD, $C_4$-AD |
| M1-M7 | $C_4$-B |
| M2-M4 | $C_3$-AD, $C_4$-AD |
| M2-M7 | $C_4$-AD, $C_4$-FDI |
| M4-M5 | $C_3$-AD, $C_4$-AD |
| M4-M7 | $C_3$-AD |

## 4   Discussion

The wavelet coherence results shown in Fig. 2 present the variability between trials and subjects quantified for the 7 movements involved in the manipulation of an object. Accordingly, it can be determined that the highest coherence with respect to each EEG channel is found with the AD muscle in M6 (Release), and for the B, FD, CED, and FDI muscles in M4 (Lift). Additionally, by averaging across muscles in each EEG channel, it is found that the highest coherence is present in the Lift movement (M4). Furthermore, according to the significance level, all corticomuscular connectivity values are found to be significant, i.e., brain and muscle connections do not come from random dependence. On the other hand, a comparison between the movements is presented in Table 1 in order to determine if there are significant differences between the quantified connectivity for each movement with a significance level of 0.05, where the main movements in which these differences are found are M1, M2, M4, and M7 in AD, B, and FDI muscles.

According to the time-frequency representations presented in Fig. 3, it is possible to determine the transition of the connectivity that occurs in the continuous movement, as well as the frequency bands where this connectivity is most prominent. Taking into account the variability between subjects, as well as the independence in that each subject takes different times for execution of these tasks, only the results for a single subject are presented. Nevertheless, it is determined that the $\beta$ and $\gamma$ bands present greater coherence depending on the movements and connectivity channels established. On the other hand, it is important to mention that each movement involved in the manipulation of an object takes different execution times, as can be seen in Fig. 3, where the M5 movement presents the longest execution time with approximately 2 s. The coherence between EEG-sEMG signals may vary with increases or decreases in connectivity when presented with tasks that require a longer execution time as presented in M5. On the contrary, in tasks where the execution time is shorter, such as lifting, high coherence contents can be found, which depend on factors such as the connection between channels, frequency bands and subjects.

The results of this study are aligned with works reported in the literature that estimate connectivity in the variation of different weights. The study [7] concludes that connectivity is higher in the beta frequency band and when manipulating objects with a lower weight (0.165 kg) compared to a higher weight (0.660 kg). Additionally, it has been reported that higher levels of connectivity are present in the beta band in motor contraction tasks, and in the gamma band in continuous movements considering increments of muscle contraction [8–10].

## 5    Conclusion

In this study, it was determined how connectivity varies between the movements involved in the manipulation of an object with a weight of 0.330 kg and with sandpaper as a contact surface. Differences between movements and EEG-sEMG connectivity channels are presented, as well as time-frequency representations of connectivity. The findings of this study have great repercussions in the implementation of rehabilitation systems based on hBCI or neuroprosthesis in real conditions, due to the determination of connectivity differences between motor tasks that are continuous and differentiated, which can generate advantages related to greater controllability, greater precision in the identification of movements, ease of use, and decreased latency in the response of these systems.

Future studies will focus on the evaluation of other methods for connectivity estimation in multi-joint tasks, implementation of EEG-sEMG based hBCI systems, and physiological interpretations of the variations between channels, frequency bands, and movements involved.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Betti, S., Zani, G., Guerra, S., Castiello, U., Sartori, L.: Reach-to-grasp movements: a multimodal techniques study. Front. Psychol. **9**, 990 (2018)
2. Luciw, M.D., Jarocka, E., Edin, B.B.: Multi-channel EEG recordings during 3,936 grasp and lift trials with varying weight and friction. Sci. Data **1**(1), 1–11 (2014)
3. Batzianoulis, I., El-Khoury, S., Pirondini, E., Coscia, M., Micera, S., Billard, A.: EMG-based decoding of grasp gestures in reaching-to-grasping motions. Robot. Auton. Syst. **91**, 59–70 (2017)
4. Ma, T., et al.: The hybrid BCI system for movement control by combining motor imagery and moving onset visual evoked potential. J. Neural Eng. **14**(2), 026015 (2017)
5. Kim, B., Kim, L., Kim, Y.-H., Yoo, S.K.: Cross-association analysis of EEG and EMG signals according to movement intention state. Cogn. Syst. Res. **44**, 1–9 (2017)

6. Li, Z., Zhang, S., Pan, J.: Advances in hybrid brain-computer interfaces: principles, design, and applications. Comput. Intell. Neurosci. **2019** (2019)

7. Guerrero-Mendez, C.D., Ruiz-Olaya, A.F.: Coherence-based connectivity analysis of EEG and EMG signals during reach-to-grasp movement involving two weights. Brain-Comput. Interfaces **9**(3), 140–154 (2022)

8. Liu, J., Sheng, Y., Liu, H.: Corticomuscular coherence and its applications: a review. Front. Hum. Neurosci. **13**, 100 (2019)

9. Rossiter, H.E., et al.: Changes in the location of cortico-muscular coherence following stroke. NeuroImage Clin. **2**, 50–55 (2013)

10. Rizzo, R., Zhang, X., Wang, J.W.J.L., Lombardi, F., P.Ch. Ivanov.: Network physiology of cortico-muscular interactions. Front. Physiol. **11**, 558070 (2020)

11. Gao, L., et al.: Enhanced descending corticomuscular coupling during hand grip with static force compared with enhancing force. Clin. EEG Neurosci. **52**(6), 436–443 (2021)

12. Blanco-Diaz, C.F., Antelis, J.M., Ruiz-Olaya, A.F.: Comparative analysis of spectral and temporal combinations in CSP-based methods for decoding hand motor imagery tasks. J. Neurosci. Methods **371**, 109495 (2022)

13. Dimigen, O.: Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. Neuroimage **207**, 116117 (2020)

14. Xi, X., et al.: Enhanced EEG-EMG coherence analysis based on hand movements. Biomed. Signal Process. Control **56**, 101727 (2020)

15. Lachaux, J.-P., et al.: Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence. Neurophysiologie Clinique/Clin. Neurophysiol. **32**(3), 157–174 (2002)

16. Yu, H., Xu, W., Zhuang, Y., Tong, K., Song, R.: Wavelet coherence analysis of muscle coupling during reaching movement in stroke. Comput. Biol. Med. **131**, 104263 (2021)

# Determination of Optical Properties of Skin Tissues Using Spatial Domain Frequency Imaging and Random Forests

B. G. Silva[1](✉) ⓘ, M. R. Gonçalves[1] ⓘ, G. H. S. Alves[2] ⓘ, Á. F. G. Monte[2] ⓘ, and D. M. Cunha[2] ⓘ

[1] School of Electrical Engineering, Federal University of Uberlândia, Uberlândia, Brasil
`bruno.garbim@ufu.br`
[2] Physics Institute, Federal University of Uberlândia, Uberlândia, Brasil

**Abstract.** The evaluation of optical properties of biological tissues has been pointed as an important tool for detection and diagnosis of tissue alterations. The Spatial Frequency Domain Imaging (SFDI) provides quantitative information about light absorption and scattering properties in tissues from measurements of diffuse reflectance. This technique requires the proper correlation between the measured values of diffuse reflectance of light by the tissue, $R_d$, at different spatial frequencies and the corresponding pair of absorption and reduced scattering coefficients $\mu_a$ and $\mu'_s$, respectively. In this work, the machine learning technique of Random Forests was applied to provide a regression model that efficiently computes $\mu_a$ and $\mu'_s$ from $R_d$ values. The database employed consisted of training and testing values of $R_d$ at different spatial frequencies for different combinations of $\mu_a$ and $\mu'_s$, obtained from Monte Carlo simulations. Results showed that the correlation coefficient $R^2$ between predicted and expected values from the test group were 0.96 and 0.97, for $\mu_a$ and $\mu'_s$, respectively. The relative average errors for each coefficient were, respectively, 1% and 0.004%, with standard deviations of 11% and 7%. These results point to the good accuracy and precision of the developed models. These models were applied to an in vivo study, where values of $R_d$ from the dorsal region of the hand of a volunteer were obtained with SFDI equipment using light wavelength of 650 nm. The obtained images of $\mu_a$ and $\mu'_s$ showed enhanced contrast of blood vessels, pointing to the potential of the technique to identify vascular tissue alterations.

**Keywords:** Spatial frequency domain imaging · Machine learning · Random forests

## 1 Introduction

Nowadays, cancer is one of major concerns in public health all over the world. Recent projections estimate a worldwide incidence of 18 million new cases per year [1, 2]. In Brazil, for each year of the triennium 2020–2022, an occurrence of 625 thousand new cases of cancer is estimated. Of this total, non-melanoma skin cancer is the most

incident, corresponding to an estimated total of 177,000 new cases per year [2]. Accurate and early diagnosis of the disease is essential to define the course of treatment, as well as its effectiveness [3]. Diagnostic techniques are of fundamental importance, for example, for the appropriate characterization of actinic keratosis-like lesions in cases of non-melanoma skin cancer, thus contributing to the reduction of mortality from the disease [4].

The Spatial Frequency Domain Imaging (SFDI) technique has emerged as a promising technique for the identification of morphological and functional changes in skin tissues in early stages. This imaging technique allows the evaluation of the properties of absorption and scattering of light by the tissues in a fast and non-invasive way [5]. Image acquisition in SFDI is performed by illuminating the tissue with an incident light beam of spatially modulated intensity at different spatial frequencies, and capturing the reflected beam with a camera. Then, the different obtained images are processed, thus reconstituting the optical properties of the tissues, which are quantitatively measured by the absorption and reduced scattering coefficients $\mu_a$ and $\mu'_s$, respectively. The evaluation of these parameters at different light wavelengths can enhance visualization of different skin structures, making eventual alterations in the tissue evident [5].

The proper correlation of light diffusion properties of a tissue with its optical properties is one of the main steps in SFDI imaging processing. Although pre-compiled look-up tables can be employed to this purpose [6, 7], it has been pointed out that machine learning techniques could accomplish this task more efficiently [8, 9]. In particular, the use of different Artificial Neural Network models has been reported in the literature [10–14]. Recently, Panigraphi et al. [9] showed that the random forest method for regression can also provide optical properties from diffuse reflectance with high accuracy. In particular, the authors considered diffuse reflectance at two spatial frequencies for the determination of values of $\mu_a$ and $\mu'_s$ [9], and the performance of the model was evaluated for noiseless test data. Alternatively, Zhao et al. [15] recently showed that using more than two spatial frequencies can improve the performance of the machine learning models [15] in determining tissue chromophore concentration from noisy diffuse reflectance data.

In this work, we investigate the use of random forests for pattern recognition in the SFDI technique, to determine optical properties of tissues from their light reflection properties. The developed model considered values of diffuse reflectance at three spatial frequencies as input data, and the performance of the model was evaluated considering noisy input values.

## 2   Materials and Methods

### 2.1   Experimental Set-Up for Image Acquisition in SFDI

Figure 1 shows the schematic representation of the experimental set-up employed for image acquisition in SFDI [5]. It comprises a digital projector as light source (model Vivitek® D555WH) a ThorLabs® filter wheel and CCD camera (DCC3240C). A white light beam is emitted from the source, reaching the sample and being reflected by the latter. The spatial distribution of the emitted beam is modulated according to a sinusoidal pattern, at different spatial frequencies $f$ (0.05; 0.2 and 0.4 mm$^{-1}$) [5]. The reflected

beam by the sample is filtered to a given wavelength by the filter, and it is then detected by the camera, which stores the collected signal.
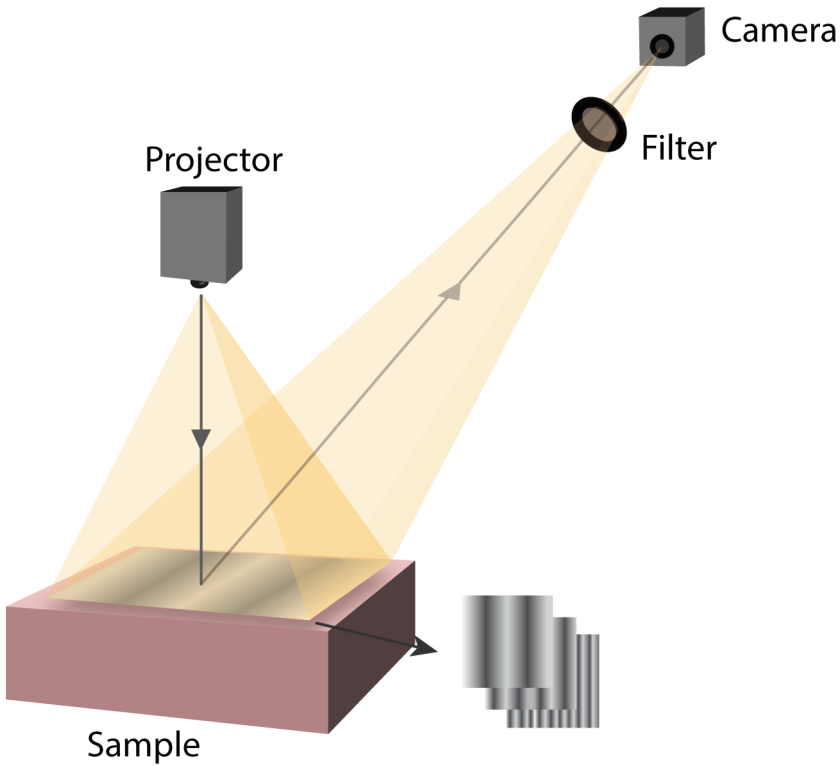


**Fig. 1.** Experimental set-up employed for image acquisition in SFDI.

The collected images by the camera must be processed since they carry dependence not only on the sample composition, but also on the emitted beam and camera properties [6]. To remove these dependencies on the imaging system, the obtained images were demodulated and calibrated with a phantom object, following the methodology proposed by Cuccia et al. [6]. The obtained signal therefore represents the diffuse reflectance, $R_d$, at the different spatial frequencies employed, which is the physical quantity that expresses the dependence of the reflected light on the absorption and scattering properties of the sample [5, 6]. Determination of the sample absorption and reduced scattering coefficients $\mu_a$ and $\mu'_s$ from the obtained $R_d$ values was performed using machine learning methods with random forests, as described in Sects. 2.2 and 2.3.

## 2.2  Generating the Database of Diffuse Reflectance Values

Values of $R_d$ as a function of the pair of coefficients $\mu_a$ and $\mu\prime_s$ were obtained using Monte Carlo (MC) simulations for the transport of light photons through matter. A MC code was developed to simulate the processes of absorption and scattering of light by a uniform material, according to the methodology proposed by Prahl et al. [16]. A material medium was characterized by its coefficients $\mu_a$ and $\mu\prime_s$, and also by its refraction index and anisotropy factor. A pencil beam impinged perpendicularly on the material surface, and the spatial distribution of the backscattered photons was obtained as a point spread function. The values of $R_d$ at different spatial frequencies were then obtained by applying a Fourier Transform on this spread function [6]. Different material compositions were simulated, by varying the values of $\mu_a$ between 0.001 and 1.8 mm$^{-1}$, and $\mu\prime_s$ between 0.3 and 6.3 mm$^{-1}$ [6]. The values of the refraction index and anisotropy factor were fixed at 1.4 and 0.71, respectively [6, 17].

Figure 2 illustrates the contour plot of $R_d$ values obtained from the MC simulations, at a spatial frequency of 0.2 mm$^{-1}$, for different combinations of $\mu_a$ and $\mu\prime_s$.

## 2.3  Random Forest Regression Models

The random forest is an assembly method, based on the concept of decision trees, which can be employed for classification or regression problems [18]. It is a supervised machine learning technique, which combines a large number of decision trees for prediction purposes. For each tree, features and samples are randomly selected from the training data, with replacement, to determine the splitting of each node, in a process called bagging (bootstrap aggregation). For regression problems, the predicted output of the model is computed as the average of the value provided by each tree [19]. The main hyperparameters of a random forest include the number of trees in the forest, the maximum depth of the trees, the minimum number of samples at leaf and minimum number of samples at split. A more detailed description about Random Forests, including mathematical aspects, can be found in the work of Breiman [19].

In this work, the Random Forest technique was employed for regression (RFR), using Python language with the scikit-learn library [20]. For each sample, input data consisted of the $R_d$ values at three different spatial frequencies f (0.05; 0.2 e 0.4 mm$^{-1}$), obtained from the MC simulations, while the output data were the corresponding $\mu_a$ and $\mu\prime_s$ coefficients, as illustrated in Fig. 3.
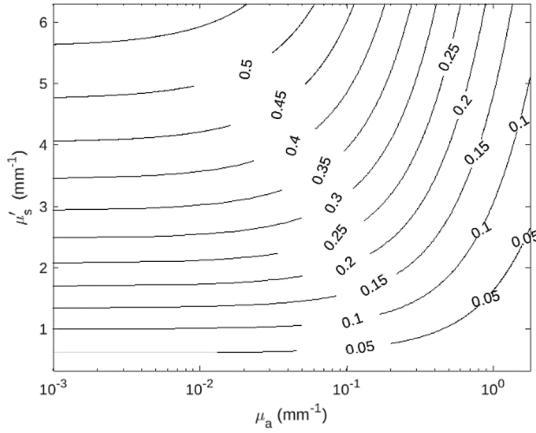
**Fig. 2.** Contour plot of $R_d$ values for different pair values of $\mu_a$ and $\mu'_s$ at spatial frequency of $0.2\ \mathrm{mm}^{-1}$.

The dataset consisted of 108.360 samples. The samples were divided into training and testing data, at the proportions of 67% and 33%, respectively. Besides, for the testing data, a Gaussian noise of 3% was added, to simulate an experimental uncertainty, and also to evaluate the generalization capacity of the trained models [15]. Optimization of the model hyperparameters was performed employing K-fold cross-validation, with 5 subsets and 50 iterations, corresponding to 250 training loops. For the $\mu_a$ coefficient, the following optimum hyperparameters were obtained: the number of trees in the forest was 200, the maximum depth of the trees was set to 21, the minimum number of samples at leaf nodes was 4 and the minimum number of samples at split was 2. For the $\mu'_s$ coefficient, these values were, respectively, 250, 11, 4 and 2. In both cases, bootstrap sampling was used to train the decision trees.

## 2.4   In Vivo Measurements

The experimental set-up for image acquisition, described in Sect. 2.1, and the RFR machine learning models, described in Sects. 2.2 and 2.3, were applied to an in vivo measurement of the absorption and scattering coefficients of the dorsal region of the hand of a volunteer subject (Fig. 4). The rectangular region of interest (ROI) shown in figure indicates the area imaged. Values of $R_d$ at the different spatial frequencies investigated were obtained for the ROI shown in Fig. 4, and the trained RFR model was then employed to compute the corresponding values $\mu_a$ and $\mu'_s$ at each pixel.

The present study in human beings was approved by the Research Ethics Committee of the Federal University of Uberlândia (process no. 85363417.9.0000.5152).
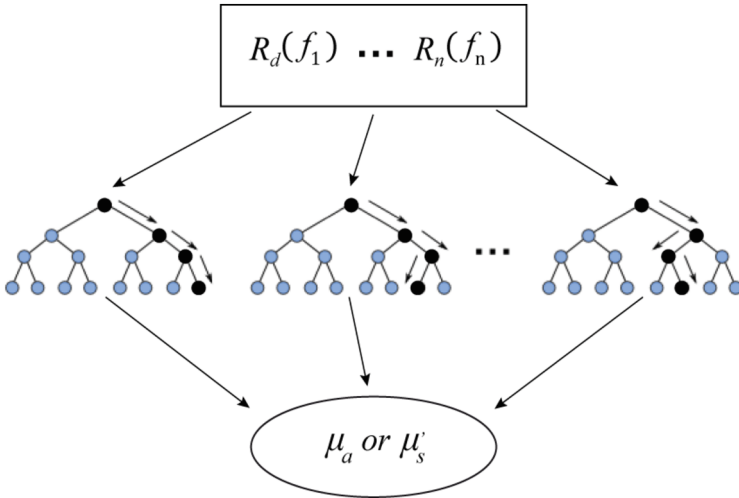
**Fig. 3.** Schematic representation of the RFR models for the determination of $\mu_a$ or $\mu'_s$ coefficients.
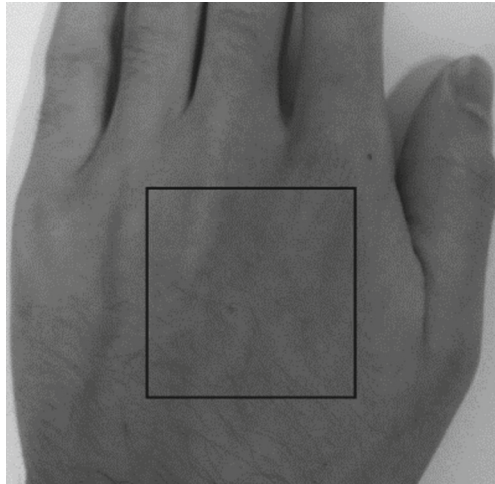


**Fig. 4.** Photography of the dorsal region of the volunteer's hand. The rectangular ROI indicates the area analyzed.

## 3   Results

The results of the application of the trained RFR on the testing data for determination of $\mu_a$ and $\mu'_s$ are shown in Fig. 5. The scatter plots represent the predicted values of $\mu_a$ and $\mu'_s$ versus the corresponding expected values.

Figure 6a and 6b show the histograms of the distribution of relative percent errors for the predicted values of $\mu_a$ and $\mu'_s$, respectively.
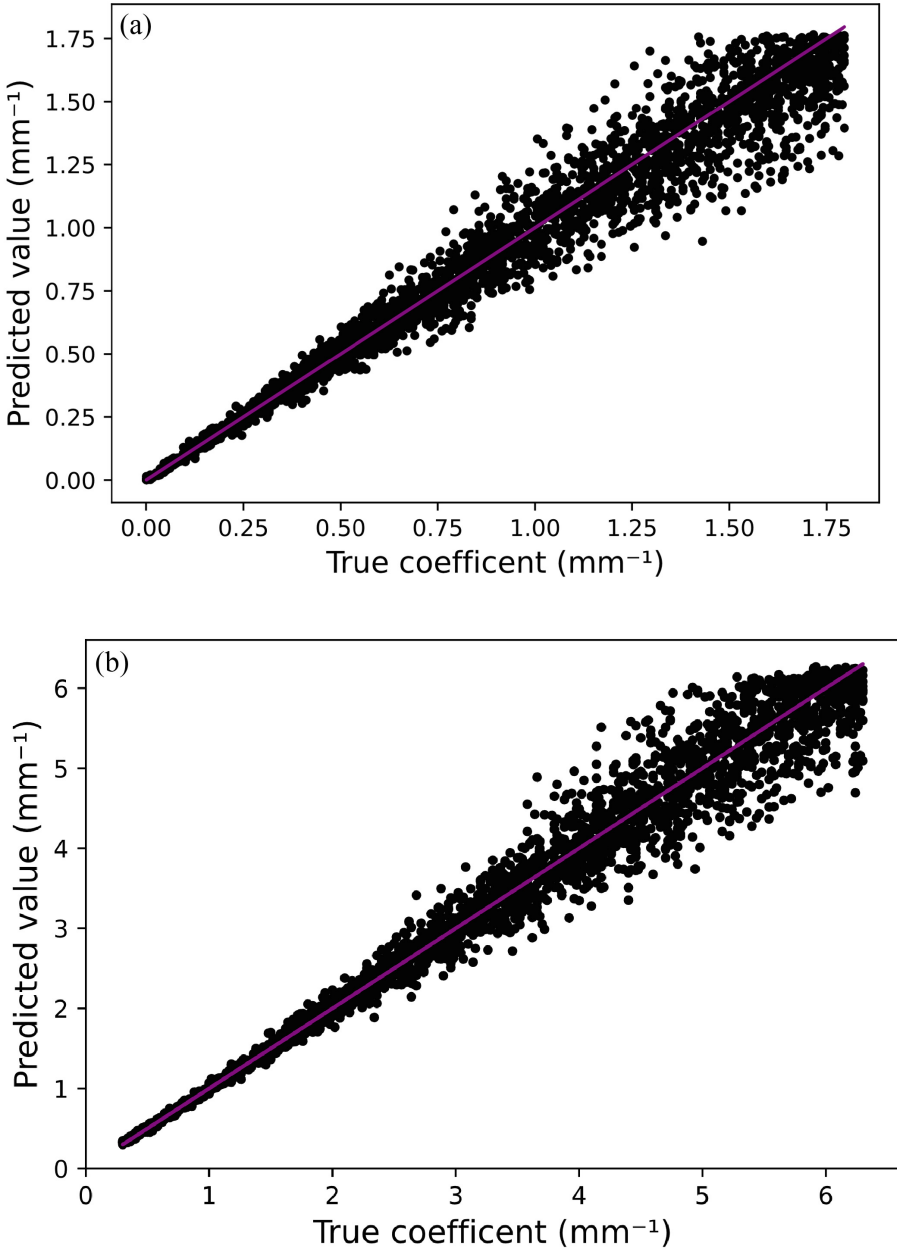
**Fig. 5.** Values of (a) $\mu_a$ and (b) $\mu'_s$ predicted by RFR models, as a function of the true values. Solid line indicates the expected values.
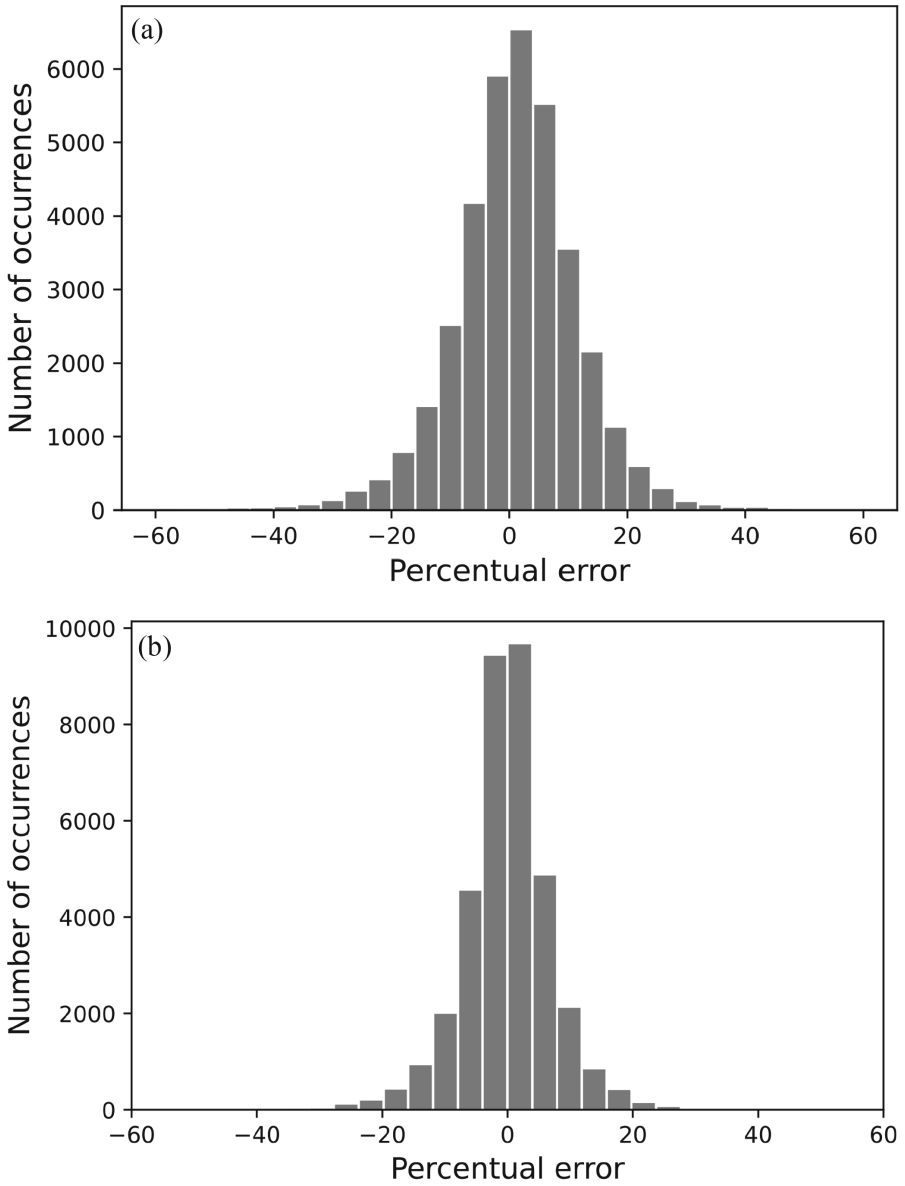
**Fig. 6.** Distribution of relative percent errors for the predicted values of (a) $\mu_a$ and (b) $\mu'_s$, respectively.

Figure 7 shows the results for the in vivo evaluation of $\mu_a$ and $\mu_s\prime$ for the back of the hand, for light wavelength of 650 nm.
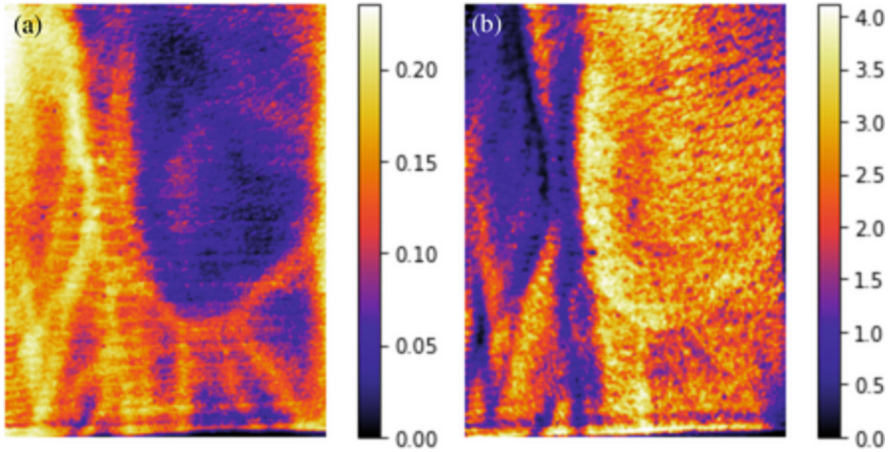
**Fig. 7.** Images of (a) $\mu_a$ e (b) $\mu'_s$ for the ROI analized. Values of both coefficients are given in units of mm$^{-1}$.

## 4   Discussion

Evaluation of the performance of the RFR prediction models for $\mu_a$ and $\mu'_s$ values was performed through the determination of the $R^2$ correlation coefficient for the linear regression between expected and predicted values. A high degree of correlation between these quantities was observed, with values of $R^2$ of 0.96 e 0.97 for $\mu_a$ and $\mu'_s$, respectively.

A Levene's test was applied on the predicted values of $\mu_a$ and $\mu'_s$ showed in Fig. 5, to test for homogeneity of variances. For each coefficient, predicted values were divided in three subgroups, each one corresponding to a different interval within the whole range of values. The variances of these subgroups were then compared among each other. Results showed that, for each coefficient, there were no statistically significant differences between variances among subgroups, considering the significance level of 0.05. This result indicates that the increase in the errors observed in Fig. 5a and 5b is actually a graph scale effect, due to differences between the orders of magnitude of the values of each coefficient, and the performance of the RFR model is independent on the coefficient values.

From Fig. 6, it can be observed that the error distributions show symmetry around the null error value, indicating that the obtained models provided unbiased results. The average percent errors for these distributions were 1% and 0.004%, for $\mu_a$ and $\mu'_s$, respectively, with standard deviations of 11% and 7%. These results indicate that, even in the presence of noisy input data, the models show good accuracy and precision in the determination of $\mu_a$ and $\mu'_s$.

Compared to the literature, values of average percent errors obtained in this work were higher than those reported by Panigraphi et al. [9]. Nevertheless, it should be pointed out that the authors did not considered noise in their input test data. By neglecting noise, our model provides performance comparable to those reported by the authors. Results were also comparable to those reported by Song et al. [14], who considered a deep

neural network model for the determination of optical properties, for test data with 2% Gaussian noise.

Figure 7 shows that the developed models, combined with the SFDI system, are capable of providing values of $\mu_a$ and $\mu'_s$ for in vivo measurements. In particular, at a light wavelength of 650 nm, Fig. 7a shows that values of the absorption coefficient highlight blood vessels. This effect occurs due to the higher contribution of oxy- and deoxyhemoglobin to the absorption coefficients at this wavelength [21], and it could be useful to identify vascular changes associated with melanocytic or nonmelanocytic skin cancers [22].

## 5   Conclusions

In this work, the machine learning method of Random Forests for regression (RFR) was applied to the determination of light absorption and scattering properties of tissues from values of light diffuse reflectance in the SFDI technique. The developed models showed good performance in determining $\mu_a$ and $\mu'_s$ coefficients from $R_d$ values at different spatial frequencies. The average percent errors were 1% and 0.004% for $\mu_a$ e $\mu'_s$, respectively, with standard deviations of 11% and 7%. The RFR models were applied to an in vivo measurement for the determination of light absorption and scattering coefficients of the back of a hand. In particular, values of $\mu_a$ obtained for light wavelength of 650 nm provided enhanced contrast of blood vessels. These results point to the potential of this imaging technique for detection and identification of vascular changes in skin tissues, which could be related to different types of skin tumors. Future works should be conducted, e.g. using tissue equivalent materials with known optical coefficients, in order to compare the estimated coefficients with true tissue values, and thus evaluating the accuracy of the technique for different skin tissue compositions.

## References

1. Bray, F., Ferlay, J., Soerjomataram, I., et al.: Global cancer statistics: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **68**(6), 394–424 (2018)
2. INCA (2021). https://www.inca.gov.br/numeros-de-cancer. Accessed 29 Aug 2022
3. Travers, J.B., Poon, C., Rohrbach, D.J., et al.: Noninvasive mesoscopic imaging of actinic skin damage using spatial frequency domain imaging. Biomed. Opt. Express **8**(6), 3045–3052 (2017)
4. Reinehr, C.P.H., Bakos, R.M.: Actinic keratoses: review of clinical, dermoscopic and therapeutic aspects. An. Bras. Dermatol. **94**(6), 637–657 (2019)
5. Cuccia, D.J.: Spatial frequency domain imaging (SFDI): a technology overview and validation of an LED-based clinic friendly device. In: Douglass, M.R., Oden, P.I. (eds.) CONFERENCE 2012, Emerging Digital Micromirror Device Based Systems and Applications IV, vol. 825405, pp. 1–6. Proc. SPIE, San Francisco (2012)
6. Cuccia, D.J., Bevilacqua, F., Durkin, A.J., et al.: Quantitation and mapping of tissue optical properties using modulated imaging. J. Biomed. Opt. **45**, 5–32 (2009)
7. Hennessy, R.J., Lim, S.L., Markey, M.K., et al.: Monte Carlo lookup table-based inverse model for extracting optical properties from tissue-simulating phantoms using diffuse reflectance spectroscopy. J. Biomed. Opt. **18**(3), 037003–1–037003–4 (2013)

8. Jacques, S.L.: Optical properties of biological tissues: a review. Phys. Med. Biol. **58**(11), R37–R61 (2013)
9. Panigrahi, S., Gioux, S.: Machine learning approach for rapid and accurate estimation of optical properties using spatial frequency domain imaging. J. Biomed. Opt. **24**, 071606–1–071606–6 (2018)
10. Yudovsky, D., Nguyen, J.Q.M., Durkin, A.J.: In vivo spatial frequency domain spectroscopy of two layer media. J. Biomed. Opt. **17**(10), 071606–1–071606–6 (2012)
11. Wang, Q., Le, D., Roman, J.R., et al.: Broadband ultraviolet-visible optical property measurement in layered turbid media. Biomed. Opt. Express **3**(6), 1226–1240 (2012)
12. Jäger, M., Foschum, F., Kienle, A.: Application of multiple artificial neural networks for the determination of the optical properties of turbid media. J. Biomed. Opt. **18**(5), 071606–1–071606–6 (2013)
13. Tsui, S.Y., Wang, C.Y., Huang, T.H., et al.: Modelling spatially-resolved diffuse reflectance spectra of a multi-layered skin model by artificial neural networks trained with Monte Carlo simulations. Biomed. Opt. Express **9**(4), 1531–1544 (2018)
14. Song, B., Jia, W., Zhao, Y., et al.: Ultracompact deep neural network for ultrafast optical property extraction in spatial frequency domain imaging (SFDI). Photonics **9**(5), 1–12 (2022)
15. Zhao, Y., Deng, Y., Yue, S., et al.: Direct mapping from diffuse reflectance to chromophore concentrations in multi-fx spatial frequency domain imaging (SFDI) with a deep residual network (DRN). Biomed. Opt. Express **12**(1), 433–443 (2021)
16. Prahl, S.A., Keijzer, M., Jacques, S.L., et al.: A Monte Carlo model of light propagation in tissue. Proc. SPIE Int. Soc. Opt. Eng. **10305**, 102–111 (1989)
17. Yudovsky, D., Durkin, A.J.: Spatial frequency domain spectroscopy of two layer media. J. Biomed. Opt. **16**(10), 107005–1–107005–10 (2011)
18. Müller, A.C., Guido, S.: Introduction to Machine Learning with Python: A Guide for Data Scientists. 1nd edn. O'Reilly Media, Sebastopol, United States of America (2016)
19. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
20. Pedregosa, F., Varoquaux, G., Gramfort, A., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res .**12**, 2825–2830 (2011)
21. Mazhar, A., Dell, S., Cuccia, D.J., et al.: Wavelength optimization for rapid chromophore mapping using spatial frequency domain imaging. J. Biomed. Opt. **15**(6), 107005–1–107005–10 (2010)
22. Argenziano, G., Zalaudek, I., Corona, R., et al.: Vascular structures in skin tumors: a dermoscopy study. Arch. Dermatol. **140**(12), 1485–1489 (2004)

# Understanding Brain Magnetic Resonance Images from Automatically Generated Interval-Valued Membership Functions

Diego S. Comas[1,2]($\boxtimes$) ⓘ, Gustavo J. Meschino[3] ⓘ, and Virginia L. Ballarin[1] ⓘ

[1] Image Processing Lab, Institute of Scientific and Technological Research in Electronics (ICyTE), National University of Mar del Plata-CONICET, Mar del Plata, Argentina
`diego.comas@fi.mdp.edu.ar`

[2] National Council of Scientific and Technical Research (CONICET), Buenos Aires, Argentina

[3] Bioengineering Lab, Institute of Scientific and Technological Research in Electronics (ICyTE), National University of Mar del Plata-CONICET, Mar del Plata, Argentina

**Abstract.** Medical images are representations of tissues and parts of the human body which play a crucial role in diagnosis assistance and human anatomy examination. New processing requirements related to problems of classification or segmentation are unceasingly generated. In this context, methods for segmentation that enables interpretable knowledge discovery can lead to significant contributions to the study and solution of certain medical problems. In a previous work, we proposed a data classification method called Type-2 Label-based Fuzzy Predicate Classification (T2-LFPC) which automatically generates interval-valued membership functions and predicates. In the present work, a methodology for interpreting brain magnetic resonance images in sequences PD, T1, and T2 with different levels of additive noise is proposed. Three measures on interval-valued membership functions are proposed and analyzed. Both simulated and real images are considered. The segmentation performance is consistent with the obtained with the test methods. The major contributions are a) the definition of attributes on the features and the association of them to each tissue, b) the description of relationships between attributes and tissues providing linguistic interpretation, c) the identification, quantification and description of both vagueness associated with the attributes and spread of intensities of pixels belonging to each tissue. Nevertheless, the knowledge achieved is consistent with what is known in the field of brain magnetic resonance images, which indicates that the methodology proposed constitutes a sound approach for knowledge discovery. Therefore, it could be extended to other medical imaging domains, making it a general approach for understanding medical images.

**Keywords:** Brain magnetic resonance images · Fuzzy membership functions · Knowledge discovery · Segmentation

# 1 Introduction

Medical image analysis assists in diagnosis and human anatomy examination, being the specialist who is able to discover and interpret the information in images containing representations of tissues and parts of the human body acquired through different technologies (modalities) [1].

The technological advance of the past decades brought new modalities and improved the existing ones, substantially increasing the available information [1, 2]. Because of this, new processing requirements related to new problems of segmentation are unceasingly generated [3]. The goal of a segmentation method is partitioning images looking for regions, but it is also a requirement to understand how the algorithm works in order to discover knowledge both about the problem and the image modality. In this sense, any method which not only solves the segmentation but, also, allows to discover interpretable knowledge, can lead to significant contributions to the study and solution of certain medical problems. In this context, interpretable knowledge is that expressed in a language comprehensible to humans.

Image segmentation can be addressed through pixel classification. In previous works [4, 5], we proposed a classification method called Type-2 Label-based Fuzzy Predicate Classification (T2-LFPC) which automatically generates interval-valued membership functions and fuzzy predicates from labelled data. From the outcome of the T2-LFPC method, in the present work it is proposed a methodology for interpreting brain Magnetic Resonance Images (MRI) in sequences PD, T1, and T2 with additive noise. Three measures on interval-valued membership functions are proposed and analyzed. As a result, understanding of the membership functions is achieved, allowing to define attributes and to associate them with tissues. In addition, it is possible to quantify and describe both vagueness associated with the attributes and spread of intensity of pixels belonging to each tissue.

# 2 Material and Methods

## 2.1 Type-2 Label-Based Fuzzy Predicate Classification (T2-LFPC)

The next definitions are adopted in this paper:

*Definition #1:* An interval-valued membership function $\overline{\mu}_{\mathbf{A}}$ on a discourse universe $X$ is a function $\overline{\mu}_{\mathbf{A}} : X \rightarrow \chi$, being $\mathbf{A}$ a property (an attribute) and $\chi$ the set of all the closed intervals contained in [0, 1] [6]. For a specific value $x \in X$, $\overline{\mu}_{\mathbf{A}}(x)$ is an interval of truth values $\overline{\mu}_{\mathbf{A}}(x) = [\varphi^-_{\overline{\mu}_{\mathbf{A}}}(x), \varphi^+_{\overline{\mu}_{\mathbf{A}}}(x)]$ which defines the degree of truth in which $x$ satisfies the property $\mathbf{A}$. The functions $\varphi^-_{\overline{\mu}_{\mathbf{A}}} : X \rightarrow [0, 1]$ and $\varphi^+_{\overline{\mu}_{\mathbf{A}}} : X \rightarrow [0, 1]$ are type-1 membership functions respectively called Lower Membership Function (LMF) and Upper Membership Function (UMF) [6].

*Definition #2:* The Footprint of Uncertainly (FOU) of an interval-valued membership function $\overline{\mu}_{\mathbf{A}}$ is [6, 7]:

$$\text{FOU}_{\overline{\mu}_{\mathbf{A}}} = \bigcup_{x \in X} \left\{ \left[ \varphi^-_{\overline{\mu}_{\mathbf{A}}}(x), \varphi^+_{\overline{\mu}_{\mathbf{A}}}(x) \right] \right\}. \tag{1}$$

The size of the FOU (its area) is related to the vagueness or imprecision around the attribute described by the interval-valued membership function [6, 8, 9].

The method T2-LFPC consists of four stages: A) random partition of the data, B) extraction of class prototypes, C) generation of a fuzzy predicates system, and D) optimization. A detailed formalization of the method can be found in [4, 5]. The stages A and B are focused in providing prototypes for each label in the Gold-Standard. First, a random dataset partition is applied to the dataset (stage A) generating subsets as descriptors, capturing diversity in the data, and reducing the computational cost of the next stages. At stage B, prototypes of each label, which play a key role in the discovery of characteristics about the classification problem, are extracted from an automatic clustering scheme [6, 8] capturing collections of common attributes in the data in each label. The stage C is focused on generating both membership functions and predicates, relating attributes of the features with properties observed in the prototypes. At stage D, an optimization of the parameters is performed.

## 2.2   Method Proposed for Understanding Brain Magnetic Resonance Images

The outcomes of the method T2-LFPC are interval-valued membership functions and predicates describing the classes and enabling data classification. It is possible understanding them identifying how the attributes are combined to form classes and, more important, how the data with those attributes are. It is feasible to study the shape, width, and FOU of the interval-valued membership functions for describing vagueness around the attributes (spread of the data within the classes). Also, considering the relative position of the membership functions in the feature domain, functions can be associated to attributes like "high", "medium", "low". Based on these considerations, a methodology for understanding brain MRI is proposed in the next paragraphs.

The next three measures on interval membership functions are proposed:

1. *max_MF*: Value of the feature for the maximum of UMF, mapping in the scale of the feature (in the MRI cases it is [0, 255]). It indicates the value of the feature who better satisfy the attributes of the class.
2. *diff_MF*: Difference between the values of the feature where the UMF is equal to 0.5. It is a measure of the width of the membership function associated to the spread of the values around the attribute. Small, medium, and high width means low, medium, and high spread respectively (imprecision for describing the values of the feature related to a label). While this measure could take values in the range of the feature [0, 255], in real cases it is expected values in [0, 150].
3. *area_MF*: The area of the FOU. It is a measure of vagueness about the definition of the attribute. A small area means low vagueness (meaning it is possible to be more precise when explaining the grade in what the attribute is met by the feature and its relationship with the label), a medium area means average vagueness (it is quite difficult to relate that meaning of the attribute), and a large area means high vagueness (high difficulty to define a meaning). While this measure maps on to $\mathbb{R}$, in real cases it is expected values in [0, 25].

Typically, when analyzing brain MRI, expert medical doctors' study gray intensities of images in sequences PD, T1, and T2 for assigning tissues to pixels. In this context,

experts' knowledge is expressed as linguistic descriptions, using attributes for the gray intensity such as "bright", "dark", "gray", and so on. Based on the previous considerations and the measures defined, the following methodology for understanding brain MRI is proposed:

1. Apply the T2-LFPC to a dataset containing samples of brain MRI in sequences PD, T1, and T2 and Gold-Standard.
2. Assign attributes relating each tissue with each feature improving interpretability: it is possible to map generic attributes "high", "medium", "low" (defined by visual inspection of the position of the membership functions) to experts' domain terminology as follow: very low = black, low = dark, slightly low = dark-gray, medium = gray, slightly high = light-gray, high values = bright, very high = very-bright.
3. Evaluate values of *max_MF*, *diff_MF*, and *area_MF* to quantify and describe levels of imprecision and vagueness. Perform comparison between values for function of different classes and features.

### 2.3   Datasets and Validation Methodology

The analysis was focused on two brain MRI datasets:

- *Dataset #1:* Simulated brain MRI [10], 4000 pixels randomly selected per class (12000 data, 3 classes, 3 features). Data were taken without any noise or distortion and were generated by means of computer simulation.
- *Dataset #2*: 10000 pixels per class randomly selected from real brain MRI (30000 data, 3 classes, 3 features). MRI were obtained at the Dementia Clinic of the Institute for Neurological Research "Raúl Carrea" (Buenos Aires, Argentina) with a 1.5 T system with the next protocol: coronal 3D T1-weighted gradient echoes orthogonal to the AC-PC line (TR/TE = 24/5 ms, slice thickness = 1.5 mm); and coronal proton density (PD) and T2-weighted fast spin echoes oriented (TR/TE1/TE2 = 3,500/32/96 ms, echo train length = 8, slice thickness = 3 mm). Pixels were classified using the software BRAINS [11] optimized by medical experts' criteria.

Features are intensities in sequences PD, T1, and T2. Independent-zero-mean value-gaussian-noise were considered (0%, and 10%), being 100% a *standard deviation* = 256 (images in gray levels).

The methods T1-LFPC (a variant of the T2-LFPC using type-1 membership functions) [4, 5], Probabilistic Neural Networks (PNN) [12], Multi-Layer Perceptrons (MLP) [13], and k-Nearest Neighbors (KNN) [14] were used for tests. The segmentation performance was estimated using 10-fold cross validation. Statistical tests of significance were included considering the Wilcoxon signed rank test for paired data [15].

## 3   Results

Results are presented split in dataset #1 and #2. Error, Tanimoto and Dice measures are reported as performance measures. The best of the test methods is indicated including relative difference and value of the test of significance (indicated as P). For dataset #1 (simulated brain MRI), Table 1 summarizes the segmentation results. In Fig. 1, interval-valued membership functions are shown. Table 2 shows the results of the proposed measures applied those membership functions.

**Table 1.** Results for the segmentation of dataset #1 (simulated brain MRI).

| Noise | Measure | T2-LFPC | Best of the test methods | Difference |
|-------|---------|---------|--------------------------|------------|
| 0% | Error | $0.019 \pm 0.003$ | $0.015 \pm 0.003$ (KNN) | 28.492% P = 0.008 |
| | Tanimoto | $0.962 \pm 0.007$ | $0.970 \pm 0.005$ (MLP) | −0.752% P = 0.008 |
| | Dice | $0.981 \pm 0.003$ | $0.985 \pm 0.003$ (MLP) | −0.385% P = 0.012 |
| 10% | Error | $0.086 \pm 0.005$ | $0.085 \pm 0.009$ (T1-LFPC) | 0.586% P = 0.503 |
| | Tanimoto | $0.844 \pm 0.008$ | $0.844 \pm 0.009$ (MLP) | −0.029% P = 0.751 |
| | Dice | $0.914 \pm 0.005$ | $0.915 \pm 0.005$ (T1-LFPC) | −0.127% P = 0.916 |

**Table 2.** Results of the measures on interval-membership functions proposed considering segmentation of dataset #1 (simulated brain MRI). (CSF = Cerebrospinal fluid, WM = White matter, GM = Gray matter)

| | Measure | PD Intensity | | | T1 Intensity | | | T2 Intensity | | |
|---|---------|------|------|------|------|------|------|------|------|------|
| | | CSF | GM | WM | CSF | GM | WM | CSF | GM | WM |
| noise 0% | max_MF | 235.12 | 216.94 | 187.65 | 53.74 | 108.91 | 144.89 | 215.14 | 107.98 | 74.15 |
| | diff_MF | 41.08 | 15.19 | 13.27 | 29.90 | 24.07 | 15.53 | 70.16 | 43.52 | 12.04 |
| | area_MF | 6.98 | 1.13 | 1.40 | 2.53 | 1.48 | 1.65 | 7.48 | 2.94 | 1.09 |
| noise 10% | max_MF | 234.55 | 216.68 | 186.62 | 53.07 | 109.70 | 144.26 | 215.74 | 106.83 | 73.00 |
| | diff_MF | 58.24 | 65.68 | 65.47 | 69.39 | 66.58 | 64.92 | 85.99 | 78.03 | 63.03 |
| | area_MF | 7.21 | 4.39 | 5.93 | 5.84 | 4.14 | 6.22 | 7.74 | 8.93 | 6.30 |

For noise level = 0%, the performance is very close for all the measures, with small and significant absolute differences. In the case of error, the error of the T2-LFPC is 1.9% against 1.5% of the best of the test method (KNN). For noise level = 10%, the performance differences are small and not statistically significant. The method T1-LFPC (the best the test method for 10% noise) obtained $0.017 \pm 0.004$ of error for 0% noise. Comparing performances for 0% and 10% noise, the method T2-LFPC degraded its performance by 352,63% and T1-LFPC by 400%. The other methods got worse performance. These results indicate that the method T2-LFPC is less sensitive to noise than the others.

After visually inspecting the membership functions of Fig. 1 and analyzing the Table 2, it is possible to conclude the next interpretation for noise level = 0%:

**Fig. 1.** Interval-valued membership functions generated from simulated brain MRI (dataset #1) with the method T2-LFPC for noise levels of 0%, and 10%.

- *Cerebrospinal fluid* (blue in Fig. 1): For "PD Intensity", the membership function is centered in very high values around 235, *diff_MF* is high (higher than 40) meaning high imprecision about the attribute and *area_MF* = 6.98, which is high, meaning high vagueness (it is not possible to be precise when explaining the grade in which the attribute is met by the feature). For "T1 Intensity" values are centered around low values (*max_MF* = 53.74), *diff_MF* = 29.90 meaning medium imprecision about the attribute and *area_MF* = 2.53 meaning low vagueness about the grade in which the attribute is met. Finally, for "T2 Intensity" values are centered around 215 (high values), *diff_MF* = 70.16 (very high imprecision about the attribute) and *area_MF* = 7.48 (high vagueness about the degrees of truth).
- *Gray matter* (red in Fig. 1): "PD Intensity" centered around high values, low imprecision about the attribute (*diff_MF* = 15.19), and very low vagueness about the degrees of truth (*area_MF* = 1.13). "T1 Intensity" centered around medium values, medium imprecision about the attribute (*diff_MF* = 24.07), and very low vagueness about the degrees of truth (*area_MF* = 1.48). "T2 Intensity" centered around medium values, high imprecision about the attribute (*diff_MF* = 43.52), and low vagueness about the degrees of truth (*area_MF* = 2.94).
- *White matter* (yellow in Fig. 1): "PD Intensity" centered around slightly high values, very low imprecision about the attribute (*diff_MF* = 13.27), and very low vagueness about the degrees of truth (*area_MF* = 1.40). "T1 Intensity" centered around medium values, low imprecision about the attribute (*diff_MF* = 15.53), and very low vagueness about the degrees of truth (*area_MF* = 1.65). "T2 Intensity" centered around low values, low imprecision about the attribute (*diff_MF* = 12.04), and very low vagueness about the degrees of truth (*area_MF* = 1.09).

Considering the previous descriptions about the relative positions of the membership functions and the mapping of generic to experts' domain attributes, the next predicates explaining the tissues are defined for noise level $= 0\%$:

- "The pixel belongs to cerebrospinal fluid" is equivalent to: "the pixel is very-bright in PD intensity **and** is dark in T1 intensity **and** is bright in T2 intensity".
- "The pixel belongs to gray matter" is equivalent to: "the pixel is bright in PD intensity **and** is gray in T1 intensity **and** is gray in T2 intensity".
- "The pixel belongs to white matter" is equivalent to: "the pixel is light-gray in PD intensity **and** is gray in T1 intensity **and** is black in T2 intensity".

For dataset #2 (real brain MRI), Table 3 summarizes the results for noise level of 0% and of 10%. In Fig. 2, the membership functions generated are shown. In Fig. 3, input images, Gold-Standard, and results obtained for one of the MRI are shown. Table 4 summarizes results of the measures proposed. For noise level $= 0\%$, the T2-LFPC was worse than the best of the test methods (MLP). The same occurred for noise level $= 10\%$. However, comparing the degradation of the performance between noise $= 0\%$ and noise $= 10\%$, MLP degraded its performance by 72.22% and T2-LFPC degraded by 51.85%. Therefore, as reported before for dataset #1, the method T2-LFPC is less sensitive to noise than the others.

**Table 3.** Results for the segmentation of dataset #2 (real brain MRI).

| Noise | Measure | T2-LFPC | Best of the test methods | Difference |
|-------|---------|---------|--------------------------|------------|
| 0% | Error | $0.189 \pm 0.013$ | $0.162 \pm 0.006$ (MLP) | $16.232\% P < 0.001$ |
| | Tanimoto | $0.690 \pm 0.017$ | $0.725 \pm 0.010$ (MLP) | $-4.862\%$ $P < 0.001$ |
| | Dice | $0.813 \pm 0.012$ | $0.838 \pm 0.007$ (MLP) | $-2.985\%$ $P < 0.001$ |
| 10% | Error | $0.287 \pm 0.007$ | $0.279 \pm 0.006$ (MLP) | $2.844\%$ $P = 0.026$ |
| | Tanimoto | $0.565 \pm 0.008$ | $0.575 \pm 0.007$ (MLP) | $-1.846\%$ $P = 0.008$ |
| | Dice | $0.713 \pm 0.008$ | $0.721 \pm 0.007$ (MLP) | $-1.194\%$ $P = 0.015$ |

The methodology proposed for understanding brain MRI applied in detail on the dataset #1 can be also applied here, obtaining compound fuzzy predicates, and descriptions of imprecision about the attributes and vagueness about the degrees of truth. This is omitted here to avoid overextending the manuscript. However, it is possible to do the next analysis:

- There are some changes in the positions of the membership functions compared against dataset #1. This is because simulated MRI are brighter than real ones.

- The order of the membership functions for each of the tissues and for each feature is the same than for dataset #1.

**Table 4.** Results of the three measures on interval-membership functions proposed considering segmentation of dataset #2 (real brain MRI). (CSF = Cerebrospinal fluid, WM = White matter, GM = Gray matter)

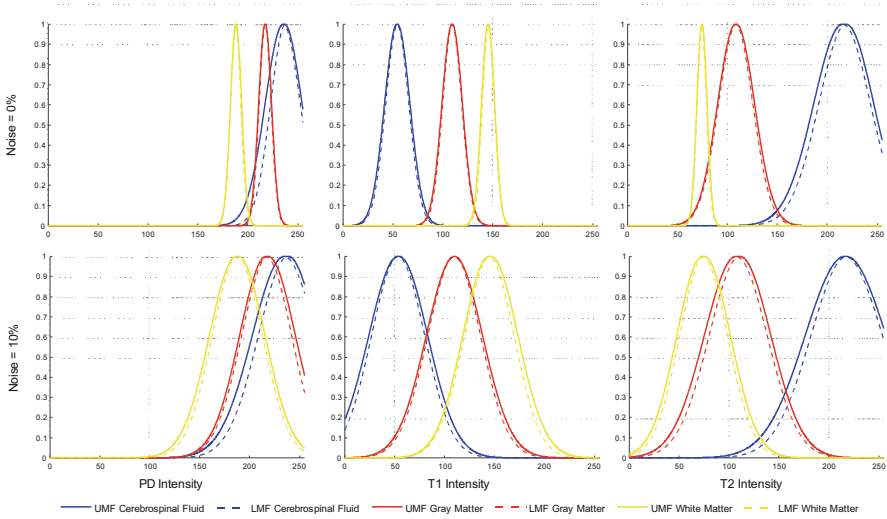| | Measure | PD Intensity | | | T1 Intensity | | | T2 Intensity | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CSF | GM | WM | CSF | GM | WM | CSF | GM | WM |
| noise 0% | max_MF | 187.45 | 161.47 | 148.96 | 38.99 | 69.74 | 89.17 | 174.34 | 108.04 | 87.53 |
| | diff_MF | 46.41 | 45.16 | 30.53 | 36.00 | 39.39 | 25.24 | 87.92 | 45.49 | 24.65 |
| | area_MF | 5.68 | 9.89 | 4.96 | 4.96 | 6.20 | 2.98 | 9.76 | 4.20 | 3.37 |
| noise 10% | max_MF | 187.32 | 160.69 | 147.52 | 39.49 | 70.54 | 88.79 | 175.35 | 107.69 | 86.37 |
| | diff_MF | 76.73 | 74.76 | 69.58 | 71.48 | 73.60 | 70.61 | 110.99 | 79.69 | 65.59 |
| | area_MF | 3.92 | 5.94 | 7.22 | 4.57 | 7.27 | 8.44 | 11.42 | 10.88 | 6.87 |
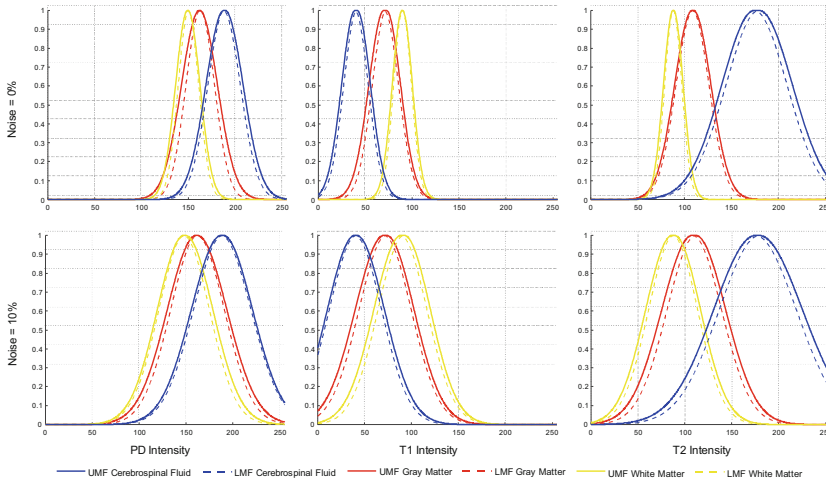


**Fig. 2.** Interval-valued membership functions generated from real brain MRI (dataset #2) with the method T2-LFPC for noise levels of 0%, and 10%.
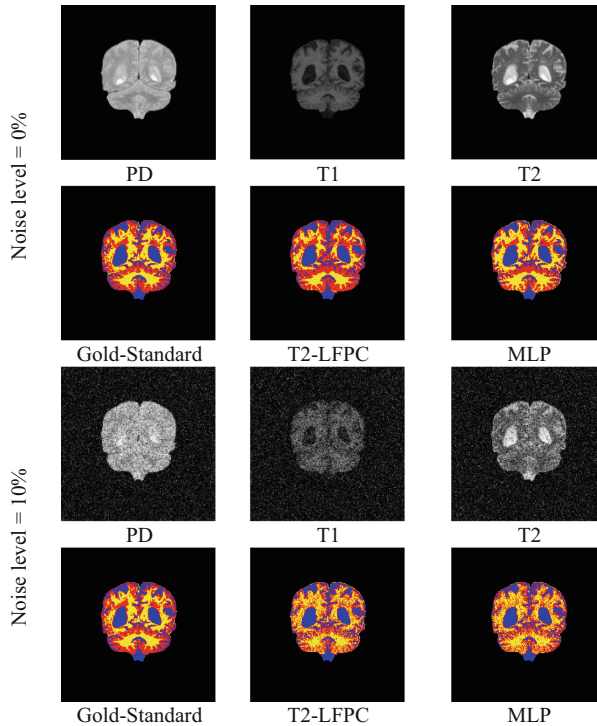
**Fig. 3.** Input images, Gold-Standard, and results obtained for one of the images of the dataset #2 (real MRI) for noise levels of 0%, and 10%.

## 4   Discussion

The major contribution of the methodology proposed is to enable understanding brain MRI, allowing: the definition of attributes on features and their association to each tissue to provide linguistic descriptions, and the identification, quantification and description of both vagueness associated with the attributes and spread of intensities of pixels in each tissue. Considering the results, the next analysis can be done:

- Regarding the positions of the membership functions, for all features and classes, there are not significant changes, as expected as noise added to the images has *mean-value* = 0. The order is the same, i.e., for "PD Intensity" is White matter-Gray matter-Cerebrospinal fluid, for "T1 Intensity" is Cerebrospinal fluid-Gray matter-White matter, for "T2 Intensity" is White matter-Gray matter-Cerebrospinal fluid. Following the previous reasoning for assigning meaning to the attributes, it is not required to change them, meaning the attributes are the same.
- Major changes are related to the imprecision and vagueness represented by the membership functions. While the attributes are the same because the positions did not change, their descriptions in terms of the membership functions (their meanings), changed as both spread of the values associated to each tissue and impression increases.

- Considering imprecision about the attribute (described by *diff_MF*), it was increased in all the cases compared to the case without noise, which is consistent with the increment of the spread of the data due to the addition of noise.
- Regarding the vagueness about the degrees of truth (described by *area_MF*), there are significant increases in most of the cases. It is important to note that considering how the method T2-LFPC works, the *area_MF*, should increase because of a significant difference observed between the class prototypes extracted from the subsets generated in the random partition (Stage A).

## 5   Conclusions

In this work, it was proposed a methodology for understanding brain MRI from automatically generated interval-valued membership functions from the outcome of our previously proposed method called T2-LFPC.

The methodology was applied both on simulated and real brain MRI considering additive noise, showing good segmentation performance compared against test methods including T1-LFPC (a variant of the T2-LFPC using type-1 membership functions), Probabilistic Neural Networks, Multi-Layer Perceptrons, and k-Nearest Neighbors. Moreover, compared against the best test method, the T2-LFPC method showed to be less sensitive to the noise.

The measures on the membership-functions proposed allow to quantify and describe levels of imprecision and vague-ness regarding the attributes and the degrees of truth of the functions generated through the T2-LFPC. The methodology proposed allows to define attributes on the features and to associate them to each tissue. It also allows to describe relationships between attributes and tissues providing linguistic descriptions to identify, quantify, and describe both vagueness associated with the attributes and spread of intensities of pixels belonging to each tissue. The knowledge achieved is consistent with what is known in the field of brain MRI segmentation.

For all said, the methodology constitutes a sound approach for knowledge discovery which could be extended to other medical imaging domains, making it a general approach for interpreting medical images. Several new tests will be run as immediate future work, considering more datasets and tests to improve the understanding the of membership functions and, therefore, of medical images.

## References

1. Wu, G., Shen, D., Sabuncu, M.R.: Machine Learning and Medical Imaging. Academic Press is an imprint of Elsevier (2016)
2. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017). https://doi.org/10.1016/J.MEDIA.2017.07.005
3. de Bruijne, M.: Machine learning approaches in medical image analysis: from detection to diagnosis. Med. Image Anal. **33**, 94–97 (2016). https://doi.org/10.1016/j.media.2016.06.032

4. Comas, D.S., Meschino, G.J., Brun, M., Ballarin, V.L.: Label-based Type-2 Fuzzy Predicate Classification applied to the design of morphological W-operators for image processing. In: First Latin American Congress on Computational Intelligence. pp. 55–60. San Carlos de Bariloche, Argentina (2014)

5. Comas, D.S., Meschino, G.J., Costantino, S., Capiel, C., Ballarin, V.L.: Interval type-2 fuzzy predicates for brain magnetic resonance image segmentation. Rev. Argentina Bioingeniería. **21**, 11–19 (2017)

6. Comas, D.S., Meschino, G.J., Nowé, A., Ballarin, V.L.: Discovering knowledge from data clustering using automatically-defined interval type-2 fuzzy predicates. Expert Syst. Appl. **68**, 136–150. springer (2017). https://doi.org/10.1016/j.eswa.2016.10.018

7. Mendel, J.M.: Type-2 fuzzy sets and systems: an overview. IEEE Comput. Intell. Mag. **2**, 20–29 (2007)

8. Comas, D.S., Pastore, J.I., Bouchet, A., Ballarin, V.L., Meschino, G.J.: Interpretable interval type-2 fuzzy predicates for data clustering: A new automatic generation method based on self-organizing maps. Knowl. Based Syst. **133**, 234–254 (2017). https://doi.org/10.1016/j.knosys.2017.07.012

9. Wu, D., Mendel, J.M.: Uncertainty measures for interval type-2 fuzzy sets. Inf. Sci. (Ny) **177**, 5378–5393 (2007)

10. Kwan, R.K.-S., Evans, A.C., Pike, G.B.: MRI simulation-based evaluation of image-processing and classification methods. IEEE Trans. Med. Imaging **18**, 1085–1097 (1999)

11. Center, I.M.H.C.R.: BRAINS Software Package. http://www.psychiatry.uiowa.edu/mhcrc/IPLpages/BRAINS.htm

12. Wasserman, P.D.: Advanced Methods in Neural Computing. Van Nostrand Reinhold, New York (1993)

13. Bishop, C.: Neural Networks for Pattern Recognition. Oxford Press, Oxford (2005)

14. Rajini, N.H., Bhavani, R.: Classification of MRI brain images using k-nearest neighbor and artificial neural network. In: International Conference on Recent Trends in Information Technology, ICRTIT 2011, pp. 563–568. IEEE Computer Society (2011). https://doi.org/10.1109/ICRTIT.2011.5972341

15. Gibbons, J.D., Chakraborti, S.: Nonparametric Statistical Inference. CRC Press, New York (2003)

# 2D Electrical Impedance Tomography Brain Image Reconstruction Using Deep Image Prior

Leonardo A. Ferreira[1(✉)] , Roberto G. Beraldo[1] , Ricardo Suyama[1] ,
Fernando S. Moura[2] , and André K. Takahata[1]

[1] Federal University of ABC, Avenida dos Estados, 5001, Santo André, SP
09210-580, Brazil
leonardo.alves@ufabc.edu.br
[2] Federal University of ABC, Alameda da Universidade, s/n, São Bernardo do
Campo, SP 09606-045, Brazil

**Abstract.** Electrical impedance tomography (EIT) is a medical imaging modality that has the potential to benefit diagnosing, monitoring, and understanding several pathological conditions. However, some regions of the body, such as the brain, are more challenging to reconstruct, demanding improvements before the technique can be used in clinical practice. In this study, we implemented and evaluated an algorithm for 2D static EIT image reconstruction based on the Deep Image Prior (DIP) method. The method was tested in measurements calculated from a computational human head model, where we included a region representing the occurrence of a stroke. The results showed that the DIP-based algorithm had some advantages compared to a more classical method, such as robustness to noise and independence of an initial solution. Therefore, this method could be better suited for real-life EIT image reconstructions.

**Keywords:** electrical impedance tomography · deep image prior · ill-posed problems · discrete inverse problems · medical image processing

## 1 Introduction

Electrical Impedance Tomography (EIT) is a technique that allows the acquisition of medical images displaying the resistivities (or impeditivities) of the interior of a body region [1]. It works by surrounding the region of interest with an electrodes array that is then used to apply electrical currents and measure the corresponding generated voltages. Repeating this procedure multiple times, each using different pairs of electrodes to apply the current, results in a set of measurements from which the resistivities image can be reconstructed. The forward problem associated with this technique consists in calculating the voltages ($\mathbf{V}_e$) under the electrodes given the resistivities ($\boldsymbol{\rho}$) of the domain and the applied currents ($\mathbf{C}$):

$$\mathbf{V}_e = f(\boldsymbol{\rho}, \mathbf{C}), \tag{1}$$

where $f(\cdot)$ is a nonlinear function. In contrast, reconstructing EIT images (i.e., solving the inverse problem) consists in obtaining the resistivities from the voltages and currents values. Determining the absolute values of the resistivities is known as absolute or static EIT, which is an ill-posed problem in the sense of Hadamard. Consequently, discrepancies, such as errors in the estimation of the forward model and noise in the electrodes measurements, lead to poor results when using a naive algorithm to search for the solution.

A strategy to mitigate these discrepancies is to reconstruct the difference between the resistivities of different instants of time (known as dynamic or time difference EIT). However, this method prevents the use of the technique for cases where we wish to image steady conditions without having measurements from before its occurrence. Other approaches to improve EIT image reconstruction algorithms include the incorporation of a priori information about the problem, which can be done by using Tikhonov regularizers [2], statistical models about the region [3], or deep learning methods [4], for example.

Compared to other medical imaging modalities, such as computed tomography (CT) or magnetic resonance imaging (MRI), EIT has the advantages of being cheaper, portable, and not presenting any known side effects [1]. These characteristics could benefit diagnosing, monitoring, and understanding several pathological conditions. For example, the portability could allow an early diagnosis of strokes, thus increasing the chances of a better prognosis for the patient [5]. However, the brain is a particularly difficult region to image with this method due to the low conductivity of the skull, which hinders the current flow to the brain.

Although many advances have been achieved to the goal of using EIT for brain imaging, there still is a need for improvements before it can be used in clinical practice. We refer to [6] for a complete review on this matter. Meanwhile, deep learning methods have been yielding state-of-the-art results in many different areas, including medical image processing [7]. Nonetheless, it can be challenging to validate this type of approach for EIT due to the lack of large real datasets and the difficulty in establishing ground truth images.

Recently, a new way of applying the deep learning tools without the need for training datasets has been proposed, named Deep Image Prior (DIP) [8]. The idea behind it is to take advantage of the regularizing properties of the convolutional neural network (CNN) structure to generate the desired solution. It works by using a random noise tensor as input to the CNN and then applying the forward model of the problem of interest to the corresponding output. The result is compared to the acquired measurements through a loss function, which is minimized by the update of the parameters of the CNN through several iterations. The hypothesis is that while the loss function is minimized, the value of the CNN output gets closer to the desired solution.

This method was shown to be effective in the solution of several low-level computer vision tasks, as well as in the image reconstruction of CT [9] and positron emission tomography (PET) [10]. Despite that, we have not found any studies considering this approach to solve the EIT inverse problem. Therefore,

our work aimed to implement and evaluate a DIP-based EIT 2D image reconstruction algorithm. Specifically, we considered the case of static EIT for brain imaging.

## 2  Materials and Methods

We implemented the presented algorithms using Python 3.7. In this study, we only evaluated the proposed method through computational simulations. Both the reconstruction algorithms and the simulated electrode measurements were calculated in a discretized domain accordingly to the Finite Element Method (FEM) [11] and considering the Complete Electrode Model (CEM) [12] to determine the voltages.

### 2.1  Mesh Generation

We generated a 2D mesh based on a young adult head model [13], from which we extracted the contours of the scalp, skull, and brain. The chosen region of the head corresponded to a transversal slice localized around the middle of the model forehead. The software GMSH[1] was used to create a mesh composed of triangular elements and containing 32 electrodes surrounding the region. Two meshes were generated: a more refined version (approximately 11000 elements and 5500 nodes) to be used in the simulation of the measurements and a coarse version (approximately 2100 elements and 1100 nodes) to be used in the image reconstruction. This difference was introduced to avoid the inverse crime through the representation of the discretization error. The position of the electrodes was the same in both meshes (i.e., we assumed the exact position is known and used by the algorithm).

### 2.2  Simulation of the Measurements

We used the resistivities of $2.50\,\Omega m$, $47.94\,\Omega m$, and $9.12\,\Omega m$ for the regions of the scalp, skull, and brain, respectively. Additionally, a circular area with a radius of 1.5 cm and resistivity of $1.42\,\Omega m$ was inserted into the brain region to represent the occurrence of a hemorrhagic stroke. Three different positions were chosen for the stroke, following what is presented in Fig. 1. The thickness of both the domain and the electrodes was considered to be 3 cm, and the electrode-skin contact impedance was set to $0.02\,\Omega m$.

The simulations considered a current of 1mA injected according to the skip-16 pattern, i.e., the electrode injection pairs had 16 other electrodes separating them. The voltages were calculated considering single-ended measurements relative to the electrical ground, which was chosen to be the center of the mesh. For every current injected, the voltages of all the electrodes (including the injection pair) were determined. Furthermore, we simulated five different additive Gaussian noise realizations for each stroke position. The distribution that originated

---
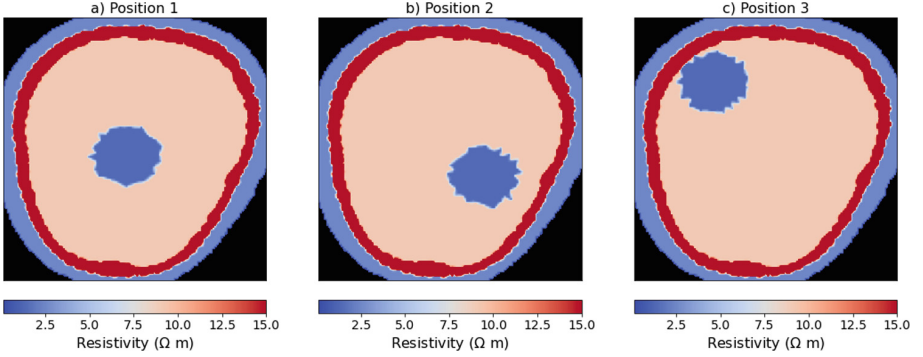
**Fig. 1.** Three cases of stroke used to generate the electrode measurements. For a better visualization of the brain, all values equal or above 15 $\Omega m$ were represented in the same color.

the noise had a mean of 0V and a standard deviation of 0.003 V, which resulted in a mean signal-to-noise ratio (SNR) of 31.71dB across the three positions. We also considered the case without noise for a base of comparison. Therefore, combining the distinct stroke positions, the different noise realizations, and the noiseless case, we obtained a total of 18 different sets of simulated voltage measurements.

## 2.3  DIP-Based EIT

The DIP implementation of our method was based on the default algorithm presented in [8]. The structure of the CNN consisted of an encoder-decoder architecture with skip connections between the encoder and decoder parts. Specific details about the architecture are presented in Table 1. The CNN input was a uniform noise with intensity values ranging from 0 to $\frac{1}{10}$. Each iteration of the algorithm used a small variation of this input, determined by adding to it a Gaussian noise with a mean equal to zero and a standard deviation equal to 0.008. Additionally, we limited the value of the CNN output to 80 $\Omega m$ to avoid the convergence of the result to high unrealistic values. Note that the ReLU activation function in the layers of the CNN also restricts the output to be greater than or equal to zero. Therefore, given $\mathbf{N}$, the noise tensor that is entered into the CNN, we have that the resistivities output image is

$$\mathbf{R_\theta} = \phi_\theta(\mathbf{N}), \tag{2}$$

where $\phi_\theta(\cdot)$ is a function that represents the CNN and $\boldsymbol{\theta}$ represents the CNN parameters. To integrate this output, which is provided in pixels, into the forward problem, which is calculated using the FEM, the first step was to determine the coordinates of each pixel. For each dimension, the coordinate of a pixel with index $i$ was calculated as

$$c_i = \frac{m_{max} - m_{min}}{l_p} i + m_{min}, \tag{3}$$

where $m_{max}$ and $m_{min}$ are the highest and lowest centroid coordinates of the mesh elements, respectively, and $l_p$ is the height or width (in pixels) of the CNN output. Secondly, the resistivity of the mesh elements was determined by

$$\boldsymbol{\rho}_t = \mathbf{T}\boldsymbol{\rho_\theta},  \tag{4}$$

where $\mathbf{T}$ is an interpolation matrix and $\boldsymbol{\rho_\theta}$ is $\mathbf{R_\theta}$ reshaped into a column vector by concatenating its lines:

$$\boldsymbol{\rho_\theta} = \mathrm{vec}(\mathbf{R_\theta}).  \tag{5}$$

We used an interpolation matrix defined by a low-pass Gaussian filter, where each element is determined by

$$[\mathbf{T}]_{j,k} = \frac{\frac{1}{2\pi\sigma}exp\left(-\frac{d(j,k)^2}{2\sigma^2}\right)}{\sum_{j,k}\frac{1}{2\pi\sigma}exp\left(-\frac{d(j,k)^2}{2\sigma^2}\right)},  \tag{6}$$

where $d(j,k)$ is the euclidean distance between the centroid of the mesh element with index $j$ and the coordinate of the pixel with index $k$, and $\sigma$ is the standard deviation of the filter. In all our simulations, $\sigma$ was set to $5 \times 10^{-7}$.

**Table 1.** Convolutional neural network architecture used for the DIP.

| Parameter | DIP |
|---|---|
| Input size | $256 \times 256 \times 32$ |
| Output size | $256 \times 256 \times 1$ |
| No. of encoder layers | 6 |
| No. of decoder layers | 6 |
| No. of filters in each encoder layer | 128 (all) |
| No. of filters in each decoder layer | 128 (all) |
| Size of filters in each encoder/decoder layer | $3 \times 3$ (all) |
| No. of skip layers | 6 |
| No. of filters in each skip layer | 32 |
| Size of filters in each skip layer | $1 \times 1$ |
| Activation function | ReLU |

Once the resistivities of the mesh elements were determined, the EIT forward model was applied to calculate the corresponding voltages $\mathbf{V_\theta}$. Therefore, we have that

$$\mathbf{V_\theta} = f(\boldsymbol{\rho}_t) = f(\mathbf{T}\,\mathrm{vec}(\phi_\theta(\mathbf{N}))),  \tag{7}$$

where $f(\cdot)$ is the forward model. Then, these voltages were compared, using the mean squared error (MSE), to the simulated electrode measurements. We also

included a total variation (TV) regularization term [14,15] in the loss function, thus resulting in

$$L(\boldsymbol{\theta}) = MSE(\mathbf{V_\theta}, \mathbf{V}_m) + w_{tv}TV(\boldsymbol{\theta}), \tag{8}$$

where $\mathbf{V}_m$ is the simulated measurements, $w_{tv}$ the TV weight, and $TV(\cdot)$ the TV regularization term. We chose $w_{tv} = 5 \times 10^{-8}$ for all of the simulated cases. The loss function was minimized by the Adam optimizer through the update of $\boldsymbol{\theta}$ with a learning rate of $5 \times 10^{-4}$. A total of 8000 iterations were executed to reconstruct each image, with the final result being the last CNN output generated. The same parameters were used for both the cases with and without noise.

## 2.4    Method for Comparison

We compared the results of the proposed method to a more classical approach to reconstruct EIT images. This approach was based on the solution of the inverse problem using the Gauss-Newton method [2] along with a TV regularization term. The equation to be optimized was

$$\boldsymbol{\rho}_{GN}^* = \arg\min_{\boldsymbol{\rho}_{GN}} \left\{ ||\mathbf{V}_m - f(\boldsymbol{\rho}_{GN})||_2^2 + \lambda_{tv}TV(\boldsymbol{\rho}_{GN}) \right\}, \tag{9}$$

where $\lambda_{tv}$ is the weight of the TV regularization. We chose $\lambda_{tv} = 1 \times 10^{-5}$ for all cases simulated with noise and $\lambda_{tv} = 1 \times 10^{-7}$ for the cases without noise. The $\ell_1$-norm was approximated through the iteratively reweighted least square (IRLS) method [16, p. 197].

To initialize the optimization we used the resistivities of $3.0\,\Omega m$, $45.0\,\Omega m$, and $9.5\,\Omega m$ for the regions of the scalp, skull, and brain, respectively. Therefore, for this method, we considered that the position of the tissues inside the head is known and that we have a good estimate of their resistivities. The optimization was executed for 40 iterations with a step size of 0.13 for the cases with noise, and for 30 iterations with a step size of 0.07 for the cases without noise.

## 2.5    Quantitative Evaluation

Three metrics were used to evaluate the results. The first one was the intersection over union (IoU) (also known as Jaccard Index), which is a common metric used in the evaluation of the similarity between two sets. It is defined as

$$IoU(\mathbf{S}_m, \mathbf{S}_g) = \frac{\mathbf{S}_m \cap \mathbf{S}_g}{\mathbf{S}_m \cup \mathbf{S}_g}, \tag{10}$$

where $\mathbf{S}_m$ and $\mathbf{S}_g$ are the binary masks of the stroke regions in the output of the method and the ground truth image, respectively. To determine the stroke region, the skull was segmented by selecting the pixels with a resistivity value greater than $14\,\Omega m$. Then, the brain mask was determined by taking all the pixels inside the interior border of the skull. Finally, the minimum resistivity of the obtained brain region was determined, and the binary mask of the stroke

consisted of the pixels in this region with resistivities lower than 1.25 times this value.

The second metric, resistivity difference (RD), had the goal of evaluating the contrast between the stroke resistivity and the rest of the brain. To achieve that, we calculated the difference between the mean resistivities of the stroke region and the remaining area of the brain. These regions were determined with the same masks defined for the IoU, with the remaining area of the brain being obtained by the subtraction between the brain and stroke masks.

The last metric we calculated had the goal of detecting the proportion of high resistivity artifacts (HRA) that could confound the image analysis. First, we determined the region of the interior of the brain by applying 10 iterations of the morphological erosion operation to the brain region mask. This step was included to disregard high values of resistivity that may appear at the border with the skull. Then, the value of the metric was the percentage of pixels in the brain interior that were above 1.25 times the mean resistivity of the region (excluding the stroke area).

## 3   Results and Discussion

Figure 2 shows the images reconstructed for the noiseless case. As it can be seen, with the DIP-based reconstruction it was possible to distinguish the region of the stroke in all of the three simulated positions. The case where the stroke was near the middle of the brain was the one where the borders of the region were less defined. This result was expected since the greater distance to the electrodes makes the measurements less sensible to the values of the center of the brain.

In this noiseless condition, the results of the Gauss-Newton reconstruction were satisfactory. Apart from a few artifacts in the scalp region, the stroke was reconstructed with a similar shape and size to what was expected, and the remaining resistivity of the brain was near to uniform. However, as can be seen in Fig. 3, the results of the Gauss-Newton method were visibly worse when additive noise was present in the measurements. The images displayed regions of high resistivity in the brain with a shape similar to the reconstructed stroke, which would be confusing for a diagnostic analysis because of the difficulty to determine if those regions are artifacts or the result of a real condition that affected the patient.

Meanwhile, the images of the DIP reconstruction in the noisy cases were similar to the ones obtained in the noiseless ones, indicating the greater robustness of this option. Therefore, this method may be better suited for real-life applications, where the presence of noise is inevitable. Another advantage observed was that we obtained all the DIP results without changing the hyperparameters, while for the Gauss-Newton algorithm we needed to change the TV weight, step size, and number of iterations for the noisy cases to obtain better results (as described in Subsection II-D). Therefore, the DIP-based reconstruction may be less sensitive to variations in the characteristics of the noise, thus not requiring different hyperparameters for each situation.
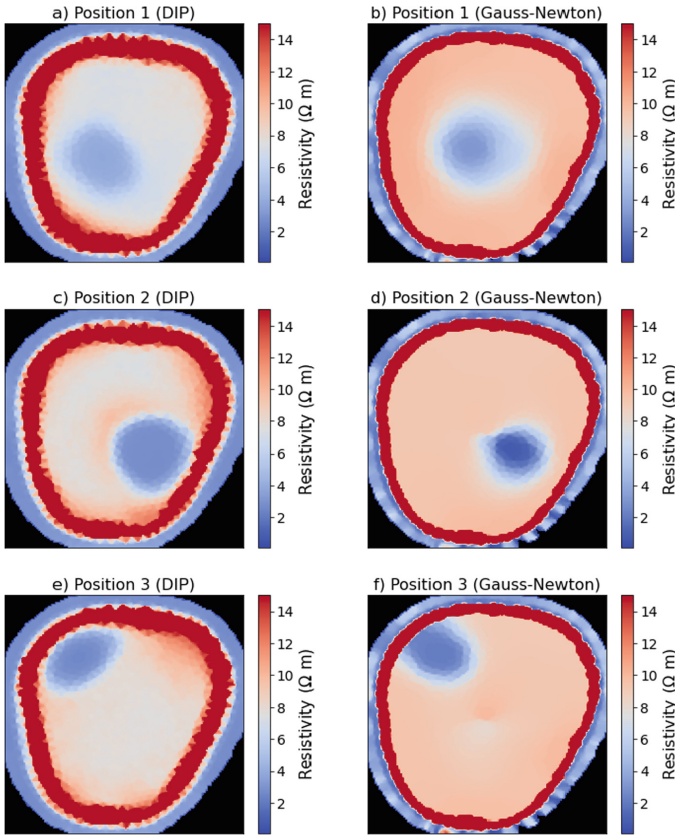
**Fig. 2.** Example of images reconstructed from the noiseless case. For a better visualization of the brain, all values equal or above 15 $\Omega m$ were represented in the same color.

A downside of the DIP-based method was the skull reconstruction, which was thicker than what was expected. Additionally, it is possible to see a region of intermediary values between the border of this region and the other tissues, which indicates the lack of sharpness of the reconstruction. However, EIT images are already known to have a poor spatial resolution, being more important the stability and general aspect of the result. For example, the thickness of the skull was similar in all images, and the strokes were visible even when they were placed close to the skull border. Therefore, this artifact would not prevent the images from being used for the application of stroke detection.

It is also worth noting that the DIP-based algorithm was initialized with a random resistivity, while the Gauss-Newton one received an initialization close to the desired solution, including the correct localization of the borders of the tissues. In our experiments, the results of the Gauss-Newton with a random initialization were far from the ground truth. Therefore, the DIP reconstruction
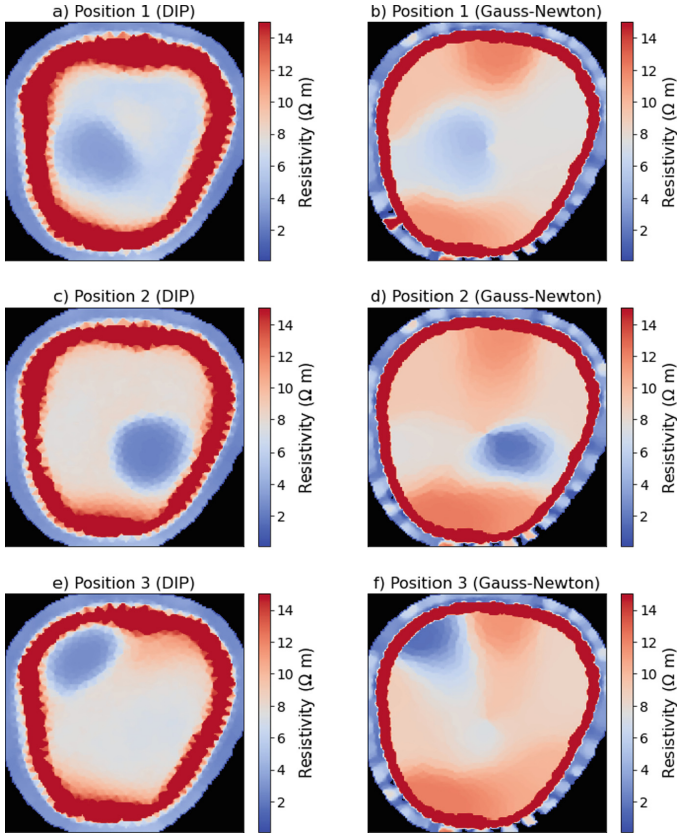
**Fig. 3.** Images reconstructed from data with additive noise. For a better visualization of the brain, all values equal or above 15 $\Omega m$ were represented in the same color.

may be advantageous in real life, where we usually have no information about the exact anatomy of the patient.

The mean and standard deviation of the metrics, calculated considering five different additive Gaussian noise realizations for each stroke position, are presented in Table 2. The metrics corroborate what was observed in the qualitative analysis, with the DIP-based reconstructions having a higher IoU and less high resistivity artifacts than the Gauss-Newton ones. The resistivity difference was higher in the Gauss-Newton reconstructions in two of the stroke positions, however, this metric was possibly influenced by the high resistivity artifacts present in the results of this method, which may increase the mean value of the region outside the stroke. Moreover, the resistivity difference obtained in the DIP reconstructions may already be sufficient for diagnosing the stroke since it is possible to identify the abnormal regions in the presented images. By analyzing the standard deviation of the metrics, it is also possible to note that the DIP results are

more consistent as they present smaller values when considering different noise realizations than the results obtained with Gauss-Newton.

**Table 2.** Intersection over union (IoU), resistivity difference (RD), and high resistivity artifacts (HRA) calculated for the reconstructions using the DIP and Gauss-Newton (GN) methods in the three stroke positions using the noisy measurements.

| Method | Pos. | IoU | RD ($\Omega m$) | HRA (%) |
|---|---|---|---|---|
| DIP | 1 | $0.46 \pm 0.05$ | $2.98 \pm 0.14$ | $1.99 \pm 1.88$ |
| | 2 | $0.77 \pm 0.03$ | $4.98 \pm 0.17$ | $0.84 \pm 0.68$ |
| | 3 | $0.54 \pm 0.06$ | $4.51 \pm 0.27$ | $2.07 \pm 1.84$ |
| GN | 1 | $0.36 \pm 0.10$ | $2.94 \pm 0.40$ | $4.20 \pm 3.06$ |
| | 2 | $0.29 \pm 0.06$ | $6.66 \pm 0.66$ | $3.78 \pm 4.47$ |
| | 3 | $0.31 \pm 0.17$ | $5.82 \pm 1.13$ | $2.65 \pm 3.10$ |

## 4   Conclusion

In this study, we implemented and evaluated a DIP-based algorithm for static EIT image reconstruction of the brain, specifically aiming at the detection of strokes. The results showed that this method presents improvements when compared to a more classical approach of inversion, presenting characteristics that are important for a real-life application of the technique, such as robustness to noise and independence of the initial solution. Future works may include a more in-depth validation of the method, including, for example, the evaluation of different stroke sizes and head shapes, more detailed anatomic models, and high resistivity (ischemic) strokes. Furthermore, the method has several points that could be further explored to improve the results, for instance, different CNN architectures, alternative regularization terms, other strategies to integrate the CNN output and the forward model, and the adaptation of the algorithm for 3D images. Therefore, the presented technique has potential to lead to new directions that could be explored in the search for better EIT reconstruction techniques.

**Conflicts of Interest.** The authors declare that they have no conflict of interest.

# References

1. Holder, D.: Electrical Impedance Tomography: Methods, History and Applications. CRC Press, 1 edition, (2004). ISBN 978-0750309523
2. Vauhkonen, M., Vadász, D., Karjalainen, P.A., Somersalo, E., Kaipio, J.P.: Tikhonov regularization and prior information in electrical impedance tomography. IEEE Trans. Med. Imaging **17**(2), 285–293 (1998). https://doi.org/10.1109/42.700740
3. Ferreira, L.A., Beraldo, R.G., Camargo, E.D.L.B., Moura, F.S.: Anatomical atlas of the human head for electrical impedance tomography. In: Bastos-Filho, T.F., de Oliveira Caldeira, E.M., Frizera-Neto, A. (eds.) CBEB 2020. IP, vol. 83, pp. 1693–1699. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-70601-2_247
4. Ren, S., Sun, K., Tan, C., Dong, F.: A two-stage deep learning method for robust shape reconstruction with electrical impedance tomography. IEEE Trans. Instrum. Meas. **69**(7), 4887–4897 (2019). https://doi.org/10.1109/TIM.2019.2954722
5. Hacke, W., et al.: Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. The Lancet **363**(9411), 768–774 (2004). https://doi.org/10.1016/S0140-6736(04)15692-4
6. Ke, X.-Y., et al.: Advances in electrical impedance tomography-based brain imaging. Mil. Med. Res. **9**(1), 1–22 (2022). https://doi.org/10.1186/s40779-022-00370-7
7. Razzak, M.I., Naz, S., Zaib, A.: Deep learning for medical image processing: overview, challenges and the future. In: Dey, N., Ashour, A.S., Borra, S. (eds.) Classification in BioApps. LNCVB, vol. 26, pp. 323–350. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-65981-7_12
8. Lempitsky, V., Vedaldi, A., Ulyanov, D.: Deep image prior. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9446–9454. IEEE (2018). https://doi.org/10.1109/CVPR.2018.00984
9. Baguer, D.O., Leuschner, J., Schmidt, M.: Computed tomography reconstruction using deep image prior and learned reconstruction methods. Inverse Prob. **36**(9), 094004 (2020). https://doi.org/10.1088/1361-6420/aba415
10. Gong, K., Catana, C., Qi, J., Li, Q.: Pet image reconstruction using deep image prior. IEEE Trans. Med. Imaging **38**(7), 1655–1665 (2018). https://doi.org/10.1109/TMI.2018.2888491
11. Woo, E.J., Hua, P., Webster, J.G., Tompkins, W.J.: Finite-element method in electrical impedance tomography. Med. Biol. Eng. Compu. **32**(5), 530–536 (1994). https://doi.org/10.1007/BF02515311
12. Cheng, K.-S., Isaacson, D., Newell, J.C., Gisser, D.G.: Electrode models for electric current computed tomography. IEEE Trans. Biomed. Eng. **36**(9), 918–24 (1989). https://doi.org/10.1109/10.35300
13. Hammond, D., Price, N., Turovets, S.: Construction and segmentation of pediatric head tissue atlases for electrical head modeling. OHBM, Vancouver, Canada (2017)
14. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–68 (1992). https://doi.org/10.1016/0167-2789(92)90242-f
15. Borsic, A., Graham, B.M., Adler, A., Lionheart, W.R.B.: In vivo impedance imaging with total variation regularization. IEEE Trans. Med. Imaging **29**(1), 44–54 (2010). https://doi.org/10.1109/TMI.2009.2022540
16. Aster, R.C., Borchers, B., Thurber, C.H.: Parameter Estimation and Inverse Problems, 3rd edn. Elsevier, Amsterdam, p. 9780128134238 (2019)

# Ultrasound Speckle Filtering Using Deep Learning

Y. Z. O. Gomez[1(✉)] and E. T. Costa[1,2]

[1] Department of Electronics and Biomedical Engineering, School of Electrical and Computer Engineering (DEEB/FEEC), University of Campinas (UNICAMP), Campinas, Brazil
`yamzugomez@yahoo.es`

[2] Center for Biomedical Engineering (CEB), University of Campinas (UNICAMP), Campinas, Brazil

**Abstract.** In this work we present a technique to reduce speckle in ultrasound images that lead to sharper image when compared to other speckle reducing techniques. To achieve this goal, Cycle Generative Adversarial Networks (CycleGAN) are used, which map specific characteristics of a set of input images that will later be used to give origin to transformed images. The discriminator used in this article tries to differentiate between images with and without speckle by the network and the set of high-quality images. The images without speckle used in the network training were obtained from online databases and from a phantom using an Ultrasonix platform and later postprocessed to remove the speckle. The quality of the generated images was evaluated using the metrics Signal to Noise Ratio (SNR) and Peak Signal to Noise Ratio (PSNR) and compared to the Perona-Malik diffusion filter. The proposed technique had an improvement of 4.9 dB in the SNR and 4.7 dB in the PSNR when compared with Perona-Malik filtering.

**Keywords:** Ultrasound · Speckle · Filtering · Deep Learning · CycleGAN

## 1 Introduction

In modern medicine, methods for visualizing the internal structures of an organism through images are widely used. Technologies for viewing images have evolved considerably and among the most prominent are magnetic resonance imaging, computed tomography, and ultrasound.

Ultrasonography uses the echo pulse generated by high frequency ultrasonic waves to visualize the internal structures of the body and has advantages over other methods such as low cost, fast execution, real-time visualization of the structures and is free from ionizing radiation. However, the images can be of low quality and with the presence of artifacts.

One of the difficulties encountered during ultrasound image analysis is the presence of speckle, whose effect is the granular aspect of the image. Speckle is a characteristic phenomenon of ultrasound images, caused by interference between coherent waves that, reflected by the tissue particle surface, arrive out of phase at the transducer.

Speckle reduction in ultrasound images is necessary to improve image contrast, which facilitates visual image analysis and image processing tasks such as classification, segmentation, and edge detection. The speckle reduction cannot be excessive as this phenomenon is a consequence of tissue characteristics, and could, for example, lead to a misdiagnosis. Moreover, speckle is considered a deterministic phenomenon, that is, two images acquired under the same conditions will have the same speckle.

There are two types of methods conventionally used to reduce speckle levels: Spatial Filtering and Transform Domain Filtering [1]. With the recent evolution of Machine Learning techniques, specifically Deep Learning [1, 2], they have also started to be used in denoising and speckle reduction of medical images.

The works described in [3–5] and [6] present deep learning techniques that have been used in ultrasound image analysis. Basic tasks such as classification, segmentation or edge detection can be performed with high precision, low computational cost and high-speed using these techniques. In [7] an architecture proposal is made for speckle reduction, preserving the characteristic details of the image. This architecture uses deep learning techniques, based on Generative Adversarial Networks (GANs), among which is the CycleGAN model [8], which was used in this work.

The algorithms, techniques and metrics are described in section II, while the results of the speckle reduction processing are presented in section III, and, finally, section IV presents the conclusions about this work.

## 2   Methods

### 2.1   CycleGAN Model

The Generative Adversarial Networks (GANs) are formed by two distinct types of networks, each one learns different properties and they compete with each other (Fig. 1). Therefore, they are called Generative Adversarial Networks. One of the networks (Generator) generates completely new data (images) from a vector of random values; the other network, the Discriminator, will learn to discriminate whether the data (image) generated by the generating network is real or not. In the end, the generating network manages to generate data (image) with fidelity like the input data of the discriminator network [9].
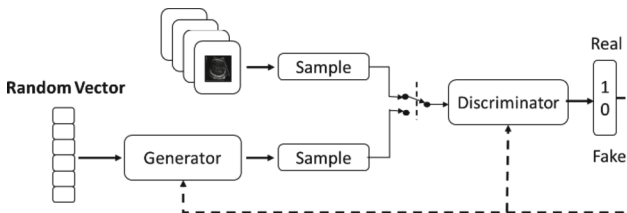


**Fig. 1.** High-level architecture of a GAN-like network

Cycle Generative Adversarial Network is a method that allows capturing the characteristics of an image and discovering how these characteristics can be converted to another

image domain, for which it is not necessary that the same image has a corresponding image in each domain. This architecture is detailed in [8].

This architecture (CycleGAN) is composed of two generator models: a generator that generates images for the first domain and a second generator that generates images for the second domain. Figure 2 shows an example of this technique with ultrasound images.
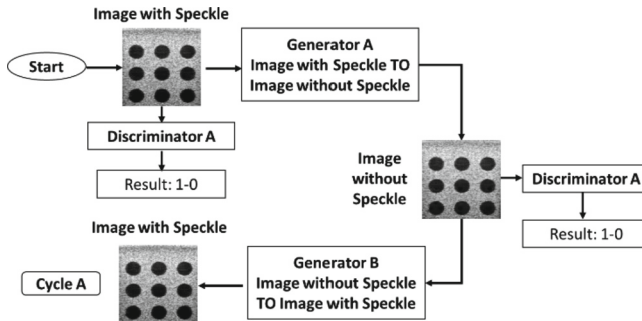


**Fig. 2.** High-level architecture of a GAN-like network

The CycleGAN model uses conditioned adversarial generative networks, that is, the generated model is conditioned by the input image. The part of the generator model is implemented by a network type U-NETs (this network is formed by a convolutional part, in which information is gradually compressed; a de-convolutional part, where it is decompressed, and a series of skip connections that connect each layer at the same level).

The generators perform the image conversion, that is, the image generation process is conditional on an input image from the other domain. Each generator has a corresponding discriminator model.

The discriminator in turn is also a type of convolutional network. However, instead of arriving at a result (single value, scalar) that indicates whether the result is a correct or false image, it returns a portion of the image, in which different patches of the original image are evaluated, a comparison is made between regions of the original and generated image. The result generated by the discriminator will allow to know if some regions of the generated image are similar or not to the original image.

The first discriminator model takes the real images of the first domain and the generator primer, from which it determines whether the generated image is real or false. The second discriminator model performs the same tasks, however, taking images from the other domain. The models, generator and discriminator are trained in the same way as in GAN-type networks. That is, generators learn to deceive discriminators and discriminators learn to better detect false images. Together, the models find a balance during the training process.

Generators not only generate the images in the target domain, but they also create versions of images in the source domain from the generated images. This task is

performed using generated images as input to the corresponding generator model and comparing the output image with the original images.

A Cycle consists of passing an image through the two generators. Together, each pair of generator models are trained to best reproduce the original source image. This process is called cycle consistency (Fig. 3).
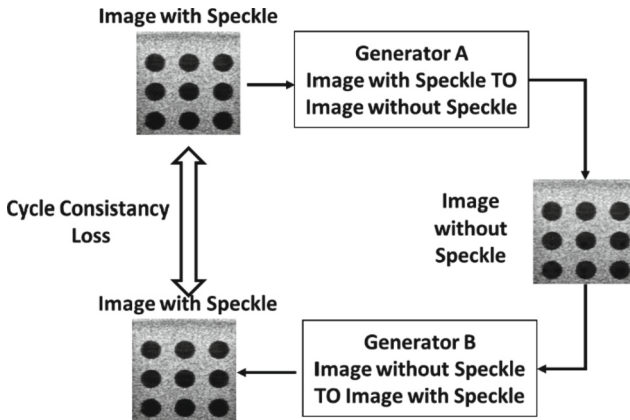


**Fig. 3.** Cycle Consistency Loss

## 2.2   Speckle Filtering Using CycleGAN

In this work, deep learning algorithms (CycleGANs model) were implemented to reduce speckle in ultrasound images. These algorithms were developed using the NVIDIA GTX 980 TI GPU and the Python Tensor-Flow and Keras libraries, due to the multiple advantages they offer, such as: optimization for GPU usage, distributed execution, and code optimization.

In general, deep learning algorithms require a database to train networks. In this work, a dataset consisting of six hundred images was used. Part of these images come from online databases: the contrast-evaluation dataset described in [10] was downloaded from [11] and reconstructed using the USTB toolbox; the breast database described in [12] was downloaded from [13] with permission. A dataset with fetal head images was obtained from [14] and another phantom database was downloaded from [15].

The other part of the training images was generated from data collected in the Center for Biomedical Engineering at UNICAMP (CEB/UNICAMP) using the Ultrasonix platform with a linear transducer and that was later reconstructed in MATLAB® using the classical B-Mode algorithms (Fig. 4).
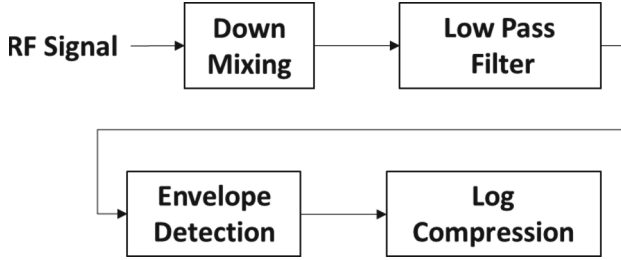
**Fig. 4.** B-Mode processing flow

Initially, the Image Despeckle Filter and Analysis (IDF) toolbox [16–19] was used to eliminate speckle from the dataset images. With this, we obtain images in the two desired domains (with and without speckle). This toolbox is based on Matlab® and provides different filters for speckle reduction, as well as several metrics for image evaluation. This toolbox was also used to calculate the quality metrics of the images generated by the implemented deep learning algorithms (Cycle GANs).

### 2.3 Metrics for Image Quality Evaluation

Signal to Noise Ratio (SNR): compares the desired signal level with the background noise level. The higher the SNR ratio, the lower the influence of background noise. The SNR in decibels is defined according to Eq. 1; where $\sigma_g^2$ corresponds to the variance of the noise-free image and $\sigma_e^2$ corresponds to the error variance (between the original and noise-free image).

$$SNR = 10log\left(\frac{\sigma_g^2}{\sigma_e^2}\right) \tag{1}$$

Peak Signal to Noise Ratio (PSNR): is the ratio between the maximum possible signal power and the power of the corrupted noise that affects the fidelity of its representation. The PSNR calculation for 8-bit grayscale images is shown in Eq. 2, where $\hat{x}$ is the signal estimate of x and MSE is the mean square error (MSE).

$$PSNR(x, \hat{x}) = 10log\left(\frac{255^2}{MSE}\right) \tag{2}$$

## 3  Results

The model training was performed using the dataset as explained in the previous section. An example of input and produced output image of an ultrasound phantom are shown in Fig. 5. Table 1 shows the metrics compared with the Perona-Malik diffusion filter executed using IDF toolbox.

The outputs generated by the CycleGAN algorithm had better SNR and PSNR, that matches results reported in [7]. However, the output images have an observable artifact
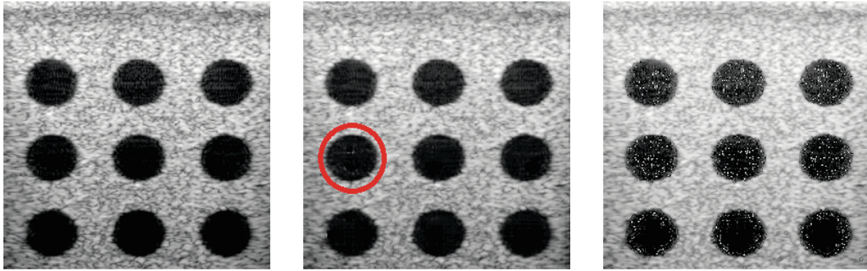
**Fig. 5.** (a) Input image; (b) Filtered image using CycleGAN highlighting the artifacts generated after training; (c) Filtered image using Perona-Malik

(highlighted in Fig. 5(b)), where a hypoechoic region is added to the output image. Similar problems have been reported in [20] and are due to the images used for training that contained large areas of black regions.

**Table 1.** Metrics of speckle reduction

| Algorithm | SNR (dB) | PSNR (dB) |
|---|---|---|
| Perona-Malik diffusion filter – IDF toolbox | 21.4 | 26 |
| CycleGAN | 26.3 | 30.7 |

## 4   Conclusions

The use of CycleGAN for filtering speckle in medical ultrasound images is promising, allowing to achieve better results than with standard techniques. However, our results also show that artifacts can be generated in the output image, requiring a more adequate algorithm training.

## References

1. Pradeep, S., Nirmaladevi, P.: A review on speckle noise reduction techniques in ultrasound medical images based on spatial domain, transform domain and CNN methods. In: IOP Conference Series: Materials Science and Engineering, vol. 1055, International Virtual Conference on Robotics, Automation, Intelligent Systems and Energy (IVC RAISE 2020), Erode, India 2020 (2021). https://doi.org/10.1088/1757-899X/1055/1/012116
2. Kaur, P., Singh, G., Kaur, P.: A review of denoising medical images using machine learning. Curr. Med. Imaging Rev. **14**(5), 675–685 (2018). https://doi.org/10.2174/1573405613666617 0428154156
3. Greenspan, H., Ginneken, B., Summer, R.M.: Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans. Med. Imaging **35**(5), 1153–1159 (2016). https://doi.org/10.1109/TMI.2016.2553401

4. Shen, D., Wu, G., Suk, H.: Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. **19**, 221–248 (2017). https://doi.org/10.1146/annurev-bioeng-071516-044442

5. Vedula, S., Senouf, O., Bronstein, A.M., et al.: Towards CT-Quality Ultrasound Imaging Using Deep Learning (2017). arXiv:1710.06304v1 https://doi.org/10.48550/arXiv.1710.06304

6. Liu, S., Wang, Y., Yang, X., et al.: Deep learning in medical ultrasound analysis: a review. Engineering **5**, 261–275 (2019). https://doi.org/10.1016/j.eng.2018.11.020

7. Mishra, D., Chaudhury, S., Sarkar, M., et al.: Ultrasound image enhancement using structure oriented adversarial network. IEEE Signal Process. Lett. **25**, 1349–1353 (2018). https://doi.org/10.1109/LSP.2018.2858147

8. Zhu, J., Park, T., Isola, P., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networKS. In: IEEE International Conference on Computer Vision (ICCV). Venice, Italy (2017). https://doi.org/10.1109/ICCV.2017.244

9. Pan, Z., Yu, W., Yi, X., et al.: Recent progress on Generative Adversarial Networks (GANs): a survey. IEEE Access **7**, 36322–36333 (2019). https://doi.org/10.1109/ACCESS.2019.2905015

10. Liebgott, H., Rodriguez-Molares, A., Cervenansky, F., et al.: Plane-wave imaging challenge in medical ultrasound. IEEE International Ultrasonics Symposium (IUS). Tours, France (2016). https://doi.org/10.1109/ULTSYM.2016.7728908

11. PICMUS challenge: simulation contrast/speckle test. https://www.ustb.no/ustb-datasets/

12. Yap, M.H., Pons, G., Marti, J., et al.: Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J. Biomed. Health Inform. (2017). https://doi.org/10.1109/JBHI.2017.2731873

13. Breast Ultrasound Dataset B. https://helward.mmu.ac.uk/STAFF/m.yap/dataset.php.

14. Grand Challenge. https://hc18.grand-challenge.org/

15. Insana Lab: Ultrasound Research Interface (URI) at http://ultrasonics.bioengineering.illinois.edu/data_phantom.asp.

16. Loizou, C.P., Petroudi, S., Pattichis, C.S., et al.: An integrated system for the segmentation of atherosclerotic carotid plaque in ultrasound video. IEEE Trans. Ultras Ferroel. Freq. Contr. **61**, 86–101 (2014). https://doi.org/10.1109/TUFFC.2014.6689778

17. Loizou, C.P., Pattichis, C.S., Christodoulou, C.I., et al.: Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery. IEEE Trans. Ultras Ferroel. Freq. Contr. **52**, 1653–1669 (2005). https://doi.org/10.1109/TUFFC.2005.1561621

18. Loizou, C.P., Theofanous, C., Pantziaris, M., et al.: Despeckle filtering software toolbox for ultrasound imaging of the common carotid artery. Comput. Methods Programs Biomed. **114**, 109–124 (2014). https://doi.org/10.1016/j.cmpb.2014.01.018

19. Loizou, C.P., Theofanous, C., Pantziaris, M., et al.: Despeckle filtering toolbox for medical ultrasound video. Int. J. Monitor. Surveill. Technol. Res. **4**, 61–79 (2013). https://doi.org/10.4018/ijmstr.2013100106

20. Cohen, J.P., Luck, M., Honari, S.: Distribution matching losses can hallucinate features in medical image translation. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. LNCS, vol. 11070. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_60

# Methods for Beam Hardening Artifacts Reduction in CT

M. Perez-Diaz(✉) [ID], A. Perez-Duran [ID], Y. Pacheco-Chanfrau [ID], and R. Orozco-Morales [ID]

Universidad Central "Marta Abreu" de Las Villas, 54830 Santa Clara, VC, Cuba
mperez@uclv.edu.cu

**Abstract.** Computed tomography is widely used for disease detection. On many occasions, image quality is affected due to the effect of hardening the X-ray beam, generating artifacts. In the present research, several methods used for the reduction of metallic artifacts (NMAR, ISMAR and LIMAR) were adapted to the reduction of artifacts by X-ray beam hardening, evaluating their effectiveness through quantitative and qualitative metrics of image quality on regions of interest (ROI). The metrics were: signal-to-noise ratio, image contrast, edge visibility, 10% of the MTF and the expert evaluation on a 8 points scale. The different methods were implemented on Matlab. To study its performance, images corrupted with this type of artifact from a physical phantom and two from patients were used. The method that showed the best performance was NMAR, but its computational efficiency for clinical routine is highly dependent on the hardware used. The methods studied do not provide satisfactory results in images that present large areas of very dense tissue.

**Keywords:** Bean hardening artifact · CT · artifact reduction methods · image quality

## 1 Introduction

Artifacts present in computed tomography (CT) images are a serious problem that must be addressed accurately and exhaustively, due to the high quality deterioration that they can cause on the images. The term "artifact" is applied to any systematic discrepancy between the Hounsfield Unit (HU) values in the reconstructed image and the true object (tissue) attenuation coefficients.

The reconstruction technique assumes that all these measurements are consistent, so any error will generally be reflected as an error in the reconstructed image [1]. Design features built into modern CT scanners minimize some types of artifacts, while others can be partially corrected by scanner software.

The effect of the beam hardening is one of the causes of different problems in CT images. It causes a significant loss in the quality of these, since it can hide structures or lesions and lead to a false negative diagnosis. This effect is due to the fact that, as the X-ray beam penetrates the tissue, the less energetic photons are completely absorbed and

thus the beam that passes and emerges has only the content of the highest frequencies of the X-ray spectrum.

Currently, there are several ways implemented to minimize the unwanted effects caused by the hardening of the X-ray beam, such as: bow-tie filtering, calibration correction and the use of correction software [2]. In recent years, multiple methods have been developed with the aim of reducing the negative effects of beam hardening, such as making corrections in the image domain [3], without requiring information on the X-ray spectrum, nor on the type of material being scanned, or using beam consistency conditions [4]. These methods have shown satisfactory results. Another approach is the deduction of the tissue attenuation coefficient by a small number of measurements, using a phenomenological model [5]. Furthermore, recent advances in the field of artificial intelligence have led to several studies involving neural networks [6–8], with very promising results in terms of effectiveness. In the year 2021, through simulation and then through tests with phantoms, an efficient solution was proposed [9], based on the linear combination of two images. It reflects the characteristics of the beam hardening during the transmission of the X-ray polychromatic beam through the material. In this solution only the unknown parameter is fitted by linear optimization. New solutions also appeared in simulations carried out with Monte Carlo [10], as well as other spectral CT methods that allow the creation of virtual monochromatic images through computational decomposition methods [11].

On the other hand, Metal Artifact Reduction (MAR) methods act on the projection space by replacing corrupted projections caused by metallic or very dense structures (generally implants), with interpolations from neighboring noncorrupted projections. Variations of these methods have been devel-oped in order to improve their results, among which we can find Linear Interpolation Metal Artifacts Reduction (LIMAR) [12], Normalized Metal Artifact Reduction (NMAR) [13] and Image Smoothing Method for Artifact Reduction (ISMAR) [14], among others.

The present work focuses on the reduction of the artifacts produced by the phenomenon of beam hardening through the application of some MAR. For three well-known metal artifact reduction methods, they were reprogrammed to bring them into the context of beam hardening artifacts in general.

## 2 Material and Methods

### 2.1 Algorithms

The chosen methods were originally designed for the reduction of metallic artifacts in CT images. The idea of adapting it to beam hardening problem comes from the presence of metallic artifacts is the most extreme possible case of X- ray beam hardening.

In each computational algorithm implementation from each method, it was necessary to implement a new part of the code for its correct performance when dealing with the phenomenon of beam hardening by non-metallic materials. It consists of the fact that after the thresholding process, it was necessary to select the region of dense tissue that produces the artifacts. This step is omitted in the original algorithms because, in the case of images with metallic implants, the entire area of the implants is the cause of the artifacts present in the image. This is not the case that occupies the present research,

where the artifacts are produced only by some parts of the dense tissue. It is necessary to clarify that the correct selection of the dense tissue that affects the image has a high impact on the results of each algorithm.

The LIMAR method is based on a sinogram interpolation technique [12]. The algorithm is based first on the generation of an initial image by reconstruction using FBP and from this an image is generated that contains only the artifact through a thresholding process; in the case of this investigation, what is segmented are the denser tissues (such as bone) in the image. By projecting the image of the artifact (dense tissue) its sinogram is generated. An artifact replacement sinogram is then generated by interpolating through the shadow of the artifact on the original sinogram. Finally, using FBP in the artifact replacement sinogram, a corrected image is obtained to which the previously segmented artifact image is added. Figure 1 details the algorithm.
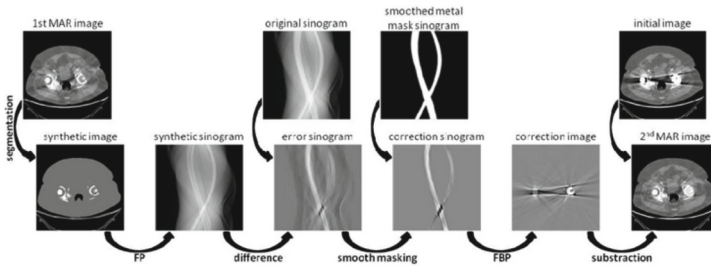


**Fig. 1.** Flowchart of LIMAR

The second method was NMAR [13]. Figure 2 shows a diagram that contains the different algorithm steps of the NMAR method.
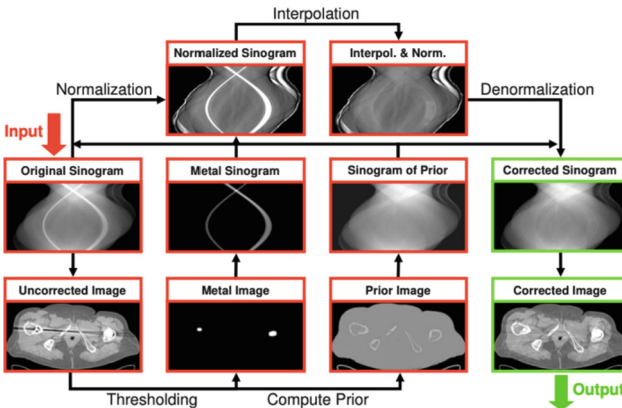


**Fig. 2.** Flowchart of NMAR

First, an uncorrected image is reconstructed using the raw data from the scanner. Using a thresholding technique, an image of the artifact is obtained. In that experiment,

the anterior image is calculated by segmenting soft tissue and bone. The projections of each of these images produce their respective sinograms. The original sinogram is normalized by dividing it by the previous image already projected, this division is carried out by means of a positive value $t_{eps}$ close to 0 as a threshold, to ensure that a division by 0 is not made. Only the values close to the traces of the artifact need to be normalized and denormalized, since only these contribute to the interpolation. The normalized projections $p^{norm}$ are then subjected to a MAR interpolation algorithm M. Next, the corrected sinogram $p^{corr}$ is obtained by denormalizing the already interpolated and normalized sinogram, which is achieved by multiplying it by the values of the projections $p^{prior}$, as described in **Error! Not a valid bookmark self-reference.**, where p also corresponds to the value of the projections and Rf to the Radon transform of the scanned object f.

$$p^{corr} = p^{prior} M p^{norm} = p^{prior} M \frac{p}{p^{prior}} = Rf^{prior} M \frac{p}{Rf^{prior}} \tag{1}$$

In this step, the structure information of the previous image is brought back to the artifact traces. The multiplication and normalization process ensures that there is no imbalance between the original data and the obtained data. After reconstruction, the artifact image is inserted back into the corrected image.

The last method implemented was called ISMAR [14], which is based on the NMAR method, described above, and adds a smoothing method to guarantee final images with better defined borders of anatomical structures. Figure 3 shows a diagram with the different steps, which consist of the segmentation of the traces of the artifact; generating a preview image; the completion of the projection and reconstruction of the images and the application of the smoothing method to obtain a more consistent final result.

To obtain the sinograms of the original, artifact and previous images, projection with parallel beam geometry is used. The sinogram of the original image is then normalized by dividing it pixel by pixel by the sinogram of the previous image. The interpolation of all the values of the sinogram of the original image that coincide with the traces of the artifact present in the sinogram of the image, is carried out following the method described in [14]. This interpolated sinogram is denormalized, obtaining the corrected sinogram, from which the corrected image is obtained by using FBP, linear interpolation and a ramp filter. To obtain a high-quality final image, the L0GM filter is used. This highlights the edges and fine anatomical structures and removes minor artifacts that may still be present.

## 2.2   Images

The set of images used comes from two patients (Images taken free from the internet, which are anonymous) and from a physical phantom. The phantom image was used to perform image quality calculations based on quantitative metrics. The images from patients were used only to appreciate how each method performed under real conditions, so it was subjected to visual inspection by a radiologist expert.

The physical phantom consisted of a 20 cm diameter cylindrical acrylic container. The CT image used for this investigation corresponds to the container filled with water
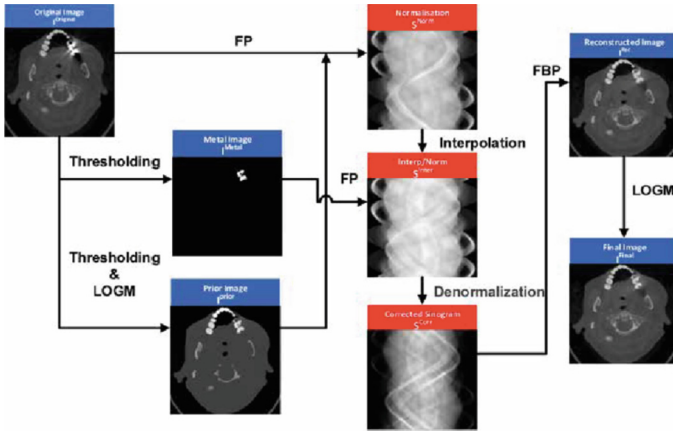
**Fig. 3.** Flow of ISMAR

(low contrast medium) and the insertion of three cylindrical objects, which simulate bone regions. The three inserted objects have a diameter of 25 mm and are composed of hydroxyapatite (HA-200), which is a mineral made up of alkaline calcium phosphate. It was used because in the human body, this material constitutes 99% of the body's calcium and 80% of the total phosphorus. These characteristics allow it to be used to simulate bone tissue.

The images used have a resolution of $512 \times 512$ pixels. Figure 4 shows both the original slice image of the phantom and a real slice image from a patient. Beam hardening artifacts are evident in both.

The selected methods were implemented in Matlab, and programming in its own M programming language. The hardware used consists of a DELL Inspiron 3541 laptop with a quad-core AMD A6 6310 processor, Cores at 1.8 GHz, a card with 8GB of DDR3, RAM at 1333 MHz and AMD Radeon R4 Graphics video chipset.
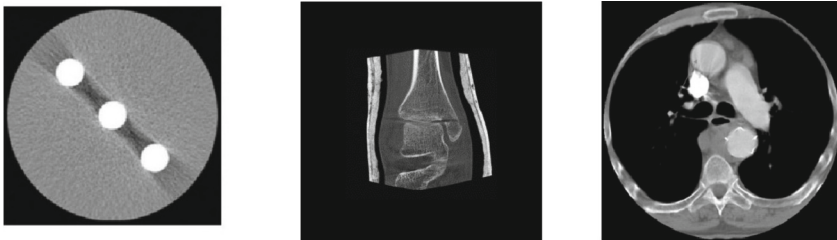


**Fig. 4.** Images

## 2.3 Image Quality Metrics

Different 3 mm regions of interest (ROI) were selected on the phantom image, both the original and those processed with each method, for the calculation of objective metrics

of image quality, in which area ROI stand out on areas of useful information (1) and areas on the beam hardening artifact (2), background ROI (3), and edge ROI (4) that also serves to calculate the MTF (where a sudden contrast change is evident in an edge region of the image). For each type of ROI, the average of the measurements was calculated.

In the selected ROI, several image quality measures were calculated, such as the signal-to-noise ratio area (SNRA) in ROI 1 and 2, image contrast (Cima) using ROI 1 and 3, edge visibility (V) in the ROI 4 and 10% of the MTF (known as detection limit) in the characteristic MTF curve, as an estimator of the spatial resolution that is lost when running each method, (ROI 4) [15].

To assess the performance of the three beam hardening artifact reduction methods tested in real conditions, a tomographic slice of a patient knee was used. The original image with artifact and those processed by the three methods were used. These were shown to an expert observer with more than 10 years of experience in Radiology. He observed each image for an unlimited time, under the same viewing conditions (environmental lighting level of 100 lx, the same monitor and distance to it of 50 cm). The panel of a DELL Inspiron 3541 laptop was used, with AMD Radeon (TM) R4 Graphics controller, which has a size of 15.6 inches, a resolution of 1366 x 768 pixels and a depth of 32 bits. The observer did not know in each case if the image was original or processed by each of the three methods, but was able to go back and forth to compare the images as many times as needed. Each image was evaluated on an 8-point scale, where: 8- No artifacts or distortion. 7- No artifact and slight distortion. 6- No artifact, but very distorted. 5- With traces of artifacts and slight distortion. 4- With artifacts and without distortion. 3- With distortion and new artifacts.2- With full artifact and great distortion.1- Considerable losses of useful information.

Good image quality was defined as that in which, having eliminated or significantly reduced the beam hardening artifact, the image contrast was maintained or improved with respect to the original image, there was high visibility of edges, a low noise level and good spatial resolution in general, coupled with a good subjective assessment of image quality.

## 3    Results and Discussion

Table 1 shows the results of image quality metrics on the different ROI selected on the phantom image.

Firstly, it can be seen that all the methods produce an increase in the SNR in the area object, due to the fact that in the process of removing artifacts, noise is also removed from the original image, a highly positive aspect from the point of view of improving image quality.

The drastic reduction seen in the SNR and image contrast values in the areas affected by the artifacts is due to the fact that in the original image these areas were covered by dark bands on the background, and when processed, the methods eliminate the artifact, leaving a minimal signal on the background. In essence, the trace of the artifact on the sinogram has been removed, replacing its content with a more uniform one.

From the point of view of the image contrast in the area of the artifact, it is expected that the method that works best is the one whose value in that area is close to 0, since

**Table 1.** Objective metrics for image quality evaluation

| Images | SNR Objet | Cima Objet (%) | SNR Artifact | Cima Artifac (%) | Edge Visibility | 10% of MTF |
|---|---|---|---|---|---|---|
| Original | 16.031 ± 0.001 | 70.254 ± 0.003 | 1.095 ± 0.855 | 18.863 ± 15.788 | 15.014 ± 4.163 | 0.61 |
| LIMAR | 27.711 ± 0.000 | 70.393 ± 0.000 | 0.809 ± 0.803 | 2.424 ± 1.235 | 17.853 ± 5.576 | 0.44 |
| ISMAR | 28.213 ± 0.000 | 71.516 ± 0.000 | 0.818 ± 0.079 | 3.096 ± 0.457 | 17.203 ± 3.887 | 0.43 |
| NMAR | 28.227 ± 0.000 | 72.480 ± 0.000 | 0.274 ± 0.112 | 0.991 ± 0.509 | 15.746 ± 3.412 | 0.70 |

this indicates that when the artifact is removed, the signal that remains is similar to the background signal. The method that best met this requirement was NMAR, from the metrics point of view.

The visibility at the edges of the objects had improvement in each method with respect to the original image, which is deduced from the thresholding process, in which the image is segmented to define which part of it produces the beam hardening effect. In this case, it was LIMAR that preserved the edges better, since, by reducing noise less, it softened the images less.

Figure 5 shows the different MTF curves for the original image and each one of those processed by the three methods. Note the detection limit with a red line. In LIMAR and ISMAR there were some frequencies for which low contrast content was not visible.
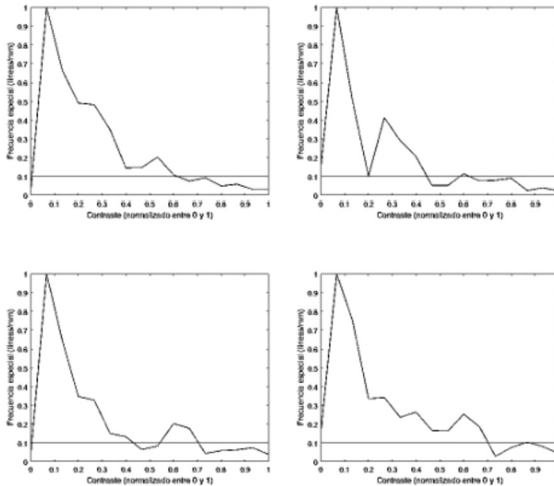


**Fig. 5.** MTFs Original (left up), LIMAR (right up), ISMAR (left down) and NMAR (right down)

In general, taking into account the results obtained by metrics, it can be verified that the method with the best overall results in this experiment has been NMAR. This method

provided results slightly superior to ISMAR for some variables and clearly superior to LIMAR.

Figure 6 shows the original image and the processed ones. All three methods were successful in removing beam hardening artifacts, with slight differences between each of them. As a limitation of all, it should be noted that slight line artifacts were generated on the background by smoothing the texture of the noise.
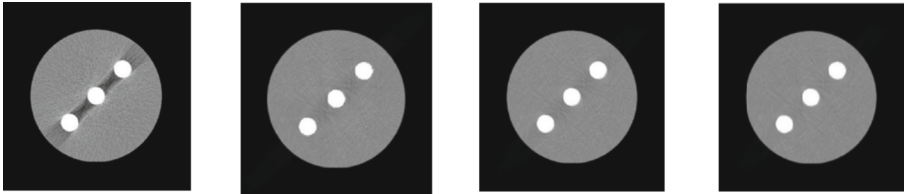


**Fig. 6.** Phantom images Original, LIMAR, ISMAR and NMAR.

In the subjective evaluation carried out by the expert radiologist, the original real images received an evaluation of 4 points, the ones processed with LIMAR received 3 points, and ISMAR and NMAR 5 points.

Figure 7 shows the performance of the three methods on an original image with high bone content. Beam hardening artifacts are generated on the surrounding muscle tissue. Multiple artifacts of this type can be seen in the original image (represented by blue arrows), consisting of dark bands. The observer considered that the original image has artifacts, but not distortion. LIMAR discreetly softens the noisy texture of the image. However, this method only reduced some of the original artifacts. On the other hand, it produced new artifacts (yellow arrows) that are due to the process of thresholding and subsequent unification of the image. ISMAR left traces of the original artifacts and introduced some distortion by smoothing out the noise. This aspect must be adjusted by optimizing the algorithm. However, a set of areas in the image where beam hardening artifacts originally existed were adequately treated by the method. Red arrows indicate regions where artifacts were successfully removed. In general, it can be argued that this method reduced part of both dark and bright artifacts. Like the rest of the algorithms, it generated line artifacts, but it did not generate any extraneous artifacts that might mask the useful signal. NMAR regarding the removal of artifacts, had a visual behavior slightly inferior to ISMAR, presenting the same drawbacks. However, in the opinion of the expert, it was the one with the best overall performance because it managed to better maintain the texture of the original background and the smoothing did not lead to marked distortions.
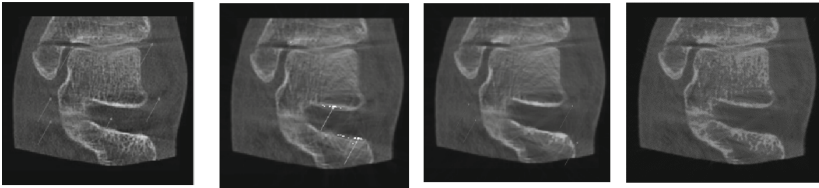
**Fig. 7.** Patient knee images: Original, LIMAR, ISMAR and NMAR

When an image is taken where the dense content is not as scattered as in the Fig. 8, the results improve even more. Figure 8 shows a CT slice showing artifacts created by the image of the vena cava on the surrounding organs and tissues. Note the performance of each method.
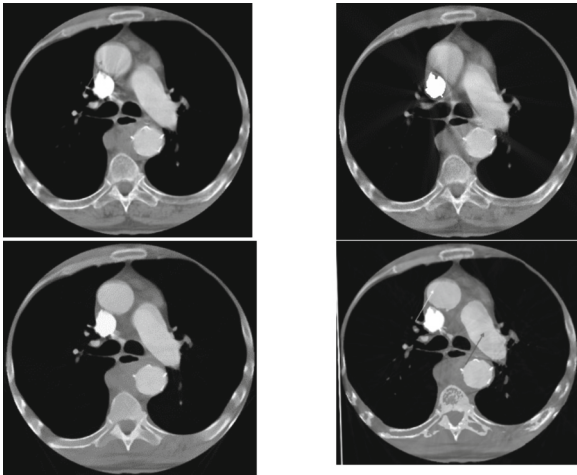


**Fig. 8.** Patient CT slice: Original (left up), LIMAR (right up), ISMAR (left down) and NMAR (right down)

Table 2 shows the computational cost of each algorithm per CT slice using the hardware described above.

**Table 2.** Algortithm's computational cost

| Algorithms | Time per image (min) |
|---|---|
| LIMAR | 5 |
| ISMAR | 20 |
| NMAR | 25 |

It can be seen that the implemented methods implied a high computational cost for the computer and software used. If one takes into account that a complete tomography can have between 60 and 120 slices, it can be seen that none of the proposed methods could be executed under efficient clinical routine conditions, with hardware and software similar to the one used in this research.

As a test, ISMAR has been run from Matlab on an Intel Core i5 computer, with a 1.80 GHz CPU and 4 GB RAM. A time of 40 s per image was obtained, which implies about 20 min of processing for a complete CT. Although this cannot yet be said to be efficient under routine clinical conditions, it is a significant advance. The next step in making methods efficient would be to pass them into a language like Python. The justification for doing this is that the biggest problem with these methods is found during the image reconstruction process, at the time of running the Radon transform in Matlab, which is not optimal. Python, for example, has libraries focused directly on image processing, which allow this transformation to be run in parallel for all projections, significantly reducing computation times.

In the consulted scientific literature, there are methods that reduce the artifacts produced by the hardening effect of the X-ray beam, which take into account the high amounts of dense regions that may appear in some CT slices [1–12, 16]. The success of some of these methods lies in the estimation of the correct polychromatic spectrum and in the use of novel techniques such as the incorporation of artificial intelligence applications, simulations or virtual images.

The strategy adopted in this research of taking methods that have been successfully tested to reduce metallic artifacts, which constitute an extreme example of beam hardening, works well when the content of the dense region that causes the beam hardening artifact is minor on the image or is concentrated in area. This follows the logic that, in most cases of metal artifacts, the trace metal content in the images is very low, so the thresholding and interpolation process is not a major inconvenience. This is not true in all cases, when the beam hardening artifact is caused by dense tissue on soft tissue.

Some algorithms for reducing beam hardening artifacts are incorporated into the software of the latest generation scanners, which is an advantage, because the image is processed during the reconstruction itself, saving time. However, there is generally no possibility to adjust them for use in different conditions, because the source code is not free.

For the purposes of this research, the most effective variant found to remove beam hardening artifacts was NMAR. The method works better the more concentrated in area the generating zone of the artifact is. The ISMAR method performs as well for these cases as NMAR in terms of artifact reduction, but smoothest the rest of the background and useful signal content of the image to a greater extent. The LIMAR method did not have good results in this investigation.

Future continuation of this work will be aimed at optimizing the NMAR and ISMAR methods by including an additional thresholding technique, to avoid the loss of valuable information on low-contrast tissue, as well as migrating the algorithms to a language of free programming such as Python, which guarantees greater computational performance and savings in terms of software licenses.

## 4 Compliance with Ethical Requirements

The authors declare that they have no conflicts of interest among themselves or with any institution.

Informed consent does not apply in this case since only an image of a physical phantom and two public real anonymous and free images on the internet were used.

## 5 Conclusions

The three implemented methods manage to eliminate beam hardening artifacts.

The method that offers the best overall image quality results both from an objective and subjective point of view is NMAR, with a similar performance to ISMAR in the task of reducing artifacts, but superior in the signal-to-noise ratio, image contrast and spatial resolution.

The methods studied do not provide satisfactory results in images that present large areas of very dense tissue such as bone, due to the loss of useful information that is produced by the flattening of the sinogram.

## References

1. Barrett, J., Keat, N.: Artifacts in CT: recognition and avoidance. Radiographics **24**, 1679–1691 (2004). https://doi.org/10.1148/rg.246045065
2. Luo, S., Wu, H., et al.: A fast beam hardening correction method incorporated in a filtered back-projection based MAP algorithm. Phys. Med. Biol. **62**, 1810–1830 (2017). https://doi.org/10.1088/1361-6560/aa56b5
3. Sarkar, S., Wahi, P., Munshi, P.: An empirical correction method for beam-hardening artifact in Computerized Tomography (CT) images NDT E Int. **102**, 104–113 (2019). https://doi.org/10.1016/j.ndteint.2018.11.009
4. Abdurahman, S., Frysch, R., et al.: Beam hardening correction using cone beam consistency conditions. IEEE Trans. Med. Imaging **37**, 2266–2277 (2018). https://doi.org/10.1109/TMI.2018.2840343
5. Zhao, W., Li, D., et al.: Multi-materials beam hardening artifacts correction for computed tomography (CT) based on X-ray spectrum estimation. arxiv.org, 2018 (2021). https://arxiv.org/abs/1812.02365. Accessed 3 July 2021
6. Ghani, M., Karl, W.: Deep learning based sinogram correction for metal artifact reduction. Electron. Imaging **XVI**, 4721-4728 (2018). https://doi.org/10.2352/ISSN.2470-1173.2018.15.COIMG-472
7. Gjesteby, L., Shan H., et al.: A dual-stream deep convolutional network for reducing metal streak artifacts in CT images. Phys. Med. Biol. **64**, 23503 (2019). https://doi.org/10.1088/1361-6560/ab4e3e
8. Pauwels, R., Cao, W., et al.: Exploratory research into reduction of scatter and beam hardening in industrial computed tomography using convolutional neural networks. In: Proceedings of the 9th International Conference on Industrial Computed Tomography, pp. 1–8. Padova (2019)
9. Jin, H., Dongjoon, K., Yeong-Gil, S., Ho, L.: Metal artifact reduction method based on a constrained beam-hardening estimator for polychromatic x-ray CT. Phys. Med. Biol. **66**, 065025 (2021)

10. Nazemi, E., Six, N., Iuso, D., De Samber, B., Sijbers, J., De Beenhouwer, J.: Monte-carlo-based estimation of the X-ray energy spectrum for CT artifact reduction. Appl. Sci. **11**, 3145 (2021). https://doi.org/10.3390/app11073145

11. Inkinen, S., Mikael, A., Juntunen, K., Ketola, J., Korhonen, K., Kotiaho, A., et al.: Virtual monochromatic imaging reduces beam hardening artefacts in cardiac interior photon counting computed tomography: a phantom study with cadaveric specimens. Biomed. Phys. Eng. Express **8**, 015029 (2021)

12. Koehler, T., Brendel, B., et al.: A New Method for Metal Artifact Reduction in CT. CT Meeting. Computer Science, pp. 29–32 (2012)

13. Meyer, E., Raupach, R., et al.: Normalized metal artifact reduction (NMAR) in computed tomography. Med. Phys. **37**, 5482–5493 (2010). https://doi.org/10.1118/1.3484090

14. Rodríguez-Gallo, Y., Orozco-Morales, R., Pérez-Díaz, M.: Methods for reducing metal artifacts in computerized tomography nucleus **65**, 11–15 (2019). https://www.researchgate.net/publication/335402443

15. Ruiz-Gonzalez, Y., Perez-Diaz, M., et al.: Objective measurements of image quality in synchrotron radiation phase-contrast imaging vs. digital mammography. Int. JCARS (2015). https://doi.org/10.1007/s11548-015-1237-7

16. Shi, H., Yang, Z., Luo, S.: Reduce beam hardening artifacts of polychromatic X-ray computed tomography by an iterative approximation approach. J. Xray Sci. Technol. **25**, 417–428 (2017). https://doi.org/10.3233/XST-16187

# Comparison Between Online and Offline Independent Component Analysis in the Context of Motor Imagery-Based Brain-Computer Interface

Paulo Viana[1]([✉]) [iD], Romis Attux[1] [iD], and Sarah N. Carvalho[2] [iD]

[1] School of Electrical and Computer Engineering, University of Campinas, Campinas, SP 13083-852, Brazil
`p263889@dac.unicamp.br`
[2] Department of Electrical Engineering, Federal University of Ouro Preto, João Monlevade, MG 35931-008, Brazil

**Abstract.** Independent component analysis (ICA) is an important tool for recovering brain activity sources from electroencephalogram (EEG) signals. It is a blind technique, hence does not require that reference signals be available. Most ICA algorithms are considered, to a certain extent, offline, as they demand parameter adjustment before the actual filtering step in new data. Online Recursive ICA (ORICA) was proposed as an adaptive algorithm that can estimate filter parameters in a real-time manner. The approach has the advantage of being more capable of tracking changes in brain dynamics and source changes. In this work, we perform a comparative analysis of six offline ICA algorithms and ORICA in the context of EEG signal preprocessing for motor imagery-based brain-computer interfaces (BCIs). The experimental results show that there was no significant difference between ORICA performance and the best offline methods, validating the feasibility of its use in motor imagery BCI.

**Keywords:** independent component analysis · blind source separation · electroencephalography · linear discriminant analysis · classification · power spectral density

## 1 Introduction

Electroencephalography (EEG) signals can be used within a variety of setups and applications. In the context of brain-computer interfaces (BCIs), different paradigms make use of the brain physiology and behavior under distinct stimuli [1]. The Motor Imagery (MI) paradigm is typically built around the premise of a subject imagining movements of limbs or tongue while his/her brain electrical activity is recorded. This activity is then classified and the motor task labels are associated with the commands of the focused application [2,3]. Left- and

right- foot, as well as arm imagery, are some of the most common MI instances in the literature [4]. It should be noted that a setup of this kind can be used in a Brain-Computer Interface to control either physical or virtual devices.

The brain patterns that occur when each task is performed are related to specific brain regions. The MI task is associated with event-related desynchronization or synchronization (ERD and ERS, respectively) in the motor cortex, predominantly in the mu (8 Hz–13 Hz) and beta (13 Hz–25 Hz) bands. These bands are defined frequency intervals in which the brain activity is detected, and the amplitude of the recorded signals is affected by the task at hand. Moreover, since the neuron path is crossed at the medulla, the response of the left and right-hand imagery can be seen in the contralateral hemispheres of the brain [5].

This physical separation of activities conditioned by the task being performed can be used to the advantage of the control system in a BCI. This will result in signals from certain electrodes containing more information about a certain task than others. However, the electrodes record a combination of different brain sources, since EEG has a low spatial resolution. Also, many other electrical physiological processes can affect EEG recording, such as ocular, muscular and cardiac activity. There are also non-physiological sources that can contaminate EEG signals, such as line noise and movement artifacts.

Independent Component Analysis (ICA) can be performed via a family of algorithms that have the goal of separating statistically independent sources from a set of observed mixtures of them [6], and is recurrently used for MI-BCI [7–9]. In general, it is assumed that the sources are linearly combined into each observed signal, which allows separation by finding the inverse of this mixing linear system. Each ICA algorithm may employ different assumptions about the characteristics of the sources, and, as a consequence, may yield different results. In a BCI application, we select the independent components (sources) that contain more information about the task being performed. However, most techniques assume stationarity of sources, which is not guaranteed in real EEG data [10].

Online Recursive ICA (ORICA) is an online algorithm that whitens the observed signals and finds the inverse of the mixing matrix with a Recursive Least Squares algorithm [11]. With this approach, changes in source patterns and non-stationarity can be addressed, since the ICA model is adaptively updated using the natural gradient of the Infomax ICA rule while the forgetting factor of the algorithm controls the adaptivity to new data and stationarity changes.

In this work, we evaluate the performance of the ORICA algorithm against six offline ICA algorithms in two classes of MI-based BCI. Source separation is used as a preprocessing step for the BCI, followed by feature extraction, feature selection, and classification by a machine learning method. We compare these methods in terms of Cohen's kappa [12], as it is the metric used in the competition which the dataset is from.
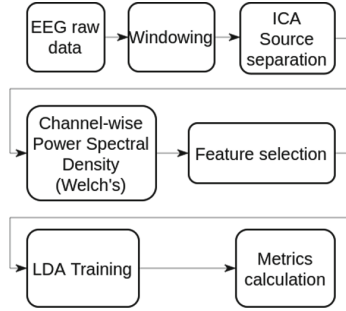
**Fig. 1.** Complete MI-based BCI pipeline.

## 2  Materials

### 2.1  Dataset

Our analysis were performed using the public database BCI Competition IV 2a [13], which has open access and is widely used in Motor Imagery benchmarks. This dataset contains two sessions of a set of motor imagery tasks. Each session comprises 6 runs, each one with twelve trials of the four classes. This yields 288 trials per session, 72 for each MI task. The tasks are the imagination of movement of left-hand, right-hand, both feet and tongue. In this work, we use only the left and right-hand imagination tasks. Each trial begins with a beep at $t = 0$ s, to capture the subject's attention, then at $t = 2$ s a visual cue is displayed on a screen display and stays for 1.25 s, which prompts the subject to perform the movement imagination for 4 s. A diagram of the steps for each trial is shown in Fig. 2.

The dataset was recorded using the international 10–20 system, with 22 EEG electrodes, with a sampling rate of 250 Hz. The signal is bandpass-filtered between 0.5 Hz and 100 Hz, and a notch filter at 50 Hz is used to suppress line noise. In addition, three EOG channels were recorded using the same setting. Apart from the MI runs, there was also an approximately 5 min session of EOG data collection that contained three blocks: two minutes with eyes open, one minute with eyes closed and one minute with eye movements.

In our experiments, this first session was used for adjusting the offline ICA methods and for warm starting the ORICA algorithm, selecting the best features (with the strategy explained in Subsect. 2.5) and training the classifier. All metrics reported referring to the test set, were not used during the training stage. The pipeline shown in Fig. 1 is done for each subject individually, characterizing an intra-subject training protocol. For each subject, each ICA algorithm is run ten times for the algorithms with stochastic design and one time for the deterministic methods, so average metrics can be calculated with confidence intervals.
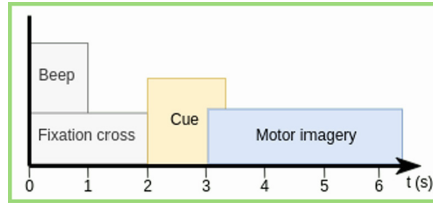
**Fig. 2.** Data acquisition protocol, adapted from [13].

## 2.2   Independent Component Analysis

Independent Component Analysis (ICA) can be defined as the problem of recovering a number of sources (or independent components), a set of mutually independent random variables, from a set of mixtures/observations. In the linear case, the mixed observed variables $\mathbf{x}$ can be expressed as $\mathbf{x} = A\mathbf{s}$, where $\mathbf{s}$ are sources that have been mixed, $A$ is the mixing matrix of the model, in matrix notation. Only $\mathbf{x}$ is observed, with $A$ also being unknown, and the goal is to estimate both $A$ and sources $\mathbf{s}$. Generally, the sources and observed variables are random vectors, and some ICA methods do not explore the time structure of $\mathbf{s}$ and $\mathbf{x}$, while other do.

We compare many offline ICA methods to the online approach. The Infomax [14] algorithm has the objective of extracting sources that have the least Mutual Information and has an "extended" variant [15] that was designed to better separate sub and super-gaussian sources. The Second-Order Blind Identification (SOBI) method [16] separates the sources with the constraint of the covariance matrices between sources being similar to the identity matrix (uncorrelated sources). The method called Joint Approximate Diagonalization of Eigenmatrices (JADE) tries to diagonalize the fourth-order cumulant tensor of the estimated sources [17] and can be seen as a non-linear decorrelation. Fast ICA is a fixed-point method that assumes that sources are non-Gaussian and finds the unmixing matrix that maximizes source non-Gaussianity, represented by the absolute value of the signal kurtosis. Picard (Preconditioned ICA for Real Data) [18] approximates the Hessian of the log-likelihood function for preconditioning the numerical optimization and has the same optimization objective as the Infomax algorithm. For all ICA methods we assumed that the number of independent components is the same as the number of channels (represented by $N$).

Infomax, Extended Infomax, FastICA and Picard are stochastic i.e. they use random initialization states which makes them yield different results each time they are used. On the other hand, SOBI, JADE and ORICA do not have this stochasticity.

## 2.3   Online Recursive Independent Component Analysis (ORICA)

ORICA is a recursive algorithm for incremental estimation of independent components [11], based on the natural gradient of Infomax. An online algorithm has

the advantage of adapting to source nonstationarity, with a lesser need for prior parameter adjusting. Hsu et al. [19] proposed a block update rule for the original algorithm that can increase processing speed without sacrificing performance and used it together with a recursive least squares (RLS) whitening algorithm. This online whitening improves the ICA convergence, and removes the need for a prewhitening step that generally is needed for ICA.

The recursive whitening method is followed by the recursive source separation [19]. The sources $\hat{\mathbf{s}}_i$ are recovered using Eq. 1. In this equation $M_i$ is the whitening matrix, $W_i$ is the source separation matrix, the $i$ index indicates the iteration number, and both matrices are iteratively updated. Firstly, the signal is whitened, so the goal is to find $W_i$ and $M_i$ so that $\hat{\mathbf{s}}_i = W_i \cdot \mathbf{v}_i$, where $\mathbf{v}_i = M_i \cdot \mathbf{x}_i$, $\hat{\mathbf{s}}_i$ is the sources estimate (or independent components, ICs) and $\mathbf{x}_i$ is the recorded multi-channel EEG.

$$\hat{\mathbf{s}}_i = W_i \cdot M_i \cdot \mathbf{x}_i \tag{1}$$

The whitening matrix $M_i$ is calculated using an RLS algorithm that decorrelates dependent signals [20]. The whitening matrix update rule is shown in Eq. 2. In this equation $\mu_i$ is the forgetting factor.

$$M_{i+1} = M_i + \frac{\mu_i}{1 - \mu_i}\left[I - \frac{\mathbf{v}_i \cdot \mathbf{v}_i^T}{1 + \mu_i(\mathbf{v}_i^T \cdot \mathbf{v}_i - 1)}\right]M_i \tag{2}$$

The unmixing matrix $W_i$ is iteratively updated using Eq. 3. In this equation $\hat{\mathbf{s}}_i = W_i \cdot M_i \cdot \mathbf{x}_i$, $i$ is the time index, and $f$ is a component-wise nonlinear function that depends on source assumptions. For separating supergaussian sources, a typical choice is $f^+(\mathbf{x}) = -2 \cdot tanh(\mathbf{x})$, and for subgaussian sources, the typical nonlinearity would be $f^-(\mathbf{x}) = tanh(\mathbf{x}) - \mathbf{x}$ [11]. ORICA assumes a fixed number $k$ of supergaussian sources, and for $k$ independent components (IC) estimates $f^-$ is applied, while $f^+$ is used for the other $N - k$ independent components.

$$W_{i+1} = W_i + \frac{\mu_i}{1 - \mu_i}\left[I - \frac{\hat{\mathbf{s}}_i \cdot f^T(\hat{\mathbf{s}}_i)}{1 + \mu_i(f^T(\hat{\mathbf{s}}_i) \cdot \hat{\mathbf{s}}_i - 1)}\right]W_i \tag{3}$$

The block update rule proposed in the literature [19] for the original ORICA is a modification to Eq. 3, so it can be run in batches of time samples, leading to an approximate algorithm that has improved speed. This update rule is shown in Eq. 4.

$$W_{i+L} \approx \left(\prod_{l=i}^{i+L-1} \frac{1}{1 - \mu_l}\right) \cdot \left[I - \sum_{l=i}^{i+L-1} \frac{\hat{\mathbf{s}}_l \cdot f^T(\hat{\mathbf{s}}_l)}{\frac{1 - \mu_l}{\mu_l} + f^T(\hat{\mathbf{s}}_l) \cdot \hat{\mathbf{s}}_l}\right]W_i \tag{4}$$

After each $W$ update, the orthogonal transformation shown in Eq. 5 is applied to it to guarantee orthogonality. In this equation $D$ is a matrix containing the eigenvalues of $W_{i+1}W_{i+1}^T$ in its diagonal and $E$ is a matrix containing the eigenvectors of $W_{i+1}W_{i+1}^T$.

$$W_{i+1} \leftarrow (DE^{-\frac{1}{2}}D^{-1})W_{i+1} \tag{5}$$

The parameter $\mu_i$ represents the forgetting factor of the algorithm, and it adjusts the weight that newer samples have in the $W$ matrix update. We used a cooling heuristic [19] for adjusting this value, as shown in Eq. 6. In this equation, $i$ is the iteration number, $\mu_0$ is a chosen initial value for $\mu_i$ and $\eta$ is a cooling factor. Lin et al. [21] applied the ORICA algorithm for separating electroencephalography (EOG) artifacts from EEG, and used $\eta = 0.6$, $\mu_0 = 0.995$ and $L = 8$. We used the same values in the training phase since EOG is one of the main contaminants in EEG signals. For the testing phase, it was set to $\mu_i = 0.001$ to simulate a real application scenario and mitigate the problem with the cooling heuristic in which $\mu_i \to 0$ when $i \to \infty$.

$$\mu_i = \frac{\mu_0}{i^\eta} \tag{6}$$

## 2.4   Feature Extraction

Hand MI produces short-lasting event related desynchronizations (ERD) in the mu (8–13 Hz) and beta (13–25 Hz) bands in the motor cortex (parietal-lobe). This activity in these specific bands is important for MI-based BCI systems, since it can be originated in the sensorimotor cortex [2], and thus can be used to infer if an imagination event is happening. The activity magnitude in each band can be estimated using a frequency analysis of the signal, and used for the motor-imagery paradigm [2]. In this work, we use Welch's method for spectral power density estimation over these bands, which can be formulated as in Eq. 7.

$$\hat{S}(\omega) = \frac{1}{K N_H U} \sum_{k=1}^{K} \left| \sum_{n=1}^{N_H} H(n) x(n + kD) e^{-j\omega n} \right| \tag{7}$$

In this equation, $K$ is the number of segments of length $N_H$ that the signal $x$ is segmented, each one shifted by $D$ points. $H(n)$ is a window function and $U$ is defined as in Eq. 8. It can be interpreted as the average spectral density over smaller windows of the original signal. We used a $N_H = 1$ s window, with $D = 250$ ms, windowed with a Hamming window.

$$U = \frac{1}{N_H} \sum_{n=1}^{N_H} |H(n)|^2 \tag{8}$$

The power over the mu and beta bands are averaged, yielding two values. This process is done for every estimated EEG source, yielding a feature vector of 44 values.

## 2.5   Feature Selection and Classification

Feature selection is made by using a wrapper [22]. This strategy works by iteratively adding features to the selected feature set, in a hill-climbing (greedy) way: each iteration a feature is added based on some performance metric of a model that uses this feature, in contrast to one that doesn't. The general algorithm works as follows.

1. Let $\mathbb{S} = \{\}$ be the selected feature set and $\mathbb{F} = \{f_1, f_2, ..., f_{N_f}\}$ be the complete feature set.
2. For each feature $f_i$ in $\mathbb{F}$, get it's score by running a classification training task (using cross-validation) using features $\mathbb{S} \cup \{f_i\}$.
3. Add the feature with the best score to $\mathbb{S}$ and remove it from $\mathbb{F}$.
4. Repeat 2) and 3) until $\mathbb{S}$ has the desired cardinality $K_{max}$ or the score stops improving.

Feature's scores were calculated as the average Cohen's kappa metric that a Linear Discriminant Analysis (LDA) classifier achieves when training with $\mathbb{S} \cup \{f_i\}$ in the K-fold scheme, with four-folds.

The selected features are used in another classifier as predictors, to predict the respective MI task. In our pipeline, the number $K_{max}$ of selected features was a training hyperparameter, which was optimized using a four-fold cross-validation over the training set. Then the final classifier is retrained in the entire training set, and evaluated in the test set. We used Linear Discriminant Analysis (with the singular vector decomposition solution) for both the wrapper and classifier, since it is one of the most used classification algorithms in BCI systems for their simplicity and low computational cost [23]. This machine learning technique consists in finding a hyperplane that best separates a set of $n$-dimensional points, by maximizing intra-class variance and minimizing inter-class variance.

## 3    Results

Figure 3 shows the Cohen's kappa classification metric for each ICA preprocessing, for each subject of the test database. The height of the bars represents the average kappa (for the deterministic ICAs, it corresponds directly to the Cohen's kappa). The black line above the bars indicates the bootstrapped confidence interval, and the presence of the red asterisk, below each bar, indicates that the kappa is significantly smaller than the best method for each subject. The bar labeled "None" indicates that no source separation was performed i.e. feature extraction was performed over the raw EEG signals, and serves as a baseline. Method "ORICA (0)" refers to ORICA assuming 0 subgaussian sources and "ORICA (1)" assumes 1 subgaussian source.

For Subjects 5, 7 and 8, ORICA was the best performing method. For Subject 9, it was the second-best performing method, and for Subject 2, it was statistically equivalent to the best method. For Subjects 1, 3, 4 and 6, ORICA presented a performance significantly lower than the best method. Figure 4 shows the summarization of the methods that had the best kappa (indicated in the bars), for each subject.
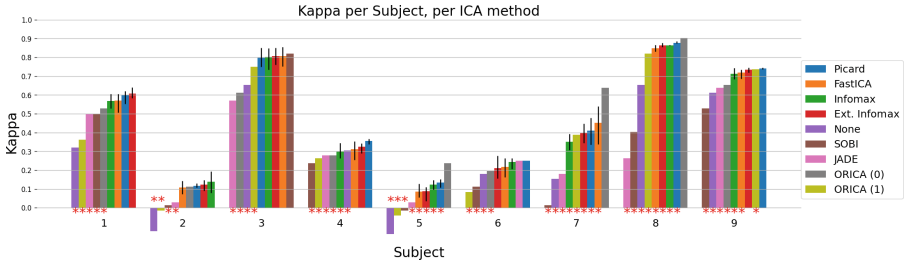
**Fig. 3.** Kappa for each subject using the eight ICA approaches and the case without ICA.
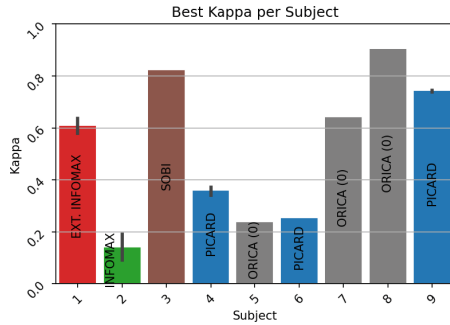


**Fig. 4.** Best performing ICA method for each subject.

Table 1 shows the average and standard deviation of the kappa metric over all subjects for each method, and the p-value for the paired two-sided Wilcoxon test with the null hypothesis that the distribution medians are equal, comparing all algorithms with ORICA (0) (for being the online method that had the highest

**Table 1.** Average kappa and standard deviation for each ICA method.

| Algorithm | Mean | SD | p-value |
|---|---|---|---|
| None | 0.290 | 0.291 | <0.001 |
| SOBI | 0.290 | 0.275 | <0.001 |
| JADE | 0.304 | 0.210 | <0.001 |
| ORICA (1) | 0.372 | 0.316 | <0.001 |
| Infomax | 0.456 | 0.279 | 0.898* |
| FastICA | 0.457 | 0.293 | 0.919* |
| ORICA (0) | 0.461 | 0.253 | – |
| Ext. Infomax | 0.462 | 0.290 | 0.997* |
| Picard | 0.476 | 0.279 | 0.707* |

average kappa). We conclude that ORICA was statistically equivalent to 4 out of the 5 top methods (indicated by an asterisk), with no statistical difference to Picard, the best average-performing ICA in terms of kappa.

## 4   Conclusion

The algorithm ORICA was shown to be a robust online method for source separation when compared to some common offline ICA algorithms. Offline methods need a reliable set of EEG recordings to estimate the unmixing matrix, which is kept unaltered during signal filtering. This could lead to errors as time passes since the brain is not a stationary system and changes in source behaviours could affect the filter performance. The online method has the advantages of being adaptive to nonstationarities, changes in EEG statistics and being subject agnostic, since the same algorithm is applied to every subject. It requires more computational resources than the offline methods during inference/filtering phase (since it requires the extra steps of updating both whitening and unmixing matrices), but more performative implementations and dedicated processing devices could be used. In our experiments we show that feature extraction and classification modules performances were equivalent to the offline methods, being the top 3 best average kappa.

Future research paths include an ablation study on how the ORICA hyperparameters (number of assumed subgaussian sources and block size) affect the source separation and the MI-based BCI performance, use of the adaptive forgetting factor proposed in [10], use of ORICA in MI tasks with a higher number of classes and electrodes, and analysis if ORICA could applied to different BCI paradigms.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Wolpaw, J., Birbaumer, N., McFarland, D., Pfurtscheller, G., Vaughan, T.: Brain-computer interfaces for communication and control. Clin. Neurophysiol. **113**, 767–791 (2002)
2. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. Proc. IEEE **89**, 1123–1134 (2001)
3. Mulder, T.: Motor imagery and action observation: cognitive tools for rehabilitation. J. Neural Trans. **114**, 1265–1278 (2007). https://doi.org/10.1007/s00702-007-0763-z

4. Singh, A., Hussain, A., Lal, S., Guesgen, H.: A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface. Sensors **21**, 2173 (2021)

5. Pfurtscheller, G., Neuper, C., Flotzinger, D., Pregenzer, M.: EEG-based discrimination between imagination of right and left hand movement. Electroencephalogr. Clin. Neurophysiol. **103**, 642–651 (1997)

6. Hyvarinen, A, Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)

7. Chen, W., Du, C., Zhang, Y., Wu, X.: Combine ICA and ensemble learning methods for motor imagery EEG classification. In: Meng, H., Lei, T., Li, M., Li, K., Xiong, N., Wang, L. (eds.) ICNC-FSKD 2020. LNDECT, vol. 88, pp. 1376–1384. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-70665-4_150

8. Varsehi, H., Firoozabadi, S.: An EEG channel selection method for motor imagery based brain-computer interface and neurofeedback using Granger causality. Neural Networks **133**, 193–206 (2021)

9. Ruan, J., Wu, X., Zhou, B., Guo, X., Lv, Z.: An automatic channel selection approach for ICA-based motor imagery brain computer interface. J. Med. Syst. **42**(12), 1–13 (2018). https://doi.org/10.1007/s10916-018-1106-3

10. Hsu, S., Pion-Tonachini, L., Jung, T-P., Cauwenberghs, G.: Tracking non-stationary EEG sources using adaptive online recursive independent component analysis. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE (2015)

11. Akhtar, M., Jung, T-P., Makeig, S., Cauwenberghs, G.: Recursive independent component analysis for online blind source separation. In: 2012 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2813–2816 (2012)

12. McHugh, M.: Interrater reliability: the kappa statistic. Biochemia Medica. **22**(3), 276–282 (2012)

13. Brunner, C., Leeb, R., Muller-Putz, G., Schlogl, A., Pfurtscheller, G.: BCI Competition IV dataset IIa at https://www.bbci.de/competition/iv/desc_2a.pdf (2004)

14. Bell, A., Sejnowski, J.: An information-maximization approach to blind separation and blind deconvolution. Neural Comput. **7**, 1129–1159 (1995)

15. Lee, T.-W., Girolami, M., Sejnowski, J.: Independent component analysis using an extended infomax algorithm for mixed Subgaussian and Supergaussian Sources. Neural Comput. **11**, 417–441 (1999)

16. Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., Moulines, E.: A blind source separation technique using second-order statistics. IEEE Trans. Sig. Process. **45**, 434–444 (1997)

17. Cardoso, J.-F., Souloumiac, A.: Blind beamforming for non-gaussian signals. In: IEE Proceedings F Radar and Signal Processing, vol. 140, p. 362 (1993)

18. Ablin, P., Cardoso, J.-F., Gramfort, A.: Faster independent component analysis by preconditioning with hessian approximations. IEEE Trans. Sig. Process. **66**, 4040–4049 (2018)

19. Hsu, S., Mullen, T., Jung, T., Cauwenberghs, G.: Online recursive independent component analysis for real-time source separation of high-density EEG. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE (2014)

20. Zhu, X.: Natural gradient-based recursive least-squares algorithm for adaptive blind source separation. Sci. China Ser. F. **47**, 55 (2004). https://doi.org/10.1360/02yf0242

21. Lin, X., Wang, L., Ohtsuki, T.: Online recursive ICA algorithm used for motor imagery EEG signal. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE (2020)
22. Kohavi, R., John, H.: Wrappers for feature subset selection. Artif. Intell. **97**, 273–324 (1997)
23. Theodoridis, S, Koutroumbas, K.: Pattern Recognition. Academic Press 3ed, San Diego, CA (2006)

# Evaluation of Explainable AI Methods in CNN Classifiers of COVID-19 CT Images

Jean P. O. Lima[1]([✉]) [iD], Roberto d'Amore[2] [iD], Marcos R. O. A. Máximo[1] [iD],
Marcus H. Victor Jr.[1] [iD], and Mônica M. S. Matsumoto[1] [iD]

[1] Medical Devices Lab (LabMed), Electronic Engineering Division, Aeronautics Institute of Technology – ITA, São José dos Campos – SP, Brazil
jean.lima@ga.ita.br

[2] Electronic Engineering Division, Aeronautics Institute of Technology – ITA, São José dos Campos – SP, Brazil

**Abstract.** In computer-aided diagnosis, many solutions based on Deep Learning were developed, but few were deployed in real environments due to the lack of transparency from intelligent models to humans. Thereafter, Explainable AI (XAI) techniques were developed to evaluate image regions with prominent influence in the decision model. In this paper, we present CAM and Grad-CAM, two XAI techniques for image classification. Particularly, we have evaluated them for interpretation of COVID-19 classification of CT images using Convolutional Neural Networks. For the classification task, we have built models with MobileNetV3, VGG-16, VGG-19, and ResNet50 using transfer learning. CT images from the Large COVID-19 CT Scan Slice Dataset were used in training and test sets. This dataset is composed of more than 17,000 CT slices labeled into three classes: covid, pneumonia and normal. This work contributed by proposing a quantitative evaluation based on Jaccard Coefficient and a proposed metric of Coverage Ratio. The metrics aim to compare the XAI regions of interest to the gold standard annotated regions by specialists. We have used another database, the COVID-19 CT Scans Dataset, for XAI evaluation. This dataset contains the regions of potential COVID-19 infection, annotated by a specialist committee. For the classification task, the MobileNetV3 model had accuracy of 97.94% and F1-Score of 98.29%. We have evaluated both XAI techniques, which yielded up to 0.7164 and 0.3105 of Coverage Ratio and Jaccard Coefficient, respectively, for CAM (MobileNetV3 classifier). And up to 0.6847 and 0.2923 of Coverage Ratio and Jaccard Coefficient, respectively, for Grad-CAM (ResNet50 classifier).

**Keywords:** Explainable AI · COVID-19 · Convolutional Neural Networks

## 1 Introduction

Coronavirus Disease 2019 (COVID-19) is a severe acute respiratory syndrome caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), whose symptoms range from respiratory distress to multiple organ failure. Due to the high transmission

rates, there was an urgent need to create new techniques and tools for prediction, detection, and accurate and early diagnosis [1]. Computational solutions to support medical diagnosis are very common, especially in the image processing field. Also according to [1], Computed Tomography (CT) images analysis has shown relevant features regarding the COVID-19 contamination detection task. It indicates a potential for automatic images classification systems, in which Deep Learning techniques, especially Convolutional Neural Networks (CNN), represent the state-of-the-art.

The main difficulty in this scenario, which delays the deployment of intelligent systems in real environments, is the lack of transparency to the specialist regarding the information processed by the computational model to reach a conclusion. Most works that approach classification of CT images to support the COVID-19 diagnosis, as observed in [2], applies the classifier model as a black box, i.e., no information other than the predicted class is provided. This represents a serious problem regarding the system's credibility.

According to [3], a computational model, to support medical diagnosis, must be transparent, understandable and explainable to gain the confidence of specialists. Explainable AI (XAI) has a fundamental role in ensuring safety, ethics and reliability in the use of Artificial Intelligence based solutions.

According to [3], Class Activation Mapping (CAM) and Gradient-weighted Class Activation Mapping (Grad-CAM) are two of the most common techniques in XAI applications of medical imaging. These techniques analyze the behavior of CNNs through feature maps extracted in the convolutional layers from a given input image. Thus, the objective of these techniques is to identify which feature maps have the greatest influence on the network classification.

The main medical imaging modalities used for lung analysis are: X-Ray Projection Radiography and X-Ray Computed Tomography (CT). CT provides a more complete representation, as it extracts several axial images (slices) from the lung. [4, 5] presents the use of CAM and Grad-CAM in the identification of nodules in lungs and to support the diagnosis of COVID-19 from CT images, respectively.

This work aims to analyze and evaluate the behavior of CT images classifier models and provide heatmaps for the original images using CAM and Grad-CAM. These heatmaps represent the main regions analyzed by the model during the inference. A similar work is presented in [1], which implements a classifier model based on the canonical CNN architecture InceptionV3 and applies Grad-CAM to generate the heatmaps.

This paper presents canonical architectures of CNNs using Transfer Learning for a supervised learning with the CT image classification task. This work contributes by proposing a quantitative performance evaluation for CAM and Grad-CAM. This proposed evaluation method is based on the comparison between the obtained results from XAI techniques and the gold standard annotated by specialists using Jaccard Coefficient and a proposed metric called Coverage Ratio. To the best of our knowledge, there is no other work that presents a quantitative evaluation for this task.

The remaining of this work is organized as follows. Section 2 presents the related works. Section 3 describes the materials and the proposed methods. Section 4 presents the results and discussions. Finally, Sect. 5 concludes the work and presents ideas for future work.

## 2   Related Works

The use of Artificial Intelligence techniques in medicine has achieved promising results, especially in computer vision tasks, in which Deep Learning approaches have predominated, as pointed out by [6]. Most studies in the field of computing applied to medicine involve medical data analysis in attempt to identify patterns and, finally, provide a pre-diagnosis that has an adequate level of certainty, as occurs in [7–9] for different medical imaging modalities. In [3], the authors comment that the main reason for the lack of tools in production environments is the lack of interpretability of the Artificial Intelligence based models. The main studies involving interpretability and classification of medical images identified in [3] present models based on CNNs for the most varied tasks and medical imaging modalities.

For the identification of lung diseases, it is no different. Reference [5], for example, presents the development of a 3D-CNN for nodules detection on complete pulmonary CT images. The interesting thing about this approach is that each CT, i.e., each set of slices, is treated as a 3D image, and the CNN extracts volumetric characteristics from the image, which enhances the model in the search for the identification of nodules. Three canonical models of CNN are presented and the results indicate an improvement of about 10% compared to the same model in its 2D version. Grad-CAM, besides being useful in model validation by verifying if the model considers coherent image volumes, is also used to detect the region where the nodule is located.

In [4], the development of a Deep Learning model for detection of COVID-19 cases from chest CT images is presented. The main contribution of their work is in the elaboration of a network architecture with high architectural diversity and selective connectivity, using a projection-expansion-projection pattern. This makes the network reach the representation capacity with less effort, without increasing the need for computational power. The result is a model that is more efficient and at the same time competitive in performance with state-of-the-art models. This result proved to be consistent, achieving similar results to ResNet50.

Reference [1] presents the development of a classifier based on CNN for classifying CT lung images to verify the existence of contamination by COVID-19. The main contribution of [1] is the creation of a database, with examples from 13 medical centers spread across the planet, containing complete CTs of 3,529 different patients who tested positive for COVID-19. The authors do not go into details regarding the implemented neural networks, but the results seem to be consistent and competitive for the slices classification task, which is due to the fact that the database is larger than those found in other works. Unfortunately, this database has not yet been made publicly available. An interesting point is that the Grad-CAM technique, in addition to being used for a subjective evaluation, is also used to evaluate the evolution of patients cases. Thus, CT images were periodically collected from patients, and the infected areas were compared using the heatmaps generated by Grad-CAM to evaluate if there was an improvement or aggravation of the case. Heatmap is a data visualization technique that allows representing the importance of regions of an image through color highlighting: regions of greatest interest are highlighted by warmer colors and those of lesser interest by colder colors.

The main disadvantage identified in the cited references is in the subjective form of evaluation of the XAI techniques. This makes comparison between studies difficult. After analyzing related works, it is possible that the best performance for the task of classifying CT images for detection of COVID-19 must come from an intelligent model that analyzes the image three-dimensionally, as with the task of identifying pulmonary nodules. However, the slice classification approach (with two-dimensional images) was chosen for this work due to the lack of public databases with enough examples to adequately conduct the training of a 3D-CNN.

## 3   Materials and Methods

### 3.1   CAM and Grad-CAM

The XAI techniques used in the development of this work were CAM and Grad-CAM. The CAM performs a weighted arithmetic mean of the feature maps of each core of the last convolutional layer with the weights that activated the output class [10]. The result can be represented by a heatmap. The warmest regions indicate the main regions of the image that influenced the classification of the network. Grad-CAM [11] seeks to generate a heatmap by calculating the Rectified Linear Unit (ReLU) of the gradients obtained in the last convolutional layer. Thus, all gradients that do not contribute to the predicted class are set to zero [11]. As well as in CAM, considering that convolutional layers preserve the spatial information of the original image, these regions obtained in the last layer can be extrapolated to the input layer and converted into a heatmap.

### 3.2   Experimental Data

Two CT image datasets were used to conduct the proposed experiments. The first one is the Large COVID-19 CT Scan Slice Dataset [12]. This dataset consists of a collection of 17,104 slices from 1,130 patients, divided into three classes: "COVID" with 7,593 samples; "NORMAL" with 6,893; and "PNEUMONIA" with 2,618. Each slice has dimension 224x224, where each voxel is quantized into 8 bits. These data were collected from 13 European and Asian countries. 59% of the images are from male patients; 32% from female patients; and 9% of the data did not contain gender information. The mean age is 51.2, 52.8, and 64.3 for COVID-19, Normal, and Pneumonia cases, respectively.

This dataset is appropriate for the classification task, as it already presents the necessary labels for supervised learning. Thus, this dataset was used to create the classifier model. To ensure the proper conduction of the training of the classifier models, the Large COVID-19 CT Scan Slice Dataset was further artificially augmented following Data Augmentation techniques [13]. The Augmentation operations were: horizontal flip; 20% zoom; and 30° rotation.

For the evaluation of CAM and Grad-CAM techniques, the COVID-19 CT scans dataset [14, 15] was used. It consists of 20 complete CTs from patients with COVID-19. Each example contains 301 slices of dimension 512x512, where each voxel is quantized into 12 bits. The main advantage of using this database to evaluate the XAI techniques is that all slices of all the examples went through a committee of specialists who performed

a segmentation on the images, identifying the lungs and the potential regions of infection. These regions are called Gold Standard Regions of Interest ($GS_{ROI}$).

Thus, in addition to a subjective evaluation of the XAI techniques performance, it is possible to perform a quantitative analysis, by comparing the warm area of the heatmap and the $GS_{ROI}$. For this analysis, the Jaccard Coefficient and Coverage Ratio metrics were used, as described in Subsect. 3.3.

### 3.3  Solution Pipeline Description

The machine learning approach consists of a supervised learning with multiclass classification task into the classes "COVID", "PNEUMONIA", and "NORMAL". For this, the Large COVID-19 CT scan slice dataset was used. First, the training and test sets were defined through a 80/20 holdout partition in the database, in which 80% of the data was used as a training set and 20% as a test set. This partition was applied to the number of patients, not the number of slices. In this way, it can be guaranteed that images from a same patient were not used in the training and test phases simultaneously. According to this partition policy, the training set was composed of 13,550 slices and the test set by 3,554. Then, the Data Augmentation technique was applied to generate more training data examples, as explained in Subsect. 3.2. In addition, 20% of training set was reserved as validation set, which was used at the end of each epoch to fine-tune the model hyperparameters and thus avoid bias. Therefore, for each trained model, the training set was presented along with the labels. So that the model parameters could be fitted, in the attempt to map the input data to the output data. At the end of training, test data was submitted to it, and their respective predicted outputs were compared with the desired outputs. Thus making it possible to evaluate the model performance. The performance metrics used were: Accuracy; Precision; Recall; and F1-Score.

The implemented classifier models were based on the canonical CNNs MobileNetV3, VGG-16, VGG-19, and ResNet50 using Transfer Learning [16], i.e., the convolutional layers, responsible for extracting features from the images, were maintained with the weights obtained from ImageNet training, and only the last convolutional layers and dense layers, directly responsible for the classification, had their weights fitted. The hidden dense layer blocks were unified for all models and were composed of 4 dense layers with ReLU activation function and 1024, 512, 128 and 64 neurons, respectively, interleaved by dropout layers with 20% rate. The output layer activation function was Softmax. Models were trained for 100 epochs, with batch size 8 and ADAM optimizer. Python programming language was used to implement these simulations using Keras framework with TensorFlow backend. Other implementation details and full code are available at [17]. The development environment specifications are: i7-8750H CPU; 16GB RAM; and NVIDIA GTX 1050 ti GPU.

Figure 1 illustrates the Explainable AI phase. In this phase, the COVID-19 CT Scans dataset were used. This dataset contains, in addition to the CT images and their labels, the $GS_{ROI}$. An evaluation of the classifier was done in a similar way to the test phase. The images were submitted to the classifier and the respective obtained outputs were compared to the labels by calculating the accuracy. In this way, it was possible to compare whether the results remain robust.
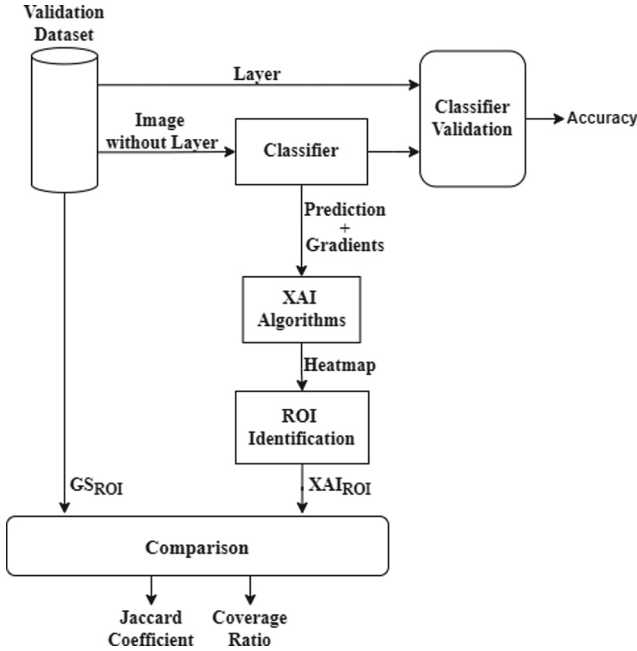
**Fig. 1.** Block diagram of the model and XAI techniques evaluation.

For each image, the classifier prediction and the gradients generated by the last con-volutional layer were submitted to the Explainable AI CAM and Grad-CAM algorithms. Then, CAM and Grad-CAM generated the heatmaps. To compare the obtained results by the XAI algorithms, a XAI Region of Interest ($XAI_{ROI}$) was calculated from each heatmap. For that, the heatmaps were normalized to 0 to 1 range. Then, a binarization with threshold 0.8 was applied. The resultant region is called $XAI_{ROI}$.

Once all $XAI_{ROI}$ were obtained, they were compared to the $GS_{ROI}$ of the images and evaluated with the metrics: Jaccard Coefficient (JC) and Coverage Ratio (CR). JC is a very common metric for object detection and image segmentation tasks. It relates the intersection and union of obtained regions ($XAI_{ROI}$, in this case) and target regions ($GS_{ROI}$, in this case) according to the equation

$$JC(GS_{ROI}, XAI_{ROI}) = \frac{GS_{ROI} \cap XAI_{ROI}}{GS_{ROI} \cup XAI_{ROI}}. \tag{1}$$

As it is not a segmentation task, but an analysis of the region that influenced a classification, it is been proposed a metric that calculates the coverage ratio of the $GS_{ROI}$ by the $XAI_{ROI}$, i.e., the proportion of the target region found by the obtained region. This metric was called Coverage Ratio and is described by the equation

$$CR(GS_{ROI}, XAI_{ROI}) = \frac{GS_{ROI} \cap XAI_{ROI}}{GS_{ROI}}. \tag{2}$$

## 4   Results and Discussions

### 4.1   Classifier Performance

After the training set was presented to the model, the test set was presented to the model and the obtained outputs were compared to the desired outputs and the performance metrics were calculated in order to verify the neural weights fit. The results of this analysis are presented in Table 1.

**Table 1.** Test set classification results.

| Classifier | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| MobileNetV3 | 0.9794 | 0.9832 | 0.9827 | 0.9829 |
| VGG-16 | 0.8729 | 0.8875 | 0.8747 | 0.8810 |
| VGG-19 | 0.9322 | 0.9621 | 0.9188 | 0.9400 |
| ResNet50 | 0.9660 | 0.9732 | 0.9735 | 0.9733 |

The neural network based on MobileNetV3 obtained an accuracy of 0.9794 and more than 0.98 for precision, recall and F1-Score. The F1-Score is the most robust metric among those chosen, as it is the harmonic mean between recall and precision. The MobileNetV3 F1-Score value indicates that the generalization power of the models remained consistent even for data from outside the training set. It characterizes the training as well conducted. In addition, the accuracy values proved to be similar to state-of-the-art works presented in Sect. 2. [1] achieved 0.80 of accuracy and [4] presented results between 0.983 and 0.991 of accuracy. The comparison was made for works with different datasets. [1] used a private dataset and [4] presented a similar dataset to the one used in this work. Therefore, this dataset was considered appropriate to this task.

VGG-16 has the simplest architecture among the models, which justifies its inferior result. The other models present similar results, including VGG-19, which presents an architectural evolution in relation to the VGG-16. It means that there is a relationship between the model architecture and the model performance. Finally, it is concluded that the MobileNetV3 convolutional block was the best feature extractor among the tested models for this task, as it achieved the best results regarding all metrics.

### 4.2   XAI Techniques Performance

For XAI techniques performance analysis, Jaccard Coefficient and Coverage Ratio metrics were used. Table 2 presents the results achieved by CAM and Grad-CAM techniques for each classifier. In addition, the classification accuracy for the evaluation dataset was also calculated, in order to verify the classifiers' robustness. The obtained accuracy was 0.991 for MobileNetV3; 0.963 for VGG-19; 0.957 for ResNet50; and 0.915 for VGG-16. These results indicate that the classifiers remained consistent even for examples from outside the training and test sets.

**Table 2.** Results regarding the evaluation metrics obtained by comparing the Regions of Interest obtained by the heatmaps and segmented regions of the image.

| Classifier | XAI Technique | Jaccard Coefficient | Coverage Ratio |
|---|---|---|---|
| MobileNetV3 | CAM | 0.3105 | 0.7164 |
| | Grad-CAM | 0.2852 | 0.6697 |
| VGG-16 | CAM | 0.2731 | 0.5845 |
| | Grad-CAM | 0.2880 | 0.6092 |
| VGG-19 | CAM | 0.2909 | 0.6679 |
| | Grad-CAM | 0.2813 | 0.6517 |
| ResNet50 | CAM | 0.2994 | 0.6932 |
| | Grad-CAM | 0.2923 | 0.6847 |

Individually, it is clear that there is not a considerable difference between the CAM and Grad-CAM techniques, as they presented similar results for each model. The Jaccard Coefficient presented results between 0.27 and 0.32. This is an expected result, since the Jaccard Coefficient relates the region of intersection between the $GS_{ROI}$ and the $XAI_{ROI}$ with their respective union. Thus, the Jaccard coefficient is sensitive to some $XAI_{ROI}$ construction factors, such as the image resolution. As XAI techniques analyze the gradients of the last convolutional layer to generate the heatmap, this information is generated based on a matrix considerably smaller than the original. In this case, the images were submitted with a dimension of $224 \times 224$, and the analyzed gradients had dimension 8x8. It means that the heatmap is generated for an 8x8 image, and then a dimensional increase is performed and, as the convolutions conserve spatial information, the heatmap holds information referring to the original images. This way, edge details are lost, which makes the difference between union and intersection of the $GS_{ROI}$ and the $XAI_{ROI}$ increase, causing the Jaccard Coefficient of the models to have low results. On the other hand, the precision Coverage Ratio is concerned with verifying whether the $GS_{ROI}$ belongs to the $XAI_{ROI}$. By using this metric it is possible to perceive that the models are strongly influenced by the regions where there are indications of infection in the lung to perform the classification, which adds greater credibility to the classifiers.

Figure 2 illustrates the heatmap generation process. From the heatmap generated for the gradients (Fig. 2A) an interpolation with smoothing is applied to obtain the heatmap for the original image (Fig. 2D). The $XAI_{ROI}$ is generated by binarizing the normalized Fig. 2D with threshold 0.8. From Fig. 2 it is possible to see that the low dimension of the original heatmap directly influences the definition of $XAI_{ROI}$. Even containing a good part of the $GS_{ROI}$ (Fig. 2C), $XAI_{ROI}$ has many false positives, i.e., pixel that belong to $XAI_{ROI}$, but do not belong to the $GS_{ROI}$. It explains the low values for the Jaccard Coefficient and high values for the Coverage Ratio.
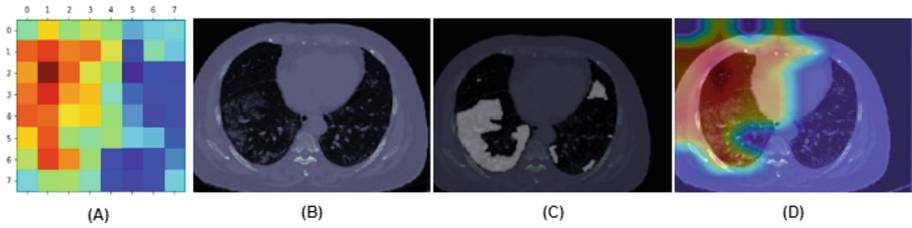
**Fig. 2.** Heatmap generation process. (A) Heatmap generated for 8X8 gradients; (B) Original image of patient with COVID-19. (C) Infection mask to be used in the comparison; and (D) CAM-delimited ROI applied to a MobileNet-based model, which will be compared with (C).

Figure 3 presents some heatmaps of correctly classified samples, in which it is noticed that there is a behavior pattern of the heatmap for each class. Figure 3A shows the regions of greatest intensity of the heatmap in the periphery of the lung along with the region that appears to have an infection. This is an expected result, as it is common for COVID pneumonia to manifest mainly in the peripheral and subpleural regions [18]. Non-COVID pneumonia cases (illustrated by Fig. 3B) present large heatmaps that, in most cases, cover a large area of the lung. It can be explained by the manifestation of Non-COVID pneumonia occupy a large region of the lung in the presented data. For normal cases (Fig. 3C) the regions of greatest intensity appear outside the lung region. It indicates that the network considers healthy lung cases in which no characteristic in the lung area makes a strong contribution to the classification.

Thus, it can be concluded that the networks obtained a good generalization, mainly for the classification task, however, they were not able to determine well-defined patterns in the image as expected, i.e., the $XAI_{ROI}$ is much larger than the $GS_{ROI}$, although in the most of the times $XAI_{ROI}$ contain it. This behavior is evidenced by the high values of the Coverage Ratio. A considerable increase in the training database should improve the feature extraction by the network and, therefore, improve the results obtained for the evaluation of XAI techniques.
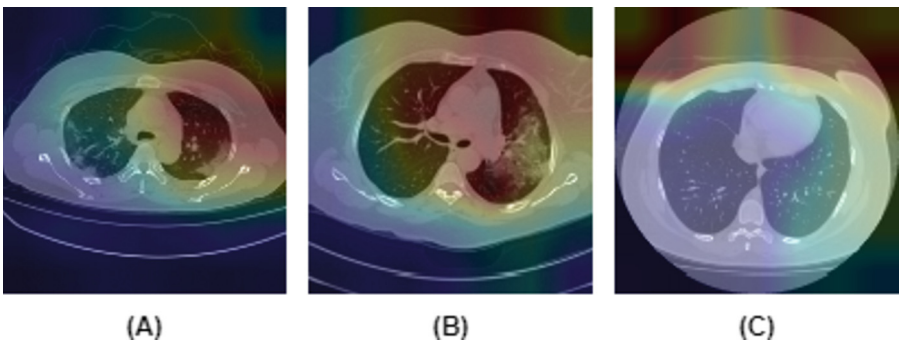


**Fig. 3.** Heatmap generated by Grad-CAM technique based on MobileNetV3 model. (A) COVID sample; (B) Pneumonia sample; and (C) Normal sample.

## 5    Conclusion

The work presented competitive results to the state-of-the-art regarding the classification task. The canonical networks proved to be powerful and capable to classify CT images from features extracted by the convolutional layers with static weights resulting from the ImageNet training. The use of the MobileNet architecture also facilitates the use of the model in embedded systems such as medical devices, which is an advantage in the context of automatic classification of medical images. On the other hand, the XAI task still has a great potential for improvement. Despite a high Coverage Ratio, it presents many false positives, which can be interpreted by the low Jaccard coefficient achieved.

For future work, some tests should be carried out in order to improve the results for the XAI task, such as the use of images with higher dimensionality; more robust architectures; training of all convolutional layers; expansion of the database; etc. These solution attempts demand greater computational power to be tested, but they must reach evolution in the presented results. In addition, the contribution of the work regarding the evaluation methods of XAI tasks allows that future works can be quantitatively compared, and, thus, document more faithfully the advances in this research field.

## References

1. Lee, E.H., Zheng, J., Colak, E., et al.: Deep COVID DeteCT: an international experience on COVID-19 lung detection and prognosis using chest CT. NPJ Digit. Med. **4**, 11 (2021). https://doi.org/10.1038/s41746-020-00369-1

2. Roberts, M., Driggs, D., Thorpe, M., et al.: Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat. Mach. Intell. **3**, 199–217 (2021). https://doi.org/10.1038/s42256-021-00307-0

3. Singh, A., Sengupta, S., Lakshminarayanan, V.: Explainable deep learning models in medical image analysis. J. Imaging **6** (2020). https://doi.org/10.48550/arXiv.2005.13799

4. Gunraj, H., Wang, L., Wong, A.: COVIDNet-CT: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images. Front. Med. **7** (2020). https://doi.org/10.3389/fmed.2020.608525

5. Joshua, E., Bhattacharyya, D., Chakkravarthy, M., Byun, Y.: 3D CNN with visual insights for early detection of lung cancer using gradient-weighted class activation. J. Healthcare Eng. 1–11 (2021). https://doi.org/10.1155/2021/6695518

6. Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. Med. Image Anal. **71**, 102062 (2021). https://doi.org/10.48550/arXiv.1910.02923

7. Pereira, S., Meier, R., Alves, V., Reyes, M., Silva, C.A.: Automatic brain tumor grading from MRI data using convolutional neural networks and quality assessment. In: Stoyanov, D., et al. (eds.) MLCN/DLF/IMIMIC -2018. LNCS, vol. 11038, pp. 106–114. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02628-8_12

8.  Young, K., Booth, G., Simpson, B., Dutton, R., Shrapnel, S.: Deep neural network or dermatologist? In: Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision. Support: 48–55. Springer, Cham 2019. https://doi.org/10.48550/arXiv.1908.06612

9.  Van Molle, P., De Strooper, M., Verbelen, T., Vankeirsbilck, B., Simoens, P., Dhoedt, B.: Visualizing convolutional neural networks to improve decision support for skin lesion classification. In: Stoyanov, D., et al. (eds.) Understanding and Interpreting Machine Learning in Medical Image Computing Applications. MLCN DLF IMIMIC 2018 2018 2018. LNCS, vol. 11038. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02628-8_13

10. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2921–2929. Las Vegas, NV, USA (2016). https://doi.org/10.1109/CVPR.2016.319

11. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618–626. Venice, Italy (2017). https://doi.org/10.1109/ICCV.2017.74

12. Maftouni, M., Law, A.C., Shen, B., Zhou, Y., et al.: A robust ensemble deep learning model for COVID-19 diagnosis based on an integrated CT scan images database. In: Proceedings of the 2021 Industrial and Systems Engineering Conference, pp. 22–25 (2021)

13. Gu, S., Pednekar, M., Slater, R.: Improve image classification using data augmentation and neural networks. SMU Data Sci. Rev. **2**, 1 (2019)

14. Paiva, O.: CORONACASES.ORG - Helping Radiologists to Help People in More Than 100 Countries. Coronavirus Cases. Coronacases.org. (2020)

15. Glick, Y.: Viewing Playlist: COVID-19 Pneumonia. Radiopaedia.Org. (2020)

16. Sarkar, D., Bali, R., Ghosh, T.: Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras. Packt Publishing Ltd (2018)

17. Lima, J.P.O.: XAI Evaluation of Covid Classifiers (2022). https://github.com/jpdol/XAIEvaluationOfCovidClassifiers

18. Liu, M., Zeng, W., Wen, Y., Zheng, Y., Lv, F., Xiao, K.: COVID-19 pneumonia: CT findings of 122 patients and differentiation from influenza pneumonia. Eur. Radiol. **30**, 5463–5469 (2020)

# Heartbeat Classification Based on PCA and CNN

Tatiane C. Ramalho$^{(\boxtimes)}$ , Carlos A. L. Ortiz , Leonardo A. A. Abrantes ,
and Jurandir Nadal

Biomedical Engineering Program, University Federal of Rio de Janeiro,
Rio de Janeiro, Brazil
`tati.tramalho@peb.ufrj.br`

**Abstract.** This work presents a method for automatic heartbeat classification based on principal component analysis and a convolutional neural network on ECG signals. We developed a database holding the first ten principal components and the relative RR intervals of P-QRS complexes from the MIT-BIH Arrhythmia Database patients. The convolutional neural network was used to obtain a model for classifying heartbeats based on this database. This model was tested and compared to other algorithms existing in the literature, and the results evidenced the relative advantages of the method.

**Keywords:** heartbeat classification · Principal Component Analysis · Convolutional Neural Network · signal processing

## 1 Introduction

The electrocardiogram (ECG) allows the representation of the heart's electric activity, presenting the contraction and the relaxation of normal cardiac muscle [1]. The ECG is composed of five wave types that repeat successively and represents the cardiac cycle. These waves are named P, Q, R, S, and T, as shown in Fig. 1. A normal ECG in rest assumes that the patient is lying down, has not made any physical effort in the last ten minutes, or smoked in the previous thirty minutes before the examination. Six electrodes fixed using adhesives to the patient's chest (usually with gel between each electrode and the patient's skin to improve the electrical detection of the signal) are used, as well as four metallic clamps placed on the wrists and ankles [2]. Figure 2 presents the distribution of electrodes and clamps for detecting the ECG signal. Each of these waves has a standard range of amplitude and duration. The detection of these values is subject to several interferences, from the placement and fixation of electrodes to interference from the electrical network.

The ECG can classify heartbeats, providing valuable information for the arrhythmia risk diagnosis or sudden death. Therefore, methods of automatic heartbeats classification are relevant and widely used. Some methods based on artificial intelligence, such as artificial neural networks and deep neural networks,
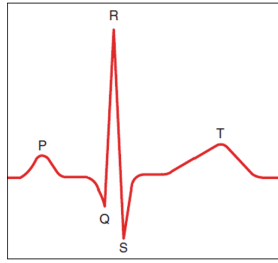
**Fig. 1.** P, Q, R, S and, T waves of an ECG signal. Adapted from [3].
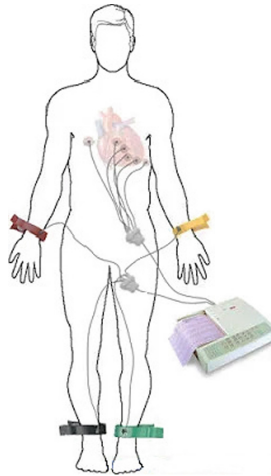


**Fig. 2.** Distribution of electrodes and clamps for detecting the ECG signal. Adapted from [2].

are very efficient for developing automatic heartbeat classifiers. The architecture most commonly employed in deep neural networks for classifying and capturing complex features of an image or signal is the convolutional neural network (CNN) [4]. The CNN does not require the morphological features of the ECG signal to be extracted in advance, which is one of its advantages. In addition, some signal processing techniques, such as principal component analysis (PCA), can reduce the dimensionality of the ECG signal database, facilitating CNN learning in recognition of different heartbeats classes.

This paper proposes a methodology for the automatic heartbeat classification based on the principal components analysis and a neural convolutional network in ECG signals implemented with the python programming language. For this, the MIT-BIH Arrhythmia Database was used, and the beats classification was grouped into the five classes of arrhythmia recommended by the Association for the Advancement of Medical Instrumentation: normal (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown

beat (Q). The obtained results are compared with some algorithms existing in the literature. As a contribution, implementing this methodology will allow for an automatic and accurate diagnosis, making possible the use of automatic monitors with low rates of false alarms.

## 2      Materials and Methods

### 2.1      MIT-BIH Arrhythmia Database

The MIT-BIH Arrhythmia Database of ECG signals from Physionet [5] collected by the arrhythmia laboratory of Beth Israel hospital was used in this study to train and test the classification models obtained by our methodology. This database contains 48 ECG records, with 30 min each, sampled at 360 Hz from 47 different patients obtained in two leads. The first lead is modified-lead II (MLII) contained in the whole records, whereas the second corresponds to V1, V2, V4, or V5, depending on the record. Only the MLII, which represents the bipolar chest lead provided by all the registers, was used.

The database was divided into four stratified datasets to represent the different heartbeats types adequately (Table 1). In this analysis, was not considered the patient 207.

**Table 1.** Stratified datasets and MIT-BIH patients.

| Datasets | MIT-BIH Patients |
| --- | --- |
| Dataset 1 | 100,101,102,109,116,122,124,200,201,220,221,228 |
| Dataset 2 | 105,106,117,118,119,121,203,213,207,217,222,234 |
| Dataset 3 | 104,111,113,115,123,202,208,209,212,215,219,223 |
| Dataset 4 | 103,107,108,112,114,205,210,214,230,231,232,233 |

The MIH-BIH database has small data variability for some classes of heartbeats. Therefore, we chose to use k-fold cross-validation with k = 4. The k-fold cross-validation used three datasets for training and validation and one for testing. Table 2 presents the training/validation and testing datasets. The training datasets were divided into two subsets, where 70% of the data was reserved for training (training dataset) and 30% for validation (validation dataset).

This paper uses the practice recommended by the Association for the Advancement of Medical Instrumentation (AAMI) [6], in which the 15 recommended classes for arrhythmia classifies into five superclasses: normal (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown beat (Q). Table 3 illustrates the 15 classes, their symbols, and the five superclasses used to implement the heartbeats classification.

**Table 2.** Training and testing datasets for k-fold cross-validation.

| Training/Validation | Testing |
|---|---|
| Group 1: Dataset 2, Dataset 3, Dataset 4 | Dataset 1 |
| Group 2: Dataset 1, Dataset 3, Dataset 4 | Dataset 2 |
| Group 3: Dataset 1, Dataset 2, Dataset 4 | Dataset 3 |
| Group 4: Dataset 1, Dataset 2, Dataset 3 | Dataset 4 |

**Table 3.** Principal classes of the MIT-BIH database and AAMI standard classes.

| AAMI class | Symbol | MIT-BIH class |
|---|---|---|
| N | N | Normal beat |
| | L | Left bundle branch block beat |
| | R | Right bundle branch block beat |
| | e | Atrial escape beat |
| | j | Nodal (junctional) escape beat |
| S | A | Atrial premature beat |
| | a | Aberrated atrial premature beat |
| | J | Nodal (junctional) premature beat |
| | S | Supraventricular premature beat |
| V | V | Premature ventricular contraction |
| | E | Ventricular escape beat |
| F | F | Fusion of ventricular and normal beat |
| Q | p or / | Paced beat |
| | f | Fusion of paced and normal beat |
| | U | Unclassifiable beat |

## 2.2  ECG Signal Preprocessing

The ECG signals in the MIT-BIH Arrhythmia Database contain the noise of different frequencies, so the first step in signal preprocessing is filtering the ECG signals. For this, we designed a two-sample moving average filter. In this way, it was possible to attenuate the high-frequency noises of the signals. We used the notes made by two Massachusetts Institute of Technology cardiologists available in the MIT-BIH Arrhythmia Database, which presented the occurrence of each beat (represented by the R wave of the QRS complex) and also selected 108 samples around each heartbeat.

The MIH-BIH Arrhythmia Database also presents heartbeats classes with imbalanced data, which is a problem for the beats classification. The typical classification algorithms assume that the data are uniformly distributed in the class and that classes are balanced. When these conditions change, the algorithm's performance measurement can be skewed. For this reason, we resampled

the data using a random oversampling approach [7]. In random oversampling, the minority class samples are increased by random sampling replication of minority class representatives. Furthermore, the principal component analysis was used to reduce the data dimensionality of the original database signals after filtering and selecting samples from the ECG signals.

### 2.3   Principal Component Analysis

The principal component analysis (PCA) applicated to biological signals is based on interpreting a temporal pattern (sequence of $N$ consecutive samples of the digitized signal) as a single point in an $N$-dimensional space. In the specific case, the temporal patterns are constituted by ECG segments composed of isolated QRS complexes. The digitized signal segments containing $N$ samples corresponding to a heartbeat interpret a vector in the space of dimension $N$, with each sample corresponding to the projection of the vector in a specific axis of this space. This technique can reduce the dimensionality of a dataset containing many highly correlated variables to maximize the representation of the variability of the dataset [8]. PCA represents a dataset through principal components, linear combinations of the $N$ random variables that compose the original dataset.

For a dataset of $N$ random variables $x_i$, the principal components form a new set with $N$ variables that are linear combinations of the initial dataset:

$$y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N$$
$$y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N$$
$$\vdots$$
$$y_N = a_{N1}x_1 + a_{N2}x_2 + \ldots + a_{NN}x_N$$

where $a_{NN}$ is the vector of eigenvalues and $x_N$ is the correlation matrix of $x$, such that the first eigenvector $y_1$, or first principal component (CP), is oriented towards the most significant variance of the original data. The second CP $y_2$ is orthogonal to the first and maximizes the representation of residual variance, and so on. With this new set of variables, it is possible to reduce the dimensionality of the data, retaining $k$ CPs that allow explaining almost all the variance of the data, with $k << n$ [9]. The value of $k$ is obtained through tests, highlighting the Broken Stick Test and the Scree Graph [8]:

$$\frac{\sum_{j=1}^{k} var(y_j)}{Total\ Variance} = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{i=1}^{n} \lambda_i} \tag{1}$$

where $\lambda_j$ corresponds to the eigenvalues of the linear transformation, taken in decreasing order.

The Broken Stick criterion [8], as described in Eq. (1), was employed to indicate the number of variables retained in the analysis.

In this study, we selected the first ten CPs computed from the MIH-BIH Arrhythmia Database, as this CPs quantity could already explain almost all
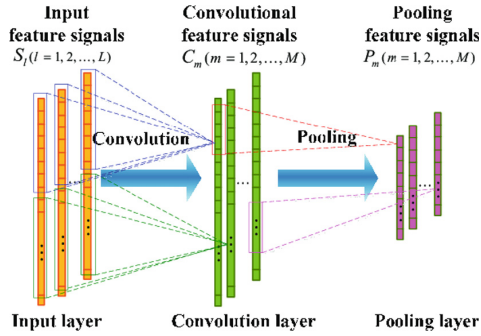
**Fig. 3.** Typical structure of CNN.

the data's variance. Thus, we created a new database containing these CPs and the relative RR intervals of P-QRS complexes. This database was employed as input to train a model for classifying heartbeats based on a convolutional neural network.

### 2.4   Convolutional Neural Network

The convolutional neural network (CNN) is the most popular deep neural network widely used to diagnose the ECG heartbeat. In the standard configuration of CNN, the neurons do not have complete connectivity, except for the last layer, which is fully connected. It makes the CNN have few parameters for learning when compared to fully-connected deep neural networks. Figure 3 shows a typical structure of CNN. Looking at Fig. 3, it is possible to understand how the CNN receives the inputs in vector form, the hidden layer that presents a convolution structure, and the pooling connected to the output fully connected (complete connectivity).

In the CNN's convolutional layer, each neuron in the first cultured layer will be connected to a small region of the input neurons, for example, a $1 \times 5$ region corresponding to 5 input data. The pooling layer receives each feature map output from the convolutional layer and prepares a reduced feature map [10]. A commonly used pooling operation is maximum pooling, which divides the input into pooling regions and then computes each region's maximum. Furthermore, a dropout layer can be inserted right after the convolution layer to help slow down the learning process, as these layers learn very fastly [10].

In this work, the structure used for CNN has one-dimensional and the following characteristics: two convolutional layers with the size of kernels of 3, the number of filters of 128, and ReLU activation (rectified linear activation function), a dropout layer of 0.35, a max-pooling layer with a pooling size of 2, a fully connected layer, and a softmax layer (responsible for the five different types of the heartbeats). The final layer of the CNN operates in the same way as a multilayer perceptron (MLP) neural network. The fully connected and the softmax layers are used to recognize and predict the classes to which each heartbeat belongs.

Furthermore, we used the Adam gradient descent stochastic optimizer for the training based on the adaptive estimation of first and second-order moments, and the learning rate used was 0.001. According to [11], this method is computationally efficient, has little memory consumption, and presents good results for large databases and large numbers of classes. We defined 100 epochs for training and validation, and in each epoch, the batch size used for the dataset was 32. We used the 4-fold cross-validation, with all combinations of three datasets for training and validation and the remaining one for testing, according to Table 2. The training datasets were divided into two subsets, where 70% of the data was reserved for training (training dataset) and 30% for validation (validation dataset). We also specify a stopping criterion for the learning. Thus, the training is stopped when there is no improvement in the specified minimum error. We opted for the number of 8 epochs without improvement to stop the training.

## 3   Results

This section presents the results obtained considering the training and testing datasets studied in this work. Comparisons between the method developed and others methodologies found in the literature that also use the CNN one-dimensional are performed. Comparisons between the method developed and others methods found in the literature with the same purpose are performed. Some measures, such as accuracy, precision, sensitivity, and specificity, are calculated to evaluate the models and are described in the Eqs. (2), (3), (4), and (5), where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

The evolution of the accuracy and loss during the training and validation for each testing group of Table 2 are illustrated in Figs. 4 and 5, respectively. Looking at Figs. 4 and 5, it can be seen that the accuracy increases and the loss decreases at each epoch for all groups used in the cross-validation. Also, Group 1 executes training during all specified epochs (100 epochs), while Groups 2 and 3 stop training at around 50 epochs, and Group 4 stops at about 25 epochs.

Table 4 shows the precision, sensitivity, specificity, and accuracy obtained in the cross-validation for each testing dataset considering AAMI standard classes.
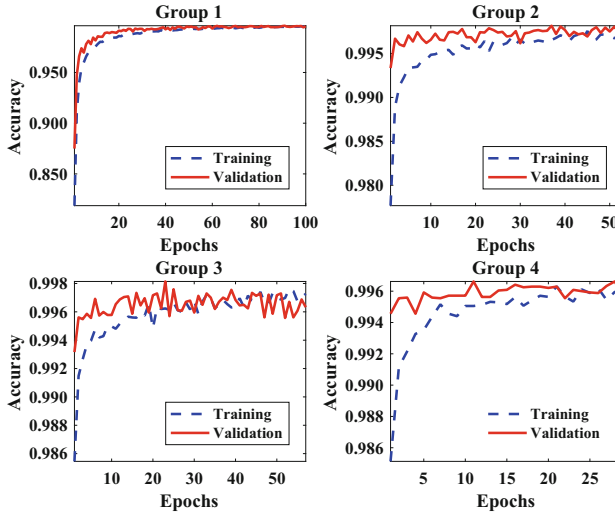
**Fig. 4.** Accuracy evolution during the training and validation for each testing group.
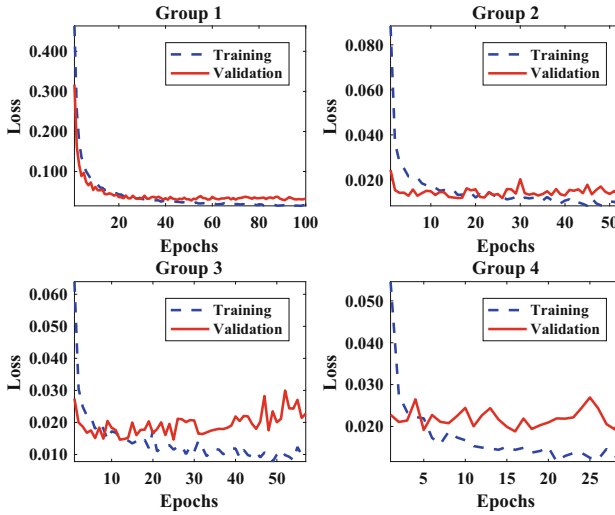


**Fig. 5.** Loss evolution during the training and validation for each testing group.

The same metrics are presented in Table 5 for all testing datasets, i.e., testing all patients of the MIT-BIH Arrhythmia Database.

The confusion matrix shows the classification results of the proposed CNN for the five heartbeats classes, exhibiting the quantity of correct and incorrect predictions (Table 6).

**Table 4.** Precision, sensitivity, specificity, and accuracy for each testing dataset considering AAMI standard classes.

| | | Dataset 1 | |
|---|---|---|---|
| Class | Precision (%) | Sensitivity (%) | Specificity (%) |
| N | 99.94 | 99.51 | 99.99 |
| S | 99.60 | 100.00 | 99.90 |
| V | 99.98 | 100.00 | 100.00 |
| F | 99.96 | 100.00 | 99.99 |
| Q | 99.96 | 99.94 | 99.99 |
| Accuracy (%) | | | 99.89 |
| | | Dataset 2 | |
| Class | Precision (%) | Sensitivity (%) | Specificity (%) |
| N | 99.92 | 99.11 | 99.98 |
| S | 99.92 | 100.00 | 99.98 |
| V | 99.71 | 99.59 | 99.93 |
| F | 99.20 | 100.00 | 99.80 |
| Q | 99.96 | 100.00 | 99.99 |
| Accuracy (%) | | | 99.74 |
| | | Dataset 3 | |
| Class | Precision (%) | Sensitivity (%) | Specificity (%) |
| N | 99.93 | 99.40 | 99.98 |
| S | 99.55 | 100.00 | 99.89 |
| V | 99.98 | 99.77 | 100.00 |
| F | 99.74 | 100.00 | 99.93 |
| Q | 99.96 | 99.98 | 99.99 |
| Accuracy (%) | | | 99.83 |
| | | Dataset 4 | |
| Class | Precision (%) | Sensitivity (%) | Specificity (%) |
| N | 98.07 | 99.32 | 99.51 |
| S | 99.52 | 97.33 | 99.88 |
| V | 98.69 | 99.31 | 99.67 |
| F | 99.81 | 100.00 | 99.95 |
| Q | 99.85 | 99.96 | 99.96 |
| Accuracy (%) | | | 99.18 |

**Table 5.** Precision, sensitivity, specificity, and accuracy for all patients of the MIT-BIH Arrhythmia Database considering AAMI standard classes.

| Class | Precision (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|
| N | 99.46 | 99.34 | 99.87 |
| S | 99.65 | 99.34 | 99.91 |
| V | 99.60 | 99.67 | 99.90 |
| F | 99.68 | 100.00 | 99.92 |
| Q | 99.93 | 99.97 | 99.98 |

| Precision (%) | Sensitivity (%) | Accuracy (%) |
|---|---|---|
| 99.87 | 99.79 | 99.66 |

**Table 6.** Confusion matrix of the developed model.

| Class | N | S | V | F | Q |
|---|---|---|---|---|---|
| N | 21361 | 54 | 36 | 38 | 14 |
| S | 91 | 21361 | 51 | 0 | 0 |
| V | 18 | 22 | 21432 | 31 | 0 |
| F | 0 | 0 | 0 | 21503 | 0 |
| Q | 6 | 0 | 0 | 0 | 21497 |

Output class / Target class

**Table 7.** Comparison of the proposed method and other algorithms published in the literature.

| Paper | Class | Preprocessing | Feature Extraction | Accuracy (%) |
|---|---|---|---|---|
| Li et al. [12] | N, L, R, A, V | Wavelet combination | 1D-CNN | 97.50 |
| Pandey et al. [13] | N, S, V, F, Q | Feature scaling | 1D-CNN | 98.23 |
| Wu et al. [14] | N, L, R, A, V | Wavelet transform | 1D-CNN + average pooling | 97.41 |
| Mahdhaoui et al. [15] | N, L, R, A, V | Moving average filter wavelet transform | 1D-CNN + residual block | 97.80 |
| Xiaolin et al. [16] | N, S, V, F, Q | ECG heartbeat segmentation | 1D-CNN | 98.12 |
| Apu et al. [17] | N, S, V, F, Q | Resample + Gaussian Combination | 1D-CNN | 98.25 |
| **Proposed method** | **N, S, V, F, Q** | **Moving average filter + Resample + PCA** | **1D-CNN** | **99.66** |

Table 7 summarizes the characteristics of the proposed method and other published algorithms in the literature. We should point out that the database separation method used in this article to train, validate and test the implemented heartbeats classifier differs from the methods used in the methodologies in Table 7. The proposed methodology uses the 4-fold cross-validation presented

in Sect. 2, which was not used in any of these articles. This technique is more efficient as it uses all data for testing independently.

## 4   Discussion

The methodology proposed in this article presented satisfactory results. Analyzing these results, we noticed that both in the training/validation and in the test of the proposed CNN network, the datasets reached high values for all metrics considered for all classes studied in this work.

During the training and validation of the network, we observed that Group 1 had slower learning because it used the total number of epochs configured. The other groups (Groups 2, 3, and 4) had a faster learning process, stopping training before the number of epochs specified.

For all datasets used in the cross-validation, precision was higher than 98%. Thus, we can infer that the heartbeats considered false positives found in the model are not more harmful than false negatives in predicting the heartbeats classes. The sensitivity and the specificity also showed high percentage values, above 97% and 99%, respectively. These values mean that the classifier can correctly identify true positive and true negative heartbeats, i.e., the classifier can accurately predict whether or not a beat belongs to a particular class.

Regarding accuracy, the proposed model stands out compared to other classification algorithms in the literature, presenting the best accuracy.

According to Table 5, the global results of the model demonstrate the robustness of the developed automatic heartbeats classifier. Furthermore, it is worth noting that the stratified division of the datasets and the balancing of the database contributed to good network performance.

The proposed method has some limitations. For example, the data must be balanced to make the classification result as satisfactory as possible. Also, there is a need to generalize the proposed approach for use in any database, which is simple if the data has previously undergone an exploratory analysis. Another issue would be the difference in the computational cost. Depending on the equipment on which the proposed algorithm is executed, it can be faster or slower. Furthermore, the CNN network can present slight differences in performance, even defining random seeds at all classification stages.

## 5   Conclusions

This paper presented a method for automatic heartbeat classification based on principal component analysis and a convolutional neural network on ECG signals. We created a database containing the first ten principal components and the relative RR intervals of P-QRS complexes from the MIT-BIH Arrhythmia Database patients, which were used as input to CNN, allowing us to obtain a classification model for this database. The method proposed showed 99.66% overall classification accuracy, 99.87% precision, and 99.79% sensitivity on the heartbeats classification in the AAMI standard classes (N, S, V, F, and Q). The

model's performance proved very satisfactory, demonstrating the advantages of the proposed method.

For future work, we aim to apply a two-dimensional CNN network to classify the heartbeats in a more significant number of classes and develop algorithms to optimize the choice of CNN parameters.

**Conflict of Interest.** The authors declare no conflicts of interest.

# References

1. Guyton, A., Hall, J.: Textbook of Medical Physiology. Elsevier Saunders, Amsterdam (2015)
2. Pinheiro, P.: Eletrocardiograma (ECG): Entenda os resultados. https://www.mdsaude.com/cardiologia/exame-eletrocardiograma-ecg/ (2021)
3. Koeppen, B., Stanton, B.: Berne & Levy Fisiologia. Elsevier, Rio de Janeiro (2009)
4. Asgharzadeh-Bonab, A., Amirani, M., Mehri, A.: Spectral entropy and deep convolutional neural network for ECG beat classification. Biocybernetics Biomed. Eng. **40**(2), 691–700 (2020)
5. Moody, G., Mark, R.: The impact of the MIT-BIH arrhythmia database. In: IEEE, Engineering in Medicine and Biology Magazine, vol. 20, pp. 45–50 (2001)
6. Aami, A.: ANSI/AAMI EC57:2012, testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. American National Standards Institute, Inc (2012)
7. Prati, R., Batista, G., Monard, M.: A study with class imbalance and random sampling for a decision tree learning system. In: IFIP, International Conference on Artificial Intelligence in Theory and Practice, vol. 276, pp. 131–140. Boston, MA (2008)
8. Jolliffe, I.: Principal Component Analysis. Springer-Verlag, New York (2002)
9. Mingoti, S.: Análise de Dados Através de Métodos de Estatística Multivariada - Uma Abordagem Aplicada. Editora UFMG, Belo Horizonte (2005)
10. Brownlee, J.: Deep Learning With Python: Develop Deep Learning Models on Theano and Tensorflow Using keras. Machine Learning Mistery, Melbourne, Australia (2016)
11. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. In: International Conference on Learning Representations (2015)
12. Li, D., Zhang, J., Zhang, Q.: Classification of ECG signals based on 1D convolution neural network. In: 19th International Conference one-health Networking. Applications and Services (healthcom), pp. 1–6. IEEE, Dalian, China (2017)
13. Pandey, S.K., Janghel, R.R., Varma, K.: Classification of ECG heartbeat using deep convolutional neural network. In: Rout, J.K., Rout, M., Das, H. (eds.) Machine Learning for Intelligent Decision Science. AIS, pp. 27–47. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-3689-2_2
14. Wu, M., Lu, Y., Yang, W.: A study on arrhythmia via ECG signal classification using the convolutional neural network. Front. Comput. Neurosci. **14**(1), 10 (2020)

15. Mahdhaoui, H., Hattay, J., Mohamed, H.: 1D convolutional neural network based ECG classification system for cardiovascular disease detection. EasyChair Preprint 6066 (2021)
16. Xiaolin, L., Cardiff, B., John, D.: A 1D convolutional neural network for heartbeat classification from single lead ECG. In: 27th International Conference on Electronics. Circuits and Systems (ICECS), pp. 1–2. Glasgow, UK, IEEE (2020)
17. Apu, M., Akter, F., Lubna, M.: ECG arrhythmia classification using 1D CNN leveraging the resampling technique and gaussian mixture model. In: IEEE, 10th International Conference on Informatics, Electronics & Vision (ICIEV), pp. 1–8. Kitakyushu, Japan (2021). https://doi.org/10.1109/ICIEVicIVPR52578.2021.9564201

# Differential Event-Related Spectral Perturbation for Left and Right Elbow Movement for Applications in a Brain-Computer Interface

André da Silva Pereira[1]([envelope]) [iD], Eric Kauati Saito[1] [iD], Paulo Victor Chagas[3] [iD], Ana Paula Fontana[2] [iD], Mario Fiorani[3] [iD], Juliana Guimarães Martins Soares[3] [iD], and Carlos Julio Tierra-Criollo[1] [iD]

[1] Laboratory of Signal and Images Processing, Biomedical Engineering Program, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
`andresilper@peb.ufrj.br`
[2] Physiotheraphy Department, UFRJ, Rio de Janeiro, Brazil
[3] Laboratory of Cognitive Physiology, Biophysics Institute Carlos Chagas Filho, UFRJ, Rio de Janeiro, Brazil

**Abstract.** In the present study, we seek to investigate how the sequence of neuronal events that include stimulus identification, motor planning, and movement execution is implemented in the cortical neural networks, using an experimental paradigm that segregates between preparation and motor activation. For this study, twenty-two right-handed subjects had EEG signals recorded and analyzed. The subjects were divided into two groups. The imagery group performed real right or left forearm movements or the imagination of these movements. The control group performed real movement of the right or left forearm or did not perform any movement (wait condition). The Event-Related Spectral Perturbation (ERSP) in each one of the experimental conditions was compared for the time intervals of preparation, execution, and after-movement end. We observed: (I) mu and beta desynchronizations (ERD) for movement and imagery and not for the wait condition, during preparation and execution periods. (II) The movement ERD is greater than the imagery ERD for both mu and beta bands. (III) Synchronization (ERS) after the movement is finished. (IV) Greater mu ERD during the execution period and beta ERS after the movement ends at the contralateral channels in the movement condition. (V) Gamma ERS after the movement ends for the movement condition, and not for wait. These findings suggest that using signals from the contralateral region of the movement to be executed might impact the accuracy of the control of devices in a brain-computer interface paradigm.

**Keywords:** Electroencephalography · Brain-Computer Interfaces · Signal Processing · Desynchronization

## 1 Introduction

Brain-computer interfaces (BCIs) allow communication between the human brain and external devices, such as orthoses or wheelchairs [1]. They are being used by an increasing number of people suffering from diseases that impair the execution of movements, to improve both the interaction with the external environment and the effectiveness of the physiotherapy approaches adopted [2].

Motor Imagery (MI) is a widely used physiotherapy strategy for improving motor control and consequently autonomy gain by patients suffering from diseases that impair motor control, such as stroke. This technique consists of the imagery of the performance of movements, without any muscular activation [3]. Neuroimaging studies have demonstrated that MI elicits similar neural activation patterns when compared to motor execution [4].

Some studies suggest that therapies involving MI promote greater autonomy gain in the daily life activities of stroke patients than conventional therapy, which does not involve MI [5]. MI also alters the cortical map, increasing the activity of the premotor area, primary motor cortex, and superior parietal cortex [6]. Furthermore, it was shown that cortical modifications produced by the MI are similar to those produced by physical therapy [7] and the combination of MI and movement execution induces greater improvement in the performance of a sequential motor skill than physical or mental therapies alone [5].

Studies have shown that controlling BCIs based on the MI of the subjects can become a realistic option for improving the life quality of patients with severe motor deficits. The BCI treatment can represent a rehabilitation tool for patients for whom there is no other therapy available. The translation of electric brain activity into control signals of BCIs is important to promote increases in the performance of daily life activities and neuroplasticity augmentation [8]. Electroencephalography (EEG) is a widely used method for obtaining cerebral patterns that must be processed and classified for controlling the devices in a non-invasive approach.

Non-invasive BCIs can be controlled by the Event-Related Spectral Perturbation (ERSP) signal, which reflects changes in Power Spectral Density along the time in a specific frequency band of the EEG signal. These changes may consist of Event-Related Desynchronization (ERD) or Event-Related Synchronization (ERS). The former corresponds to a decrease while the second corresponds to an increase of power in a specific frequency band.

During the performance of a repetitive and voluntary task, the mu band (8–13 Hz) presented an ERD initiating 2 s before the movement onset [9]. ERD in the mu and beta (13–30 Hz) frequency bands was also demonstrated for motor execution and motor imagery [10, 11]. Some studies attribute to the gamma band the function of coupling the temporal and spatial processing occurring in different brain areas, in order to produce a coherent perception [12]. A study found a gamma ERS during the performance and imagination of movements [13].

However, the question of mu and beta desynchronizations and gamma synchronization present contralateral dominance during the preparation and execution of a voluntary movement remains unclear and might differ among different experimental protocols [14].

In the present study, we compare the ERSP obtained from a forearm flexion/extension task for both movement and imagery conditions. The protocol was designed to separate and identify signals related to the preparation and execution of the tasks [15]. It is based on studies that used instructions with a delay period to study the responses of the pre-motor cortex [15, 16]. In these tasks, an instructional cue provides anticipated information about the nature of the movement to be executed after a time interval (delay). The delay allows the subjects to plan the signalized movement, reducing the amount of information processing necessary to execute the task after the cue.

## 2    Methodology

### 2.1    Experimental Paradigm

This study was conducted in the Laboratory of Cognitive Physiology of the Carlos Chagas Filho Biophysics Institute (IBCCF), Federal University of Rio de Janeiro (UFRJ). The experiment was approved by the Ethics Committee of UFRJ (Protocol number 851.521). Every subject filled the Edinburgh Handedness Inventory and gave informed consent for participating in this research.

The experiment begins with the subject sitting comfortably in an armchair, with their forearms resting on the arms of the chair. After placing each electrode in its corresponding position at the cap, the impedance was verified and maintained under 5 kΩ. For this experiment, the subjects were divided into two groups: the motor imagery group, in which the tasks are movement or motor imagery, and the control group in which the tasks are movement or simply waiting for the next cue.

The cues (arrows) were presented on a monitor, at a distance of 110 cm from the subject. The direction of the arrows indicates which member should perform the task (right or left) and the color, black or white, indicates the task to be performed (movement or imagination of the movement for motor imagery group and movement or wait for the control group, respectively). The task consists of the flexion of the forearm followed by the return to the initial position. The subjects were instructed to perform the imagination from a first-person perspective.

A schematic representation of the experimental protocol is shown in Fig. 1. The trial begins with a fixed white cross presented over a gray background. After 2 s, the cross is replaced by one of four possible arrows (Preparation period), which is presented for 2 s on the screen. The black arrows indicate that the subject should perform the movement, while the white arrows indicate imagination of the movement for the motor imagery group, or wait for the next cue, for the control group. After this time period, the fixed white cross takes over and remains on the screen from seconds 2 to 8 s. The subjects were instructed to perform the task as soon as the cross replaced the arrow on the screen.

Each experiment has 240 trials, which accounts for 40 min of EEG recording. Taking together the time needed to set the EEG montage, impedance verification, and recording time, each experiment takes about 2 h.
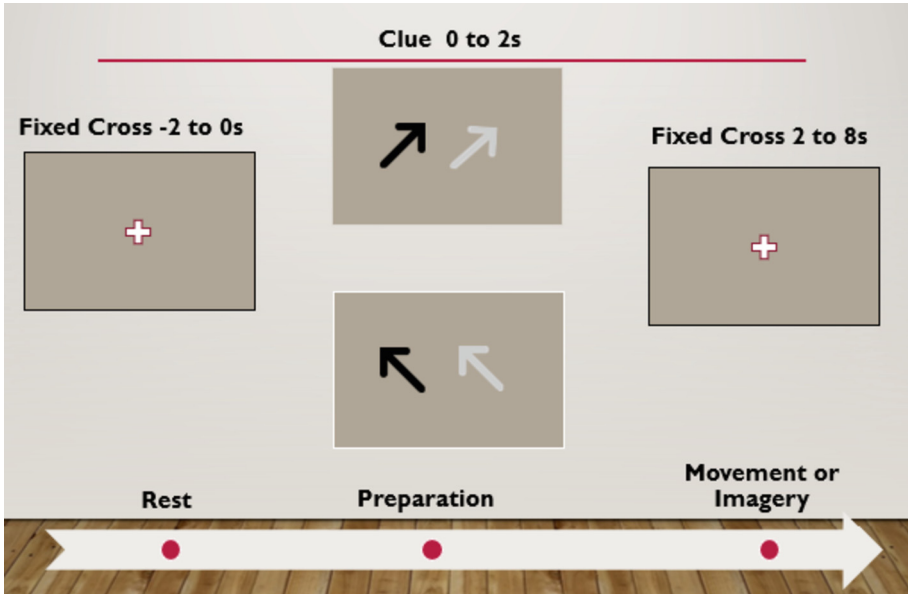
**Fig. 1.** Schematic representation of the experimental design.

## 2.2 Electroencephalographic Recordings

A 32-channel Neurosoft EEG system (NEURON-SPECTRUM-5) was used for recording the cerebral signals of 24 right-handed healthy subjects (14 women, average age = 24 ± 7,6). The sampling rate of this system is 1000 Hz. The channels were analogically filtered with a bandpass filter set to attenuate frequencies out of the interval 0.1 to 100 Hz at the moment of signal acquisition.

The electrodes distribution along the scalp was done with a 10–10 International System EEG cap (Fig. 2 - Fp1, Fpz, Fp2, F3, F4, Fz, F7, F8, FC1, FC2, FC3, FC4, C1, Cz, C2, C3, C4, C5, C6, CP1, Cpz, CP2, CP5, CP6, P3, P4, Pz, T3, T4, T5, T6, Oz, A1, A2). The reference for the system was the average activity registered by electrodes placed at the earlobes of the subjects (A1, A2). Muscular activity was also monitored by two pairs of electromyography electrodes, one for each bicep brachii muscle of the subjects, which is important for analyzing if the tasks are being properly executed.
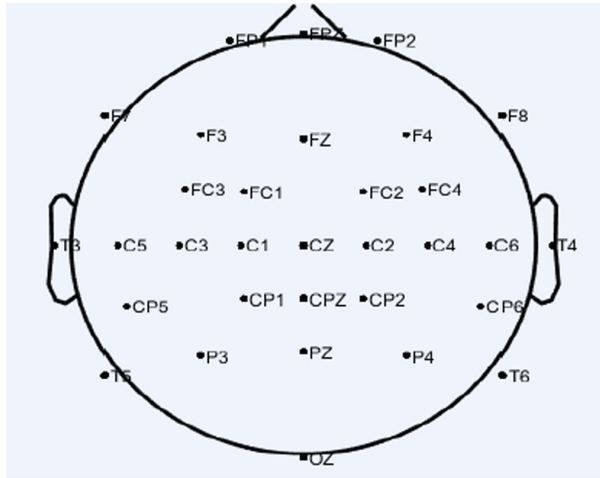
**Fig. 2.** Scalp Channel Locations Source: EEGLAB.

## 2.3   Signal Processing

After the acquisition, the EEG signal is bandpass filtered from 1 to 45 Hz using a second-order, digital, high, and low-pass Butterworth filters. The filtering is applied in the direct and reverse directions in order to result in zero-phase shift. A digital notch filter ($Q =$ 15) set to 60, 120, and 180 Hz is also applied to reduce electrical network interference and harmonics.

Independent Component Analysis (ICA) is applied using the SOBI (Second-Order Blind Identification) algorithm for removing eye blink artifacts from the signal. After ICA decomposition is done, we compare the Pearson correlation between each component and the average of the Fp1 and Fp2 channels. The component which has the largest correlation with this averaged channel is removed.

The signal is then segmented into epochs going from $-2$ to 8 s. Errors of the subjects are removed after a careful visual inspection of the signal. Artifact removal proceeds in a semi-automatic manner, in which a segment containing 20 s of the signal without any artifacts is used for defining two thresholds, a positive and a negative one, consisting of two times the standard deviation of the segment.

The artifact removal step is done based on three different criteria. 1- Each epoch was compared with the threshold and it was rejected if the following was true: having a total of 10% of the samples non-consecutively, or 5% of the samples consecutively outside of a threshold. 2- The epochs which have any sample with amplitude above 150 $\mu$V are excluded. 3- It is also performed a visual analysis of the signal by an expert and the epochs that have visual similarities to artifacts are removed.

After artifact rejection, the signals for each subject are downsampled to 600 Hz. Table 1 shows the total number of epochs after artifact rejection. One of the subjects had to be removed since the signal was largely contaminated with artifacts (over 50% of epochs had to be removed).

**Table 1.** Number of epochs after artifact rejection for each experimental condition.

| Experimental Conditions | Control Group (n = 11) | Imagery Group (n = 12) |
|---|---|---|
| Right Forearm Movement | 493 | 474 |
| Left Forearm Movement | 497 | 472 |
| Right Forearm Imagery | 0 | 441 |
| Left Forearm Imagery | 0 | 451 |
| Wait for the next cue | 946 | 0 |
| Total | 1936 | 1838 |

### 2.4 Event-Related Spectral Perturbation (ERSP)

The ERSP measures non-phased-locked event-related perturbations in the EEG amplitude [17]. The EEG signal was filtered with a zero-phased fourth-order Butterworth bandpass filter with bandpass set according to the frequency bands of interest: mu (8–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz). Each time-sample was squared to obtain the energy, and then we averaged the epochs among all the subjects.

For smoothing the data, the average of the time-samples within a sliding window of 500 ms with 50% overlap was used (PSW). We then performed a baseline normalization, according to the equation [10]:

$$ERSP = \frac{(PSW - PBL)}{PBL} X\,100,$$

where PBL is the baseline period defined as the average of a time period consisting of 1 s before the stimulus onset (from $-2$ to $-1$ s), ERSP represents the percentage of the PSW normalized by the baseline (reference). For further smoothing the data we also applied a moving average third-order filter.

### 2.5 Statistical Analysis

The statistical analysis of ERSP is done using the two-tailed Wilcoxon signed rank test, with alpha level = 0.05. The signal is divided into three periods: Preparation (0 to 2000 ms), Execution (2000 to 4500 ms), and After movement end (6000 to 7500 ms), therefore there are twelve-time windows of interest. For each subject we calculated the average value of ERSP within each time period, using a 500 ms time window, for the channels C3, C4, F3, F4, P3, and P4. The statistical tests were applied successively in each of the time windows. Two different analyses were performed. In the first one, we compared channels of the right and left regions (C3 vs. C4, F3 vs. F4, P3 vs. P4) for the same experimental conditions, in order to establish whether the contralateral cerebral activity produces statistical difference in the ERSP recordings of equivalent channels placed at different scalp hemispheres. For the second analysis, we compared the same channels for different experimental conditions (C3 MI vs. C3 MV and C3 MV vs. C3 Wait, for example) to assess if the brain processes the experimental conditions differentially.

# 3    Results

When comparing movement and wait conditions (Fig. 3), the mu band presents significant differences in both preparation and execution time periods. The mu ERD for the MV condition is significant during all the execution period at the C3 and P3, but not in the C4 and P4 channels. The beta band presented significant ERD during the execution only at the P3 channel, and significant ERS after movement ends at the contralateral side channels. The gamma band also presented a significant ERS after movement end in both hemispheres.
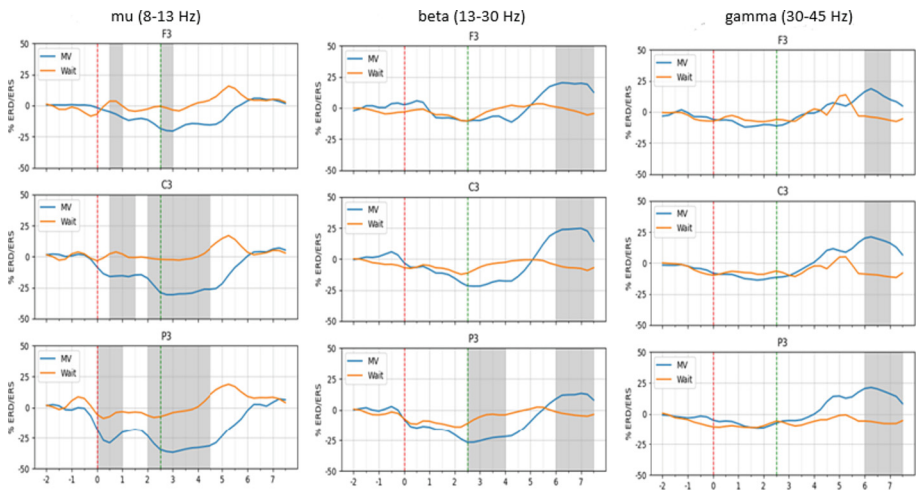


**Fig. 3.** Results of the statistical analysis of right arm movement (blue) vs. wait (orange) in the left hemisphere channels for the mu, beta, and gamma frequency bands (n = 11). The light gray areas represent time periods of statistical significance (p < 0.05).

The right arm motor imagery condition also presented an ERD in both preparation and execution periods, however with less intensity for the execution period compared to the movement condition (Fig. 4). Different from the movement, MI conditions for both arms did not present ERS after the execution of the task.

When comparing corresponding channels of different scalp hemispheres for the right arm movement (Fig. 5), we found a greater mu desynchronization at the contralateral channels during the execution time period and a greater beta ERS at the contralateral channels, after movement end. We did not find any difference in the beta band during the execution time period.

The left arm movement and wait conditions were also compared (data not shown). We found a significant mu and beta ERD during the execution of the movement and a significant beta ERS after the movement ended. These differences were equally found for both ipsi and contralateral channels. We also found a gamma ERS after movement end, that was greater for the contralateral channels. We did not find a significant mu ERD during the preparation. We then compared the left and right forearm movement
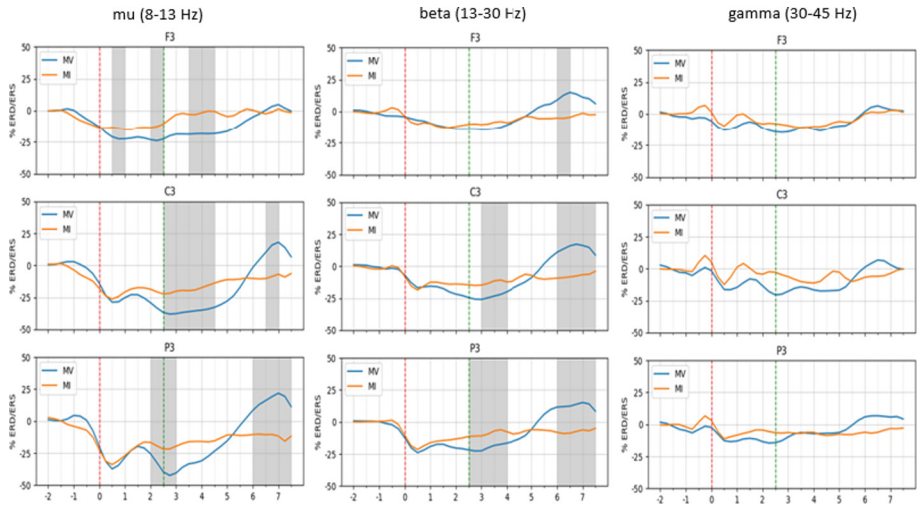
**Fig. 4.** LFCOG results of the statistical analysis for the right arm movement (blue) vs. right arm imagery (orange) in the left hemisphere channels (n = 12). The light gray areas represent time periods of statistical significance (p < 0.05).
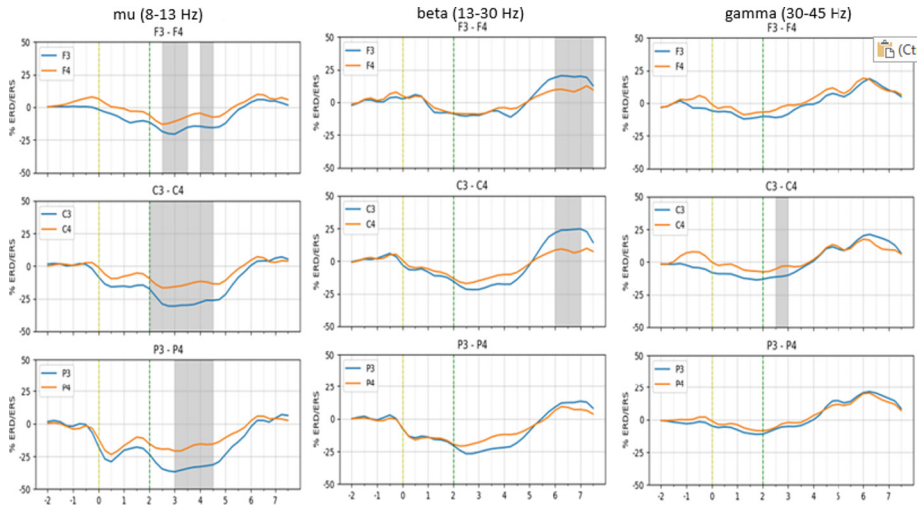


**Fig. 5.** Results of the statistical analysis for the right arm movement condition of the control group subjects, comparing right (blue) and left (orange) hemisphere channels (n = 11). The light gray areas represent time periods of statistical significance (p < 0.05).

conditions. We found a greater mu ERS for the left arm movement after the movement ended in the C4 channel. We also found a greater beta rebound in the P3 channel for the right arm movement condition. The left forearm imagery also elicited mu and beta ERD. When comparing right and left forearm imagery, we found a greater mu ERD during

execution in the C3 channel for the right arm imagery and a greater beta ERD during execution in the C4 channel for the left forearm imagery.

## 4   Discussion

This work used an experimental paradigm involving instructions with a delay period for studying the ERSP during the preparation and execution of imagery, movement, and wait conditions. As shown in previous studies [10, 11], we found that movement and imagination of the movement are associated with desynchronizations in mu and beta frequency bands, but the wait condition is not. When comparing movement and wait conditions, we found differences in the mu band for both preparation and execution time periods. For the beta band, the statistical differences are in the execution and after movement end time periods. We also found a beta ERS, known as beta rebound [14], after the movement ends. This indicates that these changes in the ERSP might represent motor processes rather than visual stimuli processing.

When comparing channels of the right and left hemispheres for the movement condition, we found greater mu ERD during the execution period and a greater beta ERS after the movement ends, both for the channel at the contralateral side from the movement. This might indicate that mu ERD during the execution and beta ERS after movement ends are more related to the lateralization of the task than the beta ERD during the execution.

The increased rhythmic activity (or ERS) of the EEG oscillations allows the synchronization of firing among populations of neurons as the ERD measures to which extent this synchronization was lost. Therefore, the mu and beta synchronization are thought to play an inhibitory role and guarantee the timing of cortical processes, while the desynchronization reflects the gradual liberation of this inhibition and is associated with activation processes [18, 19].

## 5   Conclusion

In this study we found mu and beta desynchronizations for movement and imagery conditions, but not for wait condition. The movement ERD is greater than the imagery ERD for both mu and beta bands, and we found an ERS after movement end period for mu and beta bands. These findings suggest that the sequences of neural events elicited by movement and imagination of the movement are similar, but for the imagery condition, the ERD is less pronounced due to the lack of muscular activation. The differences in the ERSP between the right and left forearm movement and imagery can be useful for BCI systems to differentiate which member should execute the movement.

# References

1. McFarland, D., Wolpaw, J.: EEG-based brain–computer interfaces. Curr. Opin. Biomed. Eng. **4**, 194–200 (2017). https://doi.org/10.1016/j.cobme.2017.11.004

2. Molinari, M., Masciullo, M.: Stroke and potential benefits of brain-computer interface. Handb. Clin. Neurol. **168**, 25–32. Elsevier (2020). https://doi.org/10.3390/s21134312

3. Sharma, N., Pomeroy, V.M., Baron, J.-C.: Motor imagery: a backdoor to the motor system after stroke? Stroke **37**, 1941–1952 (2006). https://doi.org/10.1161/01.STR.0000226902.433 57.fc

4. Fernández-Gómez, E., Sánchez-Cabeza, Á.: Imaginería motora: revisión sistemática de su efectividad en la rehabilitación de la extremidad superior tras un ictus. Rev. Neurol. **66**, 137 (2018). https://doi.org/10.33588/rn.6605.2017394

5. Santos-Couto-Paz, C., Teixeira-Salmela, L., Tierra-Criollo, C.: The addition of functional task-oriented mental practice to conventional physical therapy improves motor skills in daily functions after stroke. Braz. J. Phys. Ther. **17**, 564–571 (2013)

6. Page, S., Szaflarski, J., Eliassen, J., Pan, H., Cramer, S.: Cortical plasticity following motor skill learning during mental practice in stroke. Neurorehabil. Neural Repair **23**, 382–388 (2009). https://doi.org/10.1177/1545968308326427

7. Page, S., Levine, P., Sisto, S., Johnston, M.: A randomized efficacy and feasibility study of imagery in acute stroke. Clin. Rehabil. **15**, 233–240 (2001). https://doi.org/10.1191/026921 501672063235

8. Kansaku, K., Cohen, L., Birbaumer, N.: Clinical Systems Neuroscience. Springer, Tokyo, Japan (2015). https://doi.org/10.1007/978-4-431-55037-2

9. Pfurtscheller, G., Aranibar, A.: Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. Electroencephalogr. Clin. Neurophysiol. **46**, 138–146 (1979). https://doi.org/10.1016/0013-4694(79)90063-4

10. Pfurtscheller, G., Lopes da Silva, F.: Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol **110**, 1842–57 (1999). https://doi.org/10.1016/s1388-2457(99)00141-8

11. Galdo-Alvarez, S., Bonilla, F., González-Villar, A., Carrillo-de-la-Peña, M.: Functional equivalence of imagined vs. real performance of an inhibitory task: an EEG/ERP study. Front. Hum. Neurosci. (2016). https://doi.org/10.3389/fnhum.2016.00467

12. Tallon-Baudry, C., et al.: Oscillatory gamma activity in humans and its role in object representation. Trends Cogn. Sci. **3**(4), 151–162 (1999). https://doi.org/10.1016/s1364-6613(99)012 99-1

13. Amo Usanos, C., et al.: Induced Gamma-Band Activity during Actual and Imaginary Movements: EEG Analysis. Sensors (Basel, Switzerland), vol. 20, no 6, (2020). https://doi.org/10. 3390/s20061545

14. Kilavik, B., Zaeppfel, M., Brovelli, A., MacKay, W., Riehle, A.: The ups and downs of β oscillations in sensorimotor cortex. Exp. Neurol. **245**, 1526 (2013). https://doi.org/10.1016/ j.expneurol.2012.09.014

15. Weinrich, M., Wise, S.: The premotor cortex of the monkey. J. Neurosci. **2**, 1329–1345 (1982). https://doi.org/10.1523/JNEUROSCI.02-09-01329.1982

16. Crammond, D., Kalaska, J.: Prior information in motor and premotor cortex: activity during the delay period and effect on pre-movement activity. J. Neurophysiol. **84**, 986–1005 (2000). https://doi.org/10.1152/jn.2000.84.2.986

17. Makeig, S.: Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalogr. Clin. Neurophysiol. **86**(4), 283–293 (1993). https://doi.org/10. 1016/0013-4694(93)90110-h

18. Klimesch, W., Sauseng, P., Hanslmayr, S.: EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res. Rev. **53**, 63–88 (2007). https://doi.org/10.1016/j.brainresrev.2006.06.003
19. Takemi, M., Masakado, Y., Liu, M., Ushiba, J.: Event-related desynchronization reflects downregulation of intracortical inhibition in human primary motor cortex. J. Neurophysiol. **110**, 1158–1166 (2013). https://doi.org/10.1152/jn.01092.2012

# Assessing the Weighted Adaptive Filtering to Attenuate Eye-Blink Artefact by Means of Simulation for Brain-Computer Interface Application

Alice Fontes[(✉)] and Mauricio Cagy

Federal University of Rio de Janeiro, Rio de Janeiro RJ 21941909, Brazil
`fontes.alice@poli.ufrj.br`

**Abstract.** In Brain-Computer Interface (BCI) systems, the signal acquisition of brain's electrical activity, when non-invasive, is usually made with the electroencephalogram (EEG). EEG signals are naturally contaminated by artefacts which can significantly distort signals, altering neurological events, therefore compromising BCI control. In order to be applied in BCI systems, the method of artefact attenuation should be automatic, online, and ideally performed with few EEG recording channels. A previous work proposed a procedure for eye-blink artefact reduction based on adaptive filtering when electrooculogram data is not available. This method satisfies all the essential conditions for application in BCI systems and also addresses the bidirectional interference issue. In order to apply this technique in BCI systems, the present work aims to proceed with its assessment through changes in simulation in order to make the reproduced environment more realistic and therefore prove the reliability and effectiveness of the method. Results show that satisfactory artefact reduction is achieved even when its time occurrence overlaps the desired ERP (event-related potential). The lowest overall RMS error was achieved using a $2^{\text{nd}}$-order filter and adaptation factor set in $10^{-5}$. Furthermore, weight thresholds have a slight influence on filter performance when using plausible values. In conclusion, the proposed approach resulted in a considerable reduction of the eye-blink artefact, preserving ERP morphology and allowing easy component identification, all this in an online and automatic fashion, with few EEG recording channels and without the need for the reference channel, hence being a great choice to attenuate eye-blink artefacts in BCI systems.

**Keywords:** Artefact Attenuation · Brain-Computer Interface · Adaptive Filter · Eye Blink · Electroencephalography

## 1 Introduction

Information from neuronal activity can be translated into device control actions through brain-computer interface (BCI) systems [1]. In these systems, the measurement of the brain's electrical activity, when non-invasive, is usually performed by the electroencephalogram (EEG) [2]. The EEG is used to analyse the electrical activity of the brain

[3]. It has several applications, such as diagnosing and monitoring brain conditions and also the aforementioned BCI systems.

EEG signals have amplitudes in the range of microvolts [4], and therefore they are naturally contaminated by unwanted signals, known as artefacts [5]. Many different sources can provoke artefacts, which are present essentially in all EEGs [6]. The reliability of electrophysiological descriptions can be severely compromised by the presence of artefacts, particularly those from eye blinks [7].

Eye blinks generate a type of ocular artefact seen in almost all EEG exams performed with open eyes [3]. Ocular artefacts are caused by eye movements and blinks, which propagate through the scalp and can be recorded by the EEG [8]. The electrical activity generated by eye movement and eye blinking can be recorded by the electrooculogram (EOG). On average, the duration of an eye blink is between 100 and 400 ms [9], although this number may be influenced by several conditions, such as the spontaneous eye blink rate. It is estimated that the normal eye blink rate is between 15 and 20 blinks per minute [10]. There is usually an interval of 2 to 10 s between eye blinks [11].

Artefacts such as eye blinks can significantly modify the EEG signals, which may induce erroneous interpretations [12]. Thus, neurological events on the EEG which control the BCI systems can be easily disguised or mistaken for artefacts. In this way, artefacts can jeopardize commands and result in an unintentional control of the device [13]. Thereby, it is of great importance to attenuate EEG artefacts in order to keep only information of interest.

In this fashion, several artefact removal techniques are available in literature. In BCI applications, since real-time control of devices is required, this attenuation must be performed in a very short time, with low computational cost, in order to enable online signal processing. One should also prefer an automatic method, that is, a method that can be performed without the need for human intervention. In addition, in these systems, the possibility of using few channels in EEG recording is usually expected [14]. Therefore, methods that can be performed on single-channel EEGs are commonly preferred for this use. Most of the algorithms used for artefact removal have the disadvantage of demanding considerable processing time, usually enough to make the online operation unfeasible. When the method has low computational cost, most times it is not automatic and cannot be applied in single-channel systems.

Another important factor in the artefact removal process must also be taken into account, especially when removing ocular artefacts: just as the EEG can be corrupted by the EOG, the EOG can also be contaminated with the EEG. This event is called bidirectional interference. Without due consideration, it may remove relevant EEG information along with the artefact, hence causing a removal error [15]. This is a risk that a large parcel of the artefact attenuation techniques suffer.

In previous works [7, 16], a pre-processing framework with a noise reduction procedure based on LMS adaptive filtering has been proposed. In these studies, a weighted adaptive filtering technique was used, aiming to attenuate blinking artefacts. The goal was to facilitate as much as possible the identification of event-related potentials (ERPs), in order to use it in Stroop-like cognitive tests. The proposed approach resulted in a considerable eye-blink artefact reduction, allowing an easy identification and marking of ERP components [16]. This noise reduction technique allows online and automatic

implementation, can be used in a single-channel EEG, and also deals with bidirectional interference, hence being a great choice for BCI applications. In this work, with the aim of applying the technique as a pre-processing step in BCI systems, investigations on the weighted adaptive filtering method were carried out by means of simulated data. Changes in simulation were made in order to make the reproduced environment even more realistic and therefore prove the reliability and effectiveness of the method.

## 2  Materials and Methods

### 2.1  Weighted Adaptive Filtering

The weighted adaptive filter inherits properties from the traditional adaptive filter. Its coefficients are adaptable and can be automatically adjusted by the filter [17]. In this way, these filters can be adapted to the context defined by the signals [17], detecting any changes in the environment and accomplishing the adaptation.

Unlike traditional adaptive filters, in the scheme of the weighted adaptive filter the subtraction of the reference noise signal estimation is not completely performed all the time. Subtraction is made in a weighted fashion, with a weight fluctuating between 0 (no subtraction) and 1 (complete subtraction). The lowest weight will be given at moments when there is no blink, which are moments when the current information in EOG is only about the signal of interest, the EEG. The highest weight will be given when the eye blink is detected. Therefore, the system subtracts the noise estimation only at times when there is evidence of an eye blink occurrence.

The eye blink can be evidenced by an increase of the signal amplitude [15]. Thus, when the amplitude of the reference channel is small, the weight of the subtraction is minimised. When the amplitude is greater than expected, the subtraction is done with a higher weight. The range between 0 and 1 that will be used will depend on the amplitude and instantaneous RMS value of the signal. Fractional values between 0 and 1 can also be assigned. This is a method for weighting the subtractive process of adaptive filtering according to the amplitude of the reference signal with noise pattern.

The block diagram scheme for the proposed weighted adaptive filter for noise reduction is shown in Fig. 1. In the weighted adaptive filter, as in the traditional adaptive filter, two signals are needed [17]: the reference signal ($r[n]$), which carries information about the noise (should be a signal such as EOG); and the primary input signal ($s[n]$), modelled as a combination of pure EEG data ($x[n]$) and an artefact source ($i[n]$) (should be a signal such as EEG).

The filter (with coefficients $h[n]$) adapts itself, so that the difference between the collected noisy signal and the filtered reference noise reaches minimal energy. The filter output (error signal) converges to the noiseless signal estimation after adjusting the filter coefficients (Fig. 1: $y_a[n] = e[n]$). The effective output, however, is not given by $e[n]$. The estimation of the signal of interest is taken from another subtracting node where the filtered noise is weighted by a value between 0 and 1, depending on the instantaneous RMS value of $r[n]$ (Fig. 1: $y_b[n]$). Thus, it is avoided that a large amount of frontal neural activity is cancelled along with the noise. The filter output ($y_a[n]$) and the effective output ($y_b[n]$) can be expressed as shown in Eqs. 1 and 2, respectively.

Calculation of RMS values is implemented prior to the noise reduction procedure so that a null-phase envelope detector can be implemented. RMS values below a lower threshold produce a null weight. This weight increases proportionally up to 1, when the RMS values reach an upper threshold [7].

$$y_a[n] = e[n] = s[n] - y[n] = x[n] + i[n] - y[n] \tag{1}$$

$$y_b[n] = s[n] - \beta(\rho).y[n] \tag{2}$$
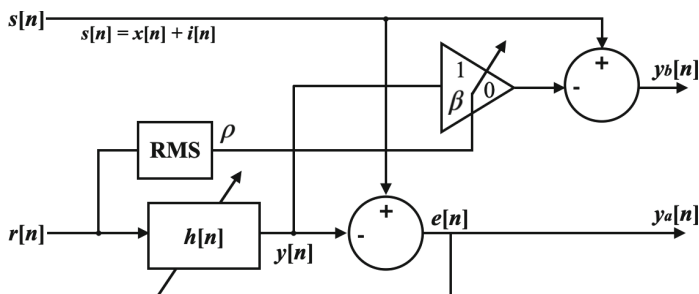


**Fig. 1.** Block diagram of the proposed weighted adaptive filter for noise reduction

## 2.2 Simulation Procedure

Routines in GNU Octave (version 7.1.0) were implemented in order to simulate the pure EEG signal, composed by the background EEG and the ERP, and also to simulate the corrupted EEG signal, composed by the background EEG, the ERP and the eye blink.

Four EEG derivations (Fp1, Fp2, Cz and Oz) were simulated in this work, for simplicity. The simulated signal mimics an EEG recording during auditory stimulation, where the goal is to perform an evoked potential analysis. A sample rate of 400 Hz and number of epochs of 240 were used.

The reference noise is usually obtained from EOG data, when available. In this work, the reference noise was estimated from the average of Fp1 and Fp2 derivations, mimicking contexts where no EOG derivation was available.

Background EEG, ERP and eye-blink artefact characteristics are reproduced in the simulated signal. The background EEG signal was simulated as a coloured noise, with higher power in the alpha band [16]. Four coloured Gaussian noises ($r_1[n]$ to $r_4[n]$) were combined using a mixing matrix as follows:

$$\begin{bmatrix} Fp1[n] \\ Fp2[n] \\ Cz[n] \\ Oz[n] \end{bmatrix} = \sqrt{\begin{bmatrix} 0.5 & 0.3 & 0.15 & 0.05 \\ 0.3 & 0.5 & 0.15 & 0.05 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.4 \end{bmatrix}} \begin{bmatrix} r1[n] \\ r2[n] \\ r3[n] \\ r4[n] \end{bmatrix} \tag{3}$$

Where $\sqrt{\bullet}$ refers to element-wise square root, and the combined noises were coloured with different bandwidths, in order to represent the differences among derivations (8–35 Hz for $r_1$ and $r_2$, 5–25 Hz for $r_3$ and 0.5–12 Hz for $r_4$) [16]. The mixing matrix is used to represent contamination between the EEG derivations. The values were chosen heuristically, based on typical situations of real signals.

An experimental ERP waveform was used as a model to represent the evoked activity, with same latencies but different amplitudes among derivations (0.5 in Fp, 1.0 in Cz and 0.8 in Oz) [16]. This simulates the intensity of ERP occurrence in each channel. The simulated evoked potential is a middle latency auditory evoked potential (MLAEP), which has a greater amplitude in Cz than in the other derivations.

The eye-blink artefact was simulated by means of the derivative of a Gaussian curve weighted by a decreasing exponential, producing a positive peak with higher amplitude than the successive negative valley, and its influence over EEG also varied among derivations (1.0 in Fp, 0.5 in Cz and 0.25 in Oz) [16].

In addition to these amplitude changes between derivations, different low-pass filters were also applied to the artefact waveform in Cz and Oz derivations, aiming at slight changes in the morphologies of the eye blink in these derivations when compared to the ones in the reference signal. Since the reference signal is obtained with the Fp1 and Fp2 derivations, filters were not applied to these channels. This behaviour reflects the high frequency attenuation that is expected with the propagation of the eye-blink artefact along the scalp. In [16], the eye-blink artefact was added to the EEG with different amplitudes but same morphology, addressing only the characteristic of different amplitudes of the artefact among derivations. In the present work, the feature of different artefact's morphologies among distinct derivations is also being addressed. The low-pass filters (2nd-order Butterworth) show cut-off frequency at 5 Hz for Cz and 3 Hz for Oz, and the resulting artefact waveforms are then added to the signal along with the ERP.

The post-stimuli instant of blinking occurrence varied randomly with a standard deviation of 250 ms around three mean values: 250 ms, 500 ms and 750 ms. Thus, eye-blinks occurred randomly, with arbitrary intervals, which mimic our eye blink rhythm.

The adaptive filter order (0 [pure gain] up to 6) and the adaptation factor ($10^{-5}$, $5 \times 10^{-6}$, $10^{-6}$) are other parameters that varied in this simulation. Those are values in which there was convergence, chosen by exploratory analysis. The parameters that yielded the lowest relative RMS error were fixed. The thresholds for null ($Th_{min}$) and full ($Th_{max}$) filtering weight, initially set at 10 μV and 60 μV, respectively, varied around these levels, within EEG's dynamic range.

In order to evaluate noise reduction performance, three levels of the simulated signal were used to produce average morphologies: background EEG plus ERP (estimation of the desired signal); background EEG plus ERP and the artefact, without using the noise reduction procedure; and background EEG plus ERP and the artefact, using the procedure.

Routines in GNU Octave were implemented in order to perform the attenuation process of eye-blink artefacts. Several parameter sets were tested for the filter. In order to evaluate the accuracy of each combination of the parameters, the RMS error was calculated for each derivation and additionally an average of the errors among derivations

was performed. In this way, one assessed the smallest RMS error for the average of all four derivations and also whether a derivation performed better in any scenario.

## 3   Results

According to the eye blink occurrence, the attenuation of the artefact was evaluated in three moments: early eye blink (total overlap with the ERP component), intermediate eye blink (partial overlap), and late eye blink (no overlap). The $6^{th}$-order filter with adaptation factor set in $10^{-5}$ yielded the worst RMS errors, hence this configuration was excluded from the following analysis.

Considering early eye blink, the lowest RMS errors for all derivations were obtained using orders 0 and 1, with adaptation factor set in $10^{-6}$. For intermediate eye blink, Fp derivations had lowest RMS error using $1^{st}$-order filter and adaptation factor of $10^{-6}$; for Cz and Oz derivations, artefact reduction performed better using $2^{nd}$- and $3^{rd}$-order filters with the adaptation factor set in $10^{-5}$. For late eye blink, low RMS error values occurred for a wide range of parameter combination: all sets for Fp derivations, and combining orders 2 up to 5 with adaptation factor of $10^{-5}$ for Cz and Oz.

Considering an overall analysis based on the average performance of the three moments of eye blink occurrence, the best scenario was $2^{nd}$-order filter (closely followed by $3^{rd}$-order filter) with adaptation factor set in $10^{-5}$. Hence, this parameter combination was set fixed in order to assess different threshold combinations; the relative RMS error varied slightly, with the lowest value obtained with $Th_{min} = 10 \,\mu V$ and $Th_{max} = 60 \,\mu V$. In this configuration, the resulting waveforms are shown in Fig. 2. In general, one can observe that the filtered waveform (continuous black line) approximates the goal ERP estimation (thick grey line), considerably reducing the influence of the eye-blink artefact (shown in dotted line). Table 1 shows the RMS errors for each derivation and moment of eye blink occurrence considering the three parameter combinations with best performance for each scenario and overall (column-wise ordered by means of the sum of average errors). The best global performance (first column) coincides with the best performance for intermediate eye blink. The second column points out the best performance for late eye blink and second best global scenario. On its turn, the third column indicates the best performance for early eye blink.

Therefore, the proposed procedure showed considerable artefact reduction and allowed the identification of ERP components that are severely deteriorated by eye-blink artefacts. A small portion of ERP interesting components was removed in frontopolar waveforms only when time occurrence of eye-blinks overlapped the entire ERP period (Fig. 2a). In spite of this fact, the waveforms were not completely removed, making their identification possible since they still remained recognizable in the resulting ERP. The Cz and Oz filtered waveforms closely follow the ERP pattern despite the presence of the high-amplitude artefact. As can be observed in Fig. 2, amplitude differences occur in the frontopolar derivations due to eye blinks and progressively decrease in the posterior derivations.
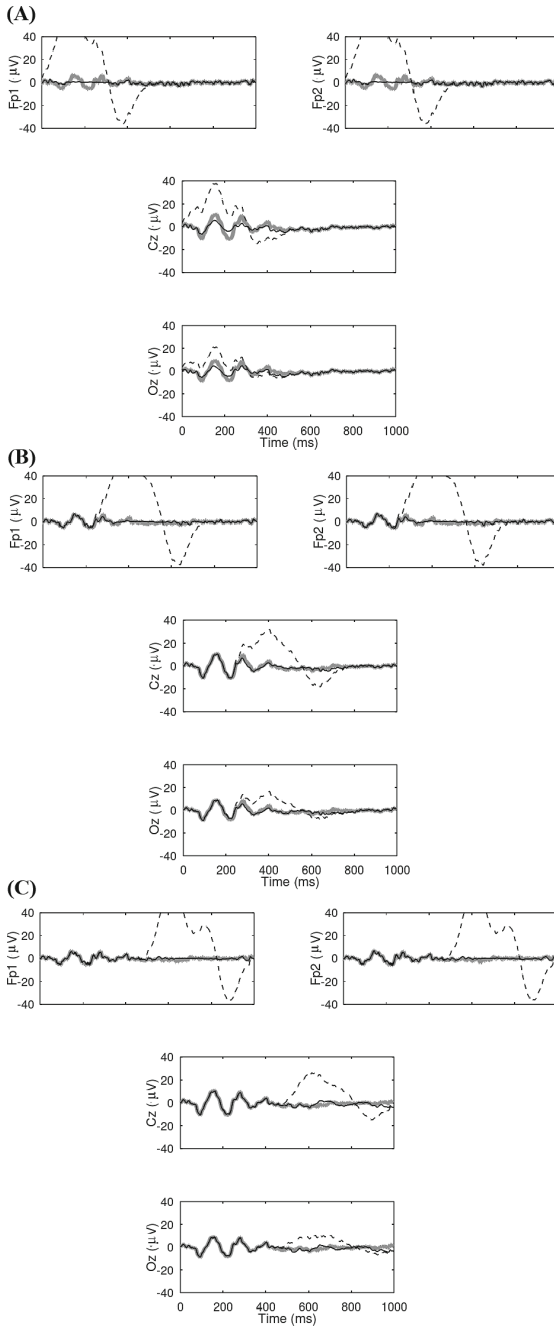
**Fig. 2.** Comparison of goal waveform (thick grey line) with those after using the developed filtering (solid black line) and not using it (dotted line). Mean time of eye-blinking is: (a) 250 ms, (b) 500 ms and (c) 750 ms.

**Table 1.** RMS errors for the best scenarios of each moment of eye blink occurrence. The performances are ordered based on total RMS error, which is the sum of average errors for each moment of eye blink occurrence. The best overall result for each derivation and average is presented in bold.

| Eye blink occurrence | Derivation | $2^{nd}$-order $10^{-5}$ | $3^{rd}$-order $10^{-5}$ | $1^{st}$-order $10^{-6}$ |
|---|---|---|---|---|
| Early | Fp1 | 1.8771 | 1.8759 | **1.8647** |
| | Fp2 | **1.8572** | 1.8575 | 1.8609 |
| | Cz | 2.5373 | 2.5713 | **2.2767** |
| | Oz | 1.8497 | 1.8827 | **1.5093** |
| | Average | 2.0303 | 2.0469 | **1.8779** |
| Intermediate | Fp1 | 0.8848 | 0.8812 | **0.8665** |
| | Fp2 | 0.8738 | 0.8713 | **0.8658** |
| | Cz | **1.2382** | 1.2983 | 1.6795 |
| | Oz | **1.3328** | 1.3495 | 1.3553 |
| | Average | **1.0824** | 1.1001 | 1.1918 |
| Late | Fp1 | 0.8909 | **0.8876** | 0.9312 |
| | Fp2 | 0.8864 | **0.8821** | 0.9297 |
| | Cz | 1.3479 | **1.3123** | 1.9444 |
| | Oz | 1.3900 | **1.3800** | 1.5442 |
| | Average | 1.1288 | **1.1155** | 1.3374 |
| Total | | **4.2415** | 4.2625 | 4.4071 |

## 4   Discussion

The artefact removal technique should be chosen according to the peculiarities and purposes of each project, since there are several techniques and each one has its particularity, advantages and disadvantages [14]. In BCI systems, since real-time control of devices is required, an online method should be chosen. It means that the data should be automatically updated as changes are made and the commands should result in responses in real time or in a sufficiently short time so that the control is not compromised. Furthermore, an automated execution is essential, that is, the technique should automatically identify and eliminate artefact components without the need for manual intervention. This enables online application and also avoids the insertion of subjectivity in the process [8]. The developed technique achieved both of these goals, hence being a good choice for artefact attenuation in BCI systems.

The number of EEG recording channels used is another important factor to be taken into account. In BCI systems, the possibility of using few channels (few electrodes) in the EEG acquisition is expected [14]. This implies ease of use and makes the EEG recording process more user-friendly. Not only are there few available techniques that

are able to perform the removal of ocular artefacts from single-channel EEG, but also most of them modify uncontaminated frequency bands of the EEG signal [18].

The adaptive filter, one of the main filtering techniques used to remove EEG artefacts [8], allows removing noise from a signal in real time [19], automatically, and it can also operate in a single-channel EEG. Nevertheless, it does not deal with bidirectional interference. Adaptive filtering requires the existence of a reference signal that has only information about the eye blink, without any reminiscence of neural activity of interest and, due to bidirectional interference, this does not occur. Consequently, a part of possibly important EEG activity might be cancelled along with the artefact.

The weighted adaptive filtering method aiming noise cancellation proposed in this work combines the possibility of online and automatic performance, which is essential in applications such as BCI systems, with the fact that it only produces some level of attenuation in the instants of eye blink, minimising attenuation of important information. Under these circumstances, erroneous removals are avoided, hence the impact of bidirectional interference is minimised.

Moreover, with the developed technique, if the reference channel is not available, it is possible to estimate it with the frontopolar derivations. In this noise estimation, usually there is the disadvantage that a large amount of frontal neural activity is also cancelled. In this work, this is mitigated by the use of weighted adaptive filtering instead of traditional adaptive filtering, in the same way that the loss of neural activity due to bidirectional interference is also diminished. Even when an EOG derivation is available, bidirectional interference may cancel out neural activity of interest, and this technique mitigates this effect.

In this way, even in the absence of EOG derivations, the developed adaptive filtering procedure was able to attenuate eye-blink artefacts. Therefore, it showed a balanced trade-off between artefact reduction and possible suppression of frontal neurological activity. The results also support that the high-amplitude activity recorded in frontopolar derivations is a good estimation of the artefact source. Thus, it can be a good alternative in BCI systems where the reference channel (EOG) is not available.

When no EOG derivation is available, the EOG data can also be estimated from Independent Component Analysis (ICA) applied to raw EEG [20]. Algorithms based on ICA are widely used to remove artefacts as well [8], but they have the disadvantage of demanding a considerable processing time, enough to make online application unattainable [14]. In addition, this technique is automatic only when combined with another technique [14]. Furthermore, it cannot be used in a single channel due to the principle of blind source separation, which says that more channels will bring better accuracy [14]. In this way, this EOG estimation approach would not be a good fit for a BCI system.

Furthermore, in the present work, the EEG and ERP characteristics found in real signals were reproduced in the simulated signals. In background EEG's simulation, both bandwidth differences and inter-channel correlations were taken into account. The ERP and artefacts distribution also sought to reflect differences between derivations. The low-pass filters applied to Cz and Oz derivations mimics the behaviour of high frequency attenuation with propagation of the eye-blink artefact along the scalp that occurs in real signals. Hence, the simulated conditions make a good representation of reality.

Additionally, this noise reduction approach could also be applied in other cases of high-amplitude artefacts that predominantly affect few derivations, such as temporal muscle activity.

Further work will be carried out in order to continue the assessment of the performance variability and robustness of the method as well as to compare this technique's performance with other eye-blink removal procedures, such as ICA.

## 5   Conclusion

The proposed adaptive filtering approach resulted in a considerable reduction in the eye-blink artefact. This allowed easy identification of ERP components that are severely deteriorated by eye-blink artefacts, even when the artefact overlaps the ERP. Therefore, the application of this technique as a pre-processing step in BCI applications can be of great importance for the reliability of the commands, hence their proper functioning. A great virtue of the developed noise reduction technique is that it can be applied in real time and therefore can be used in BCI systems. Moreover, this technique can also improve the analysis and description of ERP in different stimulation paradigms.

Future studies will be held aiming at evaluating the performance of the technique in a more exhaustive way with emphasis on robustness and performance variability, performing more simulations and applying this algorithm in real signals acquired during protocols involving brain-computer interface.

**Conflict of Interest.**   The authors declare that they have no conflict of interest.

## References

1. Wolpaw, J.R., et al.: Brain-computer interfaces for communication and control. Clin. Neurophysiol. **113**(6), 767–791 (2002)
2. Saha, S., Mamun, K., Ahmed, K., et al.: Progress in brain computer interface: challenges and opportunities. Front. Syst. Neurosci. **15**, 578–875 (2021)
3. Niedermeyer, E., Schomer, D., Lopes, F.: Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 6th edn. Lippincott Williams & Wilkins, Philadelphia (2011)
4. Bear, M.F., Connors, B.W., Paradiso, M.A.: Neurociências: desvendando o sistema nervoso. 2nd edn. Artmed, Porto Alegre (2002)
5. Cosandier-Rimele, D., Badier, J.M., Chauvel, P., Wendling, F.: A physiologically plausible spatio-temporal model for EEG signals recorded with intracerebral electrodes in human partial epilepsy. IEEE Trans. Biomed. Eng. **54**(3), 380–388 (2007)
6. Aurlien, H., Gjerde, I.O., Aarseth, J.H., et al.: EEG background activity described by a large computerized database. Clin. Neurophysiol. **115**(3), 665–673 (2004)
7. Cagy, M., Cagy, M., Infantosi, A. F. C.: Pre-Processing Framework to Help ERP Identification Applied to a Word-Colour Cognitive Test. In: Anais do 22o. Congresso Brasileiro de Engenharia Biomédica, pp. 900–903. SBEB, Tiradentes, Brasil (2010)

8. Urigüen, J.A., Garcia-Zapirain, B.: EEG artifact removal - state-of-the-art and guidelines. J. Neural Eng. **12**(13), 031001 (2015)
9. Schiffman, H.R.: Sensation and Perception, An Integrated Approach, 5th edn. Wiley, New York (2001)
10. Tsubota, K., Hata, S., Okusawa, Y., et al.: Quantitative videographic analysis of blinking in normal subjects and patients with dry eye. Arch. Ophthalmol. **114**, 715–720 (1996)
11. Bentivoglio, A.R., Bressman, S.B., Cassetta, E., et al.: Analysis of blink rate patterns in normal subjects. Mov. Disord. **12**(6), 1028–1034 (1997)
12. Antoniol, G., Tonella, P.: EEG data compression techniques. IEEE Trans. Biomed. Eng. **44**(2), 105–114 (1997)
13. Vaughan, T.M., et al.: Brain-computer interface technology: a review of the Second International Meeting. IEEE Trans. Neural Syst. Rehabil. Eng. **11**(2), 94–109 (2003)
14. Jiang, X., et al.: Removal of artifacts from EEG signals: a review. Sensors **19**(5), 987 (2019)
15. Croft, R.J., Barry, R.J.: Removal of ocular artifacts from the EEG: a review. Neurophysiol. Clin. **30**(1), 5–19 (2000)
16. Cagy, M., Cagy, M., Infantosi, A. F. C.: Variable-weight adaptive filtering as an ERP pre-processing technique. In: Jobbágy, Á. (eds.) 5th European Conference of the International Federation for Medical and Biological Engineering. IFMBE Proceedings, vol. 37. Springer, Berlin, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23508-5_58
17. Ghmati, T.S., Elhoula, A.S.: Adaptive digital FIR filters; case study: noise cancellation using LMS algorithm. AJAS **2**(1), 30–36 (2021)
18. Maddirala, A.K., Veluvolu, K.C.: Eye-blink artifact removal from single channel EEG with k-means and SSA. Sci. Rep. **11**(1), 11043 (2021)
19. Haykin, S.: Adaptive Filter Theory, 4th edn. Prentice Hall, New Jersey (2002)
20. Guerrero-Mosquera, C., Vazquez, A.: Automatic removal of ocular artifacts from EEG data using adaptive filtering and independent component analysis. In: Proceedings of the 17th European Signal Processing Conference, pp. 2317–2321. EURASIP, Glasgow, Scotland (2009)

# Classification of Autism Spectrum Disorder Using a 3D-CNN Ensemble Model and Regional Homogeneity Data from the ABIDE I Dataset

Guilherme Bauer-Negrini[1]([envelope]) [ID], Luísa Vieira Lucchese[2] [ID],
Viviane Rodrigues Botelho[3] [ID], Thatiane Alves Pianoschi[3] [ID],
and Carla Diniz Lopes Becker[3] [ID]

[1] Department of Biochemistry, Institute of Health Sciences, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, RS 90040-060, Brazil
gbnegrini@pitt.edu
[2] Institute of Hydraulic Research, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, RS 91501-970, Brazil
[3] Research Group on Intelligent Systems Applied to Health, Federal University of Health Sciences of Porto Alegre (UFCSPA), Porto Alegre, RS 90050-170, Brazil

**Abstract.** Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in communication and social interaction, and presence of repetitive behaviors, and restricted interests. Many structural and functional brain alterations can be observed in individuals with ASD. Therefore, functional magnetic resonance imaging (fMRI) has been used to uncover the neurobiology of this disorder. Numerous studies have used machine learning models to classify ASD using the Autism Brain Imaging Data Exchange fMRI dataset. Currently, the state-of-the-art accuracy is close to 70% when using multi-site samples. Here, we propose a 3D version of the SqueezeNet resource-efficient architecture. We used regional homogeneity data available from the preprocessed version of the ABIDE I dataset. A 3D-CNN ensemble model was trained using a 5-fold cross-validation procedure. Our model achieved 68.93% accuracy and 0.72 Area Under the Receiver Operating Characteristic Curve (AUC) value. Despite not achieving the highest overall accuracy, our SqueezeNet-based 3D-CNN can still outperform another 3D-CNN model based on the same dataset. Using a resource-efficient architecture, we developed a model capable of processing complex 3D medical imaging data. Moreover, we achieved comparable state-of-the-art results.

**Keywords:** Convolutional Neural Networks · fMRI · Machine Learning

## 1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in communication and social interaction and by the presence

of repetitive behaviors and restricted interests and activities [1]. It is currently estimated that between 1% and 2% of individuals are affected worldwide [2]. There is still no biomarker validated for use in clinical practice, and the diagnosis remains based on clinical observations [3]. This poses significant challenges for healthcare professionals considering the heterogeneous characteristics of this disorder.

Resting-state functional magnetic resonance imaging (rs-fMRI) is one of the most used techniques to investigate the human brain functional connectivity [4] and has been used to study the neurobiological bases of ASD. Individuals with ASD can present several structural and functional brain alterations, including increased total brain volume during early childhood [5] and regional homogeneity alterations of resting-state brain activity [6]. The International Neuroimaging Data-sharing Initiative created the Autism Brain Imaging Data Exchange (ABIDE) initiative to aggregate functional and structural brain imaging data collected from international imaging sites to accelerate ASD neuroimaging studies [7]. This consortium aggregated a collection (ABIDE I) of mostly unpublished samples from 17 independent sites and released it in August 2012 [7].

Several machine learning models were developed to classify ASD using the ABIDE dataset. Heinsfeld *et al.* (2018) [8] used deep neural networks (DNN) and feature selection from functional connectivity of ROIs to identify ASD with 70% of accuracy. This was one of the first results surpassing the prior 60% accuracy threshold for deep learning [9,10] with a comprehensive dataset that contained data from samples of several group ages, genders, and sites. More recently, Yang *et al.* (2020) [11] and Ingalhadlikar *et al.* (2021) [12] have achieved, respectively, 75.27% and 71.35% accuracy using similar feature selection procedures and artificial neural networks. However, significantly higher metrics, such as 95% accuracy, can be achieved when classification is done on subsets of age groups [13] or constrained sample sizes [14]. Aggregating samples from multiple sites, age groups, and genders to increase sample size can certainly introduce great variation in the dataset [10]. Nevertheless, an ideal biomarker or predictive model should be able to represent greater variance between subjects and imaging practices [9].

Convolutional Neural Networks (CNNs) have been shown to perform well in classification tasks using fMRI data [4,15]. Sherkatghanadet *et al.* (2020) [16] used a 2D-CNN to achieve 70.22% of accuracy using the ABIDE dataset and using functional connectivity of ROIs as features. El-Gazzar *et al.* (2019) developed a 3D-CNN to directly use 4D spatio-temporal data without the need to compute a correlation matrix of functional connectivity ROIs as in previous studies. However, using data from all sites of the ABIDE dataset, the 3D-CNN achieved only 58% of accuracy. Recently, Thomas *et al.* (2020) [17] developed an ensemble model that uses multiple statistical derivatives as input to a 3D-CNN, achieving 64% accuracy in classifying individuals with ASD in a combined ABIDE I and ABIDE II dataset.

3D-CNNs models have the advantage of working directly with 3D rs-fMRI data while keeping the spatial resolution intact. However, 3D-CNNs can require significantly more parameters to adjust compared to their 2D counterparts, mak-

ing them harder to train and prone to overfitting [18]. To overcome this challenge, one strategy is to design a 3D version of a 2D resource-efficient CNN [18].

Here, we developed a 3D version of the 2D resource-efficient SqueezeNet to classify Autism Spectrum Disorder (ASD) from regional homogeneity data from the ABIDE I dataset. With the use of this lightweight architecture, we demonstrate that it is possible to achieve state-of-the-art results while maintaining the spatial structure of the input data.

## 2  Materials and Methods

### 2.1  Data Description

We used publicly available data from the preprocessed version of the ABIDE I dataset [19]. ABIDE I is composed of structural, resting-state fMRI, and phenotypic data completely anonymized from multiple sites following international guidelines and approved by local Institutional Review Boards [7].

The preprocessing pipeline was Configurable Pipeline for the Analysis of Connectomes (CPAC). Preprocessing steps for this version included slice time and motion correction, intensity normalization (4D Global mean = 1000), nuisance signal regression, band-pass filtering (0.01–0.1 Hz), and registration to a 3mm Montreal Neurological Institute (MNI) standard template. The final fMRI volume dimensions were $61 \times 73 \times 61$ (height, width, depth).

We chose regional homogeneity (ReHo) values from the statistical derivatives available in the preprocessed ABIDE I dataset. ReHo is a measurement of the similarity of the time series of a given voxel to those of its nearest neighbors in a voxel-wise way, and it is based on the hypothesis that intrinsic brain activity is manifested by clusters of voxels rather than single voxels [20]. One interesting aspect of ReHo is that it requires no *a priori* definition of regions of interest (ROIs) and can provide information about the local/regional activity of regions throughout the brain [21]. ReHo analysis has been extensively used to investigate resting-state function connectivity alterations in many disorders, including ASD [22]. Furthermore, ReHo indirectly reflects the local encephalic activation of the region surrounding a given voxel, which may favor feature extraction by a CNN that maintains the spatial structure of the data.

### 2.2  Model Architecture

We proposed a 3D SqueezeNet-based model, similar to previously developed by Köpüklü *et al.* (2019) [18], and with added alterations to the original SqueezeNet architecture [23]. The SqueezeNet-based macro architecture (Fig. 1) is composed of 9 convolution layers, including 8 Fire modules. Here, a 3D version of the Fire module is responsible for reducing the number of input elements with $1 \times 1 \times 1$ filters (squeeze layer) and then expanding the elements with a combination of $1 \times 1 \times 1$ and $3 \times 3 \times 3$ filters (expand layer). We decreased the original number of convolutional kernels in the Conv1 layer from 96 to 64. The original architecture

increases the number of kernels in every 2 Fire modules in multiples of 128 and here it was increased in multiples of 64. MaxPooling3D layers are used to reduce input size while keeping relevant information by merging semantically similar features into one [24]. The original SqueezeNet replaces the traditional fully connected layer with a GlobalAveragePooling3D layer to reduce the overfitting phenomenon, which was observed when using fully connected layers.

The activation layers use the rectified linear unit (ReLU). ReLU inserts non-linearities in the network to obtain sparse representations and model complex functions [25]. To reduce model overfitting, one dropout layer (50%) [26] was used after the layer Fire 8, as in the original, and an extra L1 regularization penalty of 0.1 was applied to the kernel initializer of the output layer. The weight initializer used the method proposed by He *et al.* (2015), as opposed to the original SqueezeNet choice of Xavier method, due to the good performance demonstrated in combination with rectifier activation functions [27]. The output layer has a single neuron with a sigmoid activation function, compatible with the binary classification task.



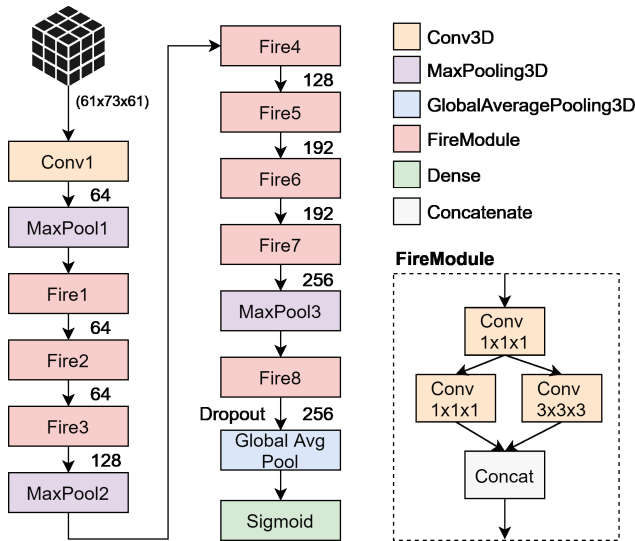**Fig. 1.** SqueezeNet-based 3D-CNN architecture. Numbers next to arrows indicate the number of convolution kernels in each layer. In the detail, the architecture of the FireModule is shown.

## 2.3   Model Training and Validation

A total of 884 samples (476 TDC and 408 ASD) were obtained from the preprocessed dataset. Samples were initially divided into two main groups following an 80/20 ratio:

- Train and validation: samples used for 5-fold cross-validation and hyper-parameter optimization (373 TDC, 334 ASD).
- Test: hold-out samples not used for any of the training or validation and used only to calculate the final metrics (103 TDC, 74 ASD).

The training procedure is illustrated in Fig. 2. We used $k$-fold cross-validation [28] as a model selection strategy. We randomly split the train set into 5-folds, in which the first segment were used as validation data, and the remaining four segments were used to train the model. Here, for each fold, 5 models were initialized with random weights and trained. The validation data ($k$ segment) of the fold was used to select the best of the 5 models based on the maximum validation accuracy value. Model weights were saved on every validation accuracy improvement. An early stopping criterion was adopted and the training was stopped if the validation accuracy had not improved in the last 50 epochs. The maximum number of epochs was fixed as 300.

All models were trained to minimize the value of the binary cross-entropy cost function. The Adam optimizer algorithm was chosen for training due to its computational efficiency and performance in problems with many parameters [29]. The learning rate was determined empirically to maximize the validation accuracy and was fixed throughout the training as $10^{-5}$. The number of training samples per batch was 64 to avoid exhausting the memory space.
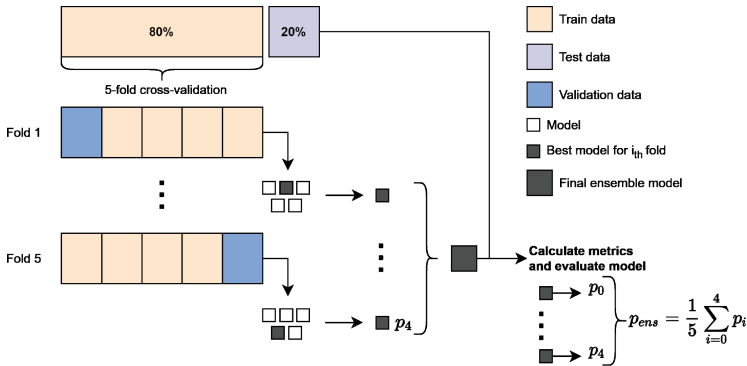


**Fig. 2.** Model development procedure. A holdout of 20% is selected to evaluate the model. Then, the remaining data is split into 5 folds and goes through cross-validation training. The ensemble model averages the predicted probabilities of the 5 models to calculate the final metrics based on the holdout test set. $p_i$ denotes the predicted probability of a model instance.

In the end, 5 3D-SqueezeNet models (the best selected models generated in each $k$ fold) were used to compose the final ensemble model. To evaluate metrics, the hold-out test set was used. The ensemble was of the bagging type, in which the average of the probability output by the 3D-SqueezeNet instances is calculated. Bagging equalizes the influence of the involved original estimators

[30]. It is also shown to decrease estimator variance in most cases, however, increasing bias (for one class or another) [31].

## 2.4  Evaluation Metrics

Metrics used to evaluate the model were based on the output probabilities. The default cut-off probability, *i.e.* threshold, to perform binary classification of balanced datasets is 0.5. However, the default value does not necessarily preserve the observed prevalence or result in the highest prediction accuracy [32]. Previous papers [33–35] calculated the threshold based on the maximum accuracy achieved. For the present analysis, we analyzed the thresholds that would provide the highest accuracy for each fold based on the validation samples. This resulted in the values of, respectively, 0.50, 0.54, 0.41, 0.51, and 0.45. Based on these thresholds, the default value of 0.5 was considered adequate to analyze the ensemble model in the test set.

Accuracy (acc), True Positive Rate (TPR, also called Sensitivity or Recall), False Positive Rate (FPR), Positive Predictive Value (PPV, also called Precision), True Negative Rate (TNR, also called Specificity) were calculated as follows:

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

$$FPR = \frac{FP}{FP + TN} \tag{3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

$$TNR = \frac{TN}{TN + FP} \tag{5}$$

The Area Under the Receiver Operating Characteristic Curve (AUC) was also employed as an evaluation metric for the results. The Receiver Operating Characteristic (ROC) curve is the plot of the TPR by the FPR for a range of classification thresholds. AUC was first proposed by DeLong et al. (1988) [36] and is the integration of the area below the ROC curve. It varies between 0.5, for classification models which cannot discriminate between classes, and 1.0, for perfect models in which the FP and FN classifications are null.

## 2.5  Computing Environment

Code was written using the Python 3.7 programming language. The model was built using the Keras (version 2.2.1) framework for TensorFlow (version 2.4.1). Hardware resources included an Intel Xeon central processing unit (2vCPU, 2.20 GHz), 12 GB of system random access memory (RAM), and one Tesla P100 graphics processing unit (GPU) with 16 GB video RAM.

# 3    Results and Discussion

The resulting evaluation metrics, based on the hold-out test sample, are shown in Table 1. Different folds of the 5-fold cross-validation procedure generated models that performed better in different metrics. The metric Recall (TPR) varies widely between the models trained for each fold. Fold 5 achieved 71.62%, however, its Precision (PPV) is the lowest among all folds. However, Precision was higher on the fold 4 model, which had a poor outcome on the Recall metric. The Specificity (TNR) also exhibits significant variance, highlighting the ensemble's balancing impact. Accuracy has higher for the fold 3, although the accuracy for the bagged model was higher. The AUC showed satisfactory values for all folds, and the ensemble AUC was higher than the folds 1 and 3 AUC by 0.80%. Regarding the performance of the models for the validation set, the Accuracy values were $65.49\% \pm 2.13$ (mean $\pm$ standard deviation) and AUC values were $0.66 \pm 0.03$ (mean $\pm$ standard deviation) across folds.

The ensemble successfully managed to increase the overall metrics of the model and decrease the variation observed for different folds and models.

**Table 1.** Resulting metrics for the best model of each k-fold and the final ensemble model. Ensemble metrics were calculated based on the average of the predicted probabilities for the hold-out test set.

| Fold | Accuracy | Precision | Specificity | Recall | AUC |
|------|----------|-----------|-------------|--------|-----|
| 1 | 64.41% | 58.21% | 72.82% | 52.70% | **0.71** |
| 2 | 65.54% | 57.83% | 54.37% | 64.86% | 0.67 |
| 3 | **66.10%** | 58.97% | **80.58%** | 62.16% | **0.71** |
| 4 | 64.97% | **61.11%** | 77.67% | 44.59% | 0.69 |
| 5 | 63.84% | 55.21% | 68.93% | **71.62%** | 0.70 |
| Ens. | 68.93% | 63.77% | 75.73% | 59.46% | 0.72 |

Figure 3 shows the Receiver Operating Characteristic (ROC) curve for the resulting ensemble model. The value of AUC = 0.72 shows that the amplitude of the results of the model is satisfactory and the symmetrical shape of the ROC curve shows that there is a reasonable balance in the classification of both output classes. Individual models showed similar AUC values. Still, the ensemble model achieved the highest value, demonstrating the benefits of this strategy.

The results of the present paper can be discussed in comparison with those of other authors who used the ABIDE I dataset under similar conditions, *i.e.* accounting for samples from multiple sites, gender, and ages (Table 2). The accuracy achieved was slightly below the ones obtained by [8,11,12,16], and above the ones achieved by [37,38]. Precision (PPV), Specificity (TNR), Recall (TPR), and AUC for the presented model are compatible with those found by previous authors. Specificity found in the present analysis is higher than in most of the
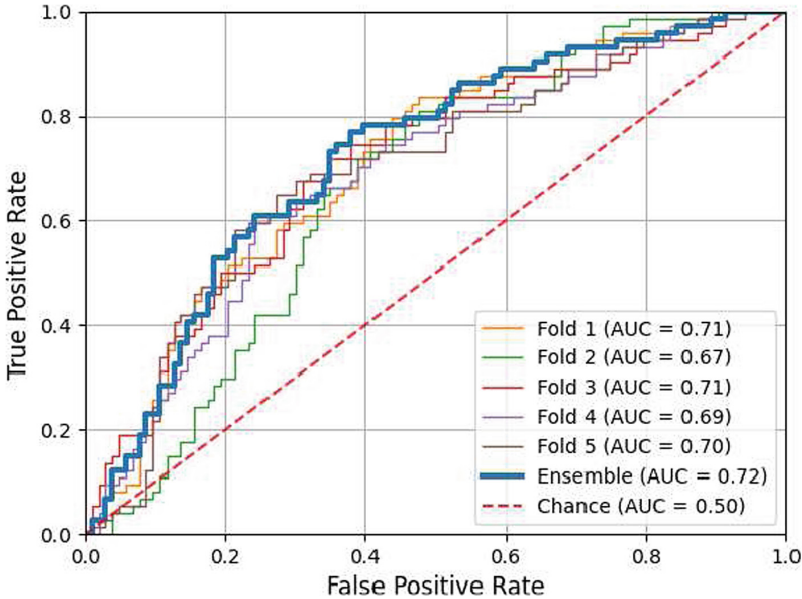
**Fig. 3.** ROC curve and AUC value for the best model of each k-fold and the final ensemble model. Curve $y = x$ (AUC = 0.5) shown for reference.

previous papers, indicating that samples indicated as negative by our model have a high potential to be of a typical developed individual. Unfortunately, not all metrics were reported by previous authors, therefore limiting the scope of our comparisons. The only metric present in all studies was accuracy.

Despite not achieving the highest overall accuracy, our SqueezeNet-based 3D-CNN is still able to outperform another 3D-CNN model based on the same dataset [38], which achieved 58% accuracy. For reference, [17] also used a 3D-CNN model and achieved an accuracy of 64% for the combined ABIDE I and ABIDE II dataset. Nevertheless, the metrics from [17] are not comparable to either [38] or those from the present study because they are based on a different dataset. Adding ABIDE II data possibly introduces more variability to the test data. However, it provides a higher number of samples to be trained on, which was observed to affect CNN models positively. The fact that we use multi-center samples, and from all age groups and genders available in ABIDE I, may contribute to a better generalization considering other settings. By all means, the resulting metrics shown are well within the expected range and show that our model was successfully developed and trained.

The most common strategy for feature selection is to calculate a correlation matrix of functional connectivity of ROIs (Table 2). We used the computed ReHo derivative without the need to select any ROIs. ReHo alterations are already reported in individuals with ASD [22,39] and our model was able

to autonomously extract features from a 3D volume of ReHo values to classify ASD.

Although positive, the results are yet distant from being replicated as diagnostic biomarkers. In fact, rs-fMRI-based classifications of ASD are generally less accurate than classifications based on behavioral observation [40]. However, the advancement of Machine Learning models may, in the future, complement the diagnosis or add a vital diagnosis screening layer.

## 4    Conclusion

We proposed a SqueezeNet-based 3D-CNN ensemble model as an alternative for ASD classification using fMRI data. Although the metrics calculated for the results obtained are slightly lower than state-of-the-art studies, our development still poses an interesting implementation because of its relative simplicity. Lightweight network architectures display a good trade-off between accuracy and the number of parameters in 2D cases. We developed a model capable of processing complex 3D medical imaging data using this type of resource-efficient architecture. Moreover, we achieved results comparable to benchmarks and even surpassed the performance of other 3D-CNN architectures using the ABIDE I dataset. Lastly, we used ReHo values as input to our network, demonstrating an alternative measure to autonomously detect ASD and someday potentially aid in patient screening and diagnosis.

**Table 2.** Comparison table for automated detection of TDC and ASD classes using samples from multiple sites, gender, and ages of the ABIDE I dataset.

| Ref. | Model | acc (%) | PPV (%) | TNR (%) | TPR (%) | AUC | Features |
|---|---|---|---|---|---|---|---|
| [8] | DNN | 70.00 | | 63.00 | 74.00 | | Functional connectivity of ROIs |
| [37] | SSAE | 62.00 | | 71.20 | 53.30 | 0.65 | Age at testing, sex, handedness, full-scale IQ, verbal IQ, performance IQ |
| [38] | 3D-CNN C-LSTM | 58.00 | | | | | 4D Spatio-Temporal fMRI |
| [11] | DNN | 75.27 | 76.88 | | 74.00 | | Functional connectivity of ROIs |
| [16] | 2D-CNN | 70.22 | | 61.82 | 77.46 | 0.70 | Functional connectivity of ROIs |
| [12] | ANN | 71.35 | | 80.60 | 59.50 | | Harmonized functional connectivity of ROIs |
| This study | 3D-Squeeze-Net CNN | 68.93 | 63.77 | 75.73 | 59.46 | 0.72 | Regional homogeneity |

$ANN$: artificial neural network, $CNN$: convolutional neural network, $C - LSTM$: convolutional long short-term memory, $DNN$: deep neural network, $ROIs$: regions of interest, $SSAE$: stacked sparse auto-encoder based neural network.

# References

1. American Psychiatric Association: Diagnostic and statistical manual of mental disorders. Am. Psychiatr. Assoc. (2013). https://doi.org/10.1176/appi.books.9780890425596

2. Centers for Disease Control and Prevention (CDC): Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, united states, 2010. Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002) **63**(2), 1–21 (2014)

3. Hayes, J., Ford, T., Rafeeque, H., Russell, G.: Clinical practice guidelines for diagnosis of autism spectrum disorder in adults and children in the UK: a narrative review. BMC Psychiatry **18**(1), 1–25 (2018)

4. Meszlényi, R.J., Buza, K., Vidnyánszky, Z.: Resting state FMRI functional connectivity-based classification using a convolutional neural network architecture. Front. Neuroinf. **11**, 61 (2017) https://doi.org/10.3389/fninf.2017.00061, https://www.frontiersin.org/article/10.3389/fninf.2017.00061

5. Courchesne, E., et al.: Unusual brain growth patterns in early life in patients with autistic disorder: an MRI study. Neurology **57**(2), 245–254 (2001)

6. Paakki, J.J., et al.: Alterations in regional homogeneity of resting-state brain activity in autism spectrum disorders. Brain Res. **1321**, 169–179 (2010)

7. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatry **19**(6), 659–667 (2014)

8. Heinsfeld, A.S., Franco, A.R., Craddock, R.C., Buchweitz, A., Meneguzzi, F.: Identification of autism spectrum disorder using deep learning and the abide dataset. NeuroImage Clin. **17**, 16–23 (2018)

9. Nielsen, J.A., et al.: Multisite functional connectivity MRI classification of autism: abide results. Front. Hum. Neurosci. **7**, 599 (2013)

10. Abraham, A., et al.: Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. Neuroimage **147**, 736–745 (2017)

11. Yang, X., Schrader, P.T., Zhang, N.: A deep neural network study of the abide repository on autism spectrum classification. Int. J. Adv. Comput. Sci. Appl. (2020)

12. Ingalhalikar, M., Shinde, S., Karmarkar, A., Rajan, A., Rangaprakash, D., Deshpande, G.: Functional connectivity-based prediction of autism on site harmonized ABIDE dataset. IEEE Trans. Biomed. Eng. **68**(12), 3628–3637 (2021). https://doi.org/10.1109/TBME.2021.3080259. Epub 2021 Nov 19. PMID 33989150; PMCID PMC8696194

13. Kazeminejad, A., Sotero, R.C.: Topological properties of resting-state FMRI functional networks improve machine learning-based autism classification. Front. Neurosci. **12**, 1018 (2019). https://doi.org/10.3389/fnins.2018.01018, https://www.frontiersin.org/article/10.3389/fnins.2018.01018

14. Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D.: Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. Neuroimage **145**, 137–165 (2017)
15. Sarraf, S., Tofighi, G.: Deep learning-based pipeline to recognize alzheimer's disease using FMRI data. In: 2016 Future Technologies Conference (FTC), pp. 816–820 (2016). https://doi.org/10.1109/FTC.2016.7821697
16. Sherkatghanad, Z., et al.: Automated detection of autism spectrum disorder using a convolutional neural network. Front. Neurosci. **13**, 1325 (2020)
17. Thomas, R.M., Gallo, S., Cerliani, L., Zhutovsky, P., El-Gazzar, A., van Wingen, G.: Classifying autism spectrum disorder using the temporal statistics of resting-state functional MRI data with 3d convolutional neural networks. Front. Psych. **11**, 440 (2020)
18. Kopuklu, O., Kose, N., Gunduz, A., Rigoll, G.: Resource efficient 3d convolutional neural networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (2019)
19. Craddock, C., et al.: The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. Front. Neuroinf. **7**, 5 (2013)
20. Zang, Y., Jiang, T., Lu, Y., He, Y., Tian, L.: Regional homogeneity approach to FMRI data analysis. Neuroimage **22**(1), 394–400 (2004)
21. Agarwal, S., Sair, H.I., Pillai, J.J.: The resting-state functional magnetic resonance imaging regional homogeneity metrics-kendall's coefficient of concordance-regional homogeneity and coherence-regional homogeneity-are valid indicators of tumor-related neurovascular uncoupling. Brain connectivity **7**(4), 228–235 (2017)
22. Vigneshwaran, S., Mahanand, B.S., Suresh, S., Sundararajan, N.: Using regional homogeneity from functional MRI for diagnosis of ASD among males. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2015). https://doi.org/10.1109/IJCNN.2015.7280562
23. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet: alexnet-level accuracy with 50x fewer parameters and ¡0.5mb model size (2016)
24. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
25. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pp. 315–323. JMLR Workshop and Conference Proceedings (2011)
26. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
27. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification (2015)
28. Refaeilzadeh, P., Tang, L., Liu, H.: Cross-Validation, pp. 532–538. Springer US, Boston, MA (2009). https://doi.org/10.1007/978-0-387-39940-9_565
29. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2017)
30. Grandvalet, Y.: Bagging equalizes influence. Mach. Learn. **55**(3), 251–270 (2004)
31. Buja, A., Stuetzle, W.: Observations on bagging. Statistica Sinica **16**, 323–351 (2006)
32. Freeman, E.A., Moisen, G.G.: A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecol. Model. **217**(1–2), 48–58 (2008)

33. de Oliveira, G.G., Ruiz, L.F.C., Guasselli, L.A., Haetinger, C.: Random forest and artificial neural networks in landslide susceptibility modeling: a case study of the fão river basin, southern brazil. Nat. Hazards **99**(2), 1049–1073 (2019)

34. Lucchese, L.V., de Oliveira, G.G., Pedrollo, O.C.: Attribute selection using correlations and principal components for artificial neural networks employment for landslide susceptibility assessment. Environ. Monit. Assess. **192**(2), 1–22 (2020)

35. Lucchese, L.V., de Oliveira, G.G., Pedrollo, O.C.: Mamdani fuzzy inference systems and artificial neural networks for landslide susceptibility mapping. Nat. Hazards **106**(3), 2381–2405 (2021)

36. DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics **44**, 837–845 (1988)

37. Parikh, M.N., Li, H., He, L.: Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data. Front. Comput. Neurosci. **13**, 9 (2019)

38. El-Gazzar, A., Quaak, M., Cerliani, L., Bloem, P., van Wingen, G., Mani Thomas, R.: A hybrid 3DCNN and 3DC-LSTM based model for 4D spatio-temporal fMRI data: an ABIDE autism classification study. In: Zhou, L., et al. (eds.) OR 2.0/MLCN -2019. LNCS, vol. 11796, pp. 95–102. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32695-1_11

39. Shukla, D.K., Keehn, B., Müller, R.A.: Regional homogeneity of FMRI time series in autism spectrum disorders. Neurosci. Lett. **476**(1), 46–51 (2010)

40. Plitt, M., Barnes, K.A., Martin, A.: Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. NeuroImage: Clin. **7** (2015). https://doi.org/10.1016/j.nicl.2014.12.013, https://www.sciencedirect.com/science/article/pii/S2213158214001983

# Performance Evaluation of the Cepstral Method to Estimate the Stable Optimal Solution of Feedforward Occlusion Cancellation in the Presence of Noise

Bruno C. Bispo[1]([✉]) and Renata C. Borges[2]

[1] Department of Electrical and Electronic Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, SC, Brazil
bruno.bispo@ufsc.br

[2] Department of Electrical Engineering, Federal University of Technology - Paraná (UTFPR), Cornélio Procópio, PR, Brazil

**Abstract.** This work aims to evaluate the performance of the cepstral method to estimate, in the presence of noise, the stable optimal solution of feedforward occlusion cancellation. The estimation is performed using the feedback structure for occlusion cancellation in a calibration process that may occur before using the hearing aid or when the user finds it convenient. Real-world occlusion path, acoustic path, and speech signals were employed. Simulations showed that the efficiency of the method reduces when decreasing the signal-to-noise ratio (SNR) at the hearing aid external microfone. For an analysis window of $2\,\mathrm{s}$, estimates with mean normalized misalignment (MIS) less than $-6$, $-14$ and $-22\,\mathrm{dB}$ are obtained for SNRs of 30, 40 and $50\,\mathrm{dB}$, respectively. For a window of $4\,\mathrm{s}$, mean MIS less than $-16$ and $-24\,\mathrm{dB}$ are achieved for SNRs of 40 and $50\,\mathrm{dB}$, respectively. The results indicate that the calibration process needs to be carried out in an acoustically controlled environment to maximize the accuracy of the stable solution estimate.

**Keywords:** Occlusion Effect · Hearing Aid · Cepstral Method · Optimal Solution · Ambient Noise

## 1 Introduction

The occlusion effect in hearing aids occurs when the ventilation opening has insufficient diameter to provide the necessary dissipation of sound energy conducted to the ear canal through the skull and jaw, when the hearing aid user speaks [1,2]. This leads to an increased low-frequency sound pressure level, thereby making the user listen to his own muffled voice [3]. The cause of the occlusion effect is modeled by the impulse response $o(n)$ in Figs. 1 and 2.

The voice $v(n)$ uttered by the hearing aid user is picked up by the hearing aid's external microphone, after traversing the acoustic path represented by the
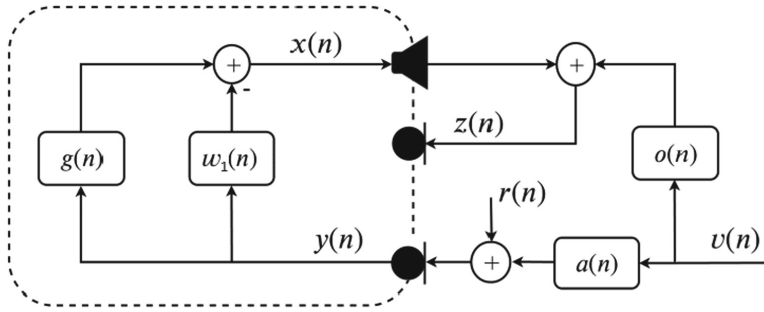
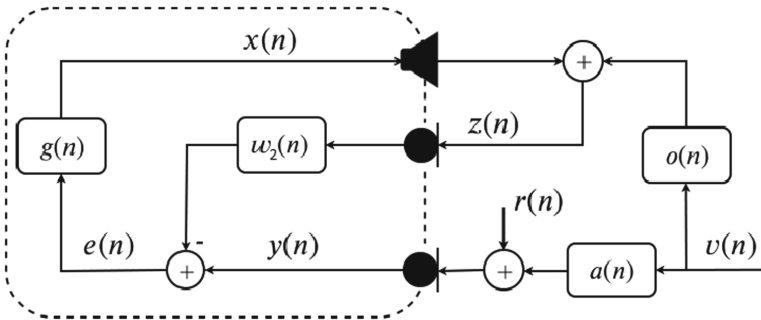**Fig. 1.** Feedforward structure for occlusion effect cancellation.



**Fig. 2.** Feedback structure for occlusion effect cancellation.

impulse response $a(n)$, along with the ambient noise $r(n)$, generating the signal $y(n)$. Disregarding the filter $w_1(n)$, the signal of the external microphone, $y(n)$, is amplified by the compensation system re-presented by the impulse response $g(n)$, resulting in the signal $x(n)$ to be played back by the hearing aid's loudspeaker. The signal $z(n)$, to be picked up by a possible internal microphone of the device, is in fact the signal to be heard by the hearing aid user and is defined as

$$
\begin{aligned}
z(n) &= x(n) + v(n) * o(n) \\
&= [g(n) * a(n) + o(n)] * v(n) + g(n) * r(n),
\end{aligned}
\tag{1}
$$

where the symbol $*$ denotes the convolution operation. Therefore, the occlusion effect is characterized by the addition of $o(n) * v(n)$ to the desired value of $z(n)$ and, thus, can be interpreted from the reverberation point of view.

Fixed and adaptive controllers have been proposed to, at least, attenuate the occlusion effect. The fixed solutions, proposed in [1,4–7], ensure the system stability but do not deal with the dynamic changes of the acoustic system, and may suffer performance losses due to variations in the ear channel or displacement of the ear mold. On the other hand, the adaptive solutions proposed to date in [4,5] present slow convergence of the adaptive filter coefficients and require constant

adaptation since the occlusion effect occurs in short time periods, when sound signals are produced by the user.

Both solutions can be implemented in feedforward or feedback structures, which are shown in Figs. 1 and 2 respectively. Note that, in the feedback structure, the hearing aid has an internal microphone to pick up and utilize the signal $z(n)$.

Among the mentioned cancellation proposals, the work presented in [7] stands out for being the only one to estimate the optimal solution for feedforward occlusion cancellation. The estimation is performed using a cepstral method and the feedback occlusion cancellation structure in a calibration process, when hearing loss compensation is not necessary and $g(n)$ and $w_2(n)$ can be freely chosen, provided that they do not result in an uncomfortable acoustic environment. However, the estimation method was evaluated considering only the absence of ambient noise in the calibration, a situation that may not be found even in controlled conditions.

This work aims to evaluate the performance of the cepstral method for estimating the stable optimal solution of occlusion cancellation in the presence of noise. This article is organized as follows: Sect. 2 discusses the optimal solution $w_o(n)$ for feedforward occlusion cancellation; in Sect. 3, the cepstral method for estimating $w_o(n)$ is described; in Sect. 4, the configurations of the performed simulations are presented; Sect. 5 presents and discusses the results obtained; and, finally, Sect. 6 concludes the work.

## 2   Feedforward Occlusion Cancellation System

In the feedforward occlusion cancellation system, illustrated in Fig. 1, the optimal frequency response, in the sense of totally removing the signal $o(n) * v(n)$ from $z(n)$, in the absence of ambient noise ($r(n) = 0$) is given by [7]

$$W_o(e^{j\omega}) = \frac{O(e^{j\omega})}{A(e^{j\omega})}, \tag{2}$$

which in the time domain corresponds to

$$w_o(n) = o(n) * a_I(n), \tag{3}$$

where $a_I(n)$ denotes the impulse response of the inverse system to the acoustic path. Due to the properties of the discrete-time Fourier transform, $w_o(n)$ is absolutely summable and therefore this optimal solution is stable.

As the acoustic path models the propagation delay from the user's mouth to the external microphone of the hearing aid, its impulse response can be written as [7]

$$a(n) = \tilde{a}(n) * \delta(n - N_a), \tag{4}$$

where $\delta(n)$ is the unit impulse function, $\tilde{a}(n) = 0$ for $n < 0$, $\tilde{a}(0) \neq 0$ and $N_a > 0$. Consequently, as demonstrated in [7], $a_I(n)$ is in general a two-sided

signal composed of left-side increasing exponentials for $n < -N_a$ and right-side decreasing exponentials for $n \geq -N_a$. But, since its energy is concentrated around $n = -N_a$, $a_I(n)$ can be considered of finite length with $a_I(n) \neq 0$ only for $A_1 \leq n \leq A_2$, where $A_1 < -N_a < A_2$. Note that $A_1$ is always negative.

Combining the above approximation of $a_I(n)$ with the fact that $o(n) \neq 0$ only for $n = 0, 1, \ldots, M - 1$, as shown in Sect. 5, $w_o(n) \neq 0$ only for $n = A_1, A_1 + 1, \ldots, M + A_2 - 1$ [7]. Hence, the stable optimal solution is non-causal.

## 3    Cepstral Method to Estimate $w_o(n)$

In the feedback cancellation system represented in Fig. 2, it can be shown that the discrete-time Fourier transform (DTFT) of the error signal $e(n)$ is given by

$$
\begin{aligned}
E(e^{j\omega}) = &\frac{1 - W_o(e^{j\omega})W_2(e^{j\omega})}{1 + G(e^{j\omega})W_2(e^{j\omega})}Y(e^{j\omega}) \\
&+ \frac{W_o(e^{j\omega})W_2(e^{j\omega})}{1 + G(e^{j\omega})W_2(e^{j\omega})}R(e^{j\omega}).
\end{aligned}
\tag{5}
$$

In the absence of ambient noise, $R(e^{j\omega}) = 0$ and (5) becomes

$$
E(e^{j\omega}) = \frac{1 - W_o(e^{j\omega})W_2(e^{j\omega})}{1 + G(e^{j\omega})W_2(e^{j\omega})}Y(e^{j\omega}).
\tag{6}
$$

In this ideal case, if $\left|W_o(e^{j\omega})W_2(e^{j\omega})\right| < 1$ and $\left|G(e^{j\omega})W_2(e^{j\omega})\right| < 1$, necessary conditions for Taylor series expansions, then the cepstrum of $e(n)$ can be defined from (6) as [7]

$$
\begin{aligned}
c_e(n) = c_y(n) &- \sum_{k=1}^{\infty} \frac{[w_o(n) * w_2(n)]^{*k}}{k} \\
&- \sum_{k=1}^{\infty} (-1)^{k+1} \frac{[g(n) * w_2(n)]^{*k}}{k},
\end{aligned}
\tag{7}
$$

where $\{\cdot\}^{*k}$ denotes the $k$th convolution power.

The cepstral method proposed in [7] explores (7) to estimate $w_o(n)$, in a fashion similar to [8,9] for acoustic feedback cancellation. The estimation is carried out in a calibration process that occurs before using the hearing aid or when the user finds it convenient. The signal $v(n)$ must be a voiced sound, usually a vowel, emitted by the hearing aid user.

The reasoning behind the method comes from speculating that $w_o(n) * w_2(n)$, the impulse response for $k = 1$ of the first time series in (7), can be extracted from $c_e(n) - c_y(n)$ by properly choosing $g(n)$ and $w_2(n)$. At the calibration process, hearing loss compensation is not required and these impulse responses can be arbitrarily chosen, as long as they do not cause acoustic disturbance to the hearing aid user.

The choice of $g(n)$ and $w_2(n)$ plays a key role as it serves three purposes [7]: ensure that the conditions for Taylor series expansions are met and, therefore, equation (7) is valid; make $w_o(n) * w_2(n)$ causal; and make the non-zero samples of $w_o(n) * w_2(n)$ do not overlap with the non-zero samples of the impulse responses for $k > 1$ of the first time series in (7).

The impulse responses of the compensation and the feedback paths are simply defined as a bandwidth gain and a delay [7], that is,

$$g(n) = k_g \delta(n - N_g) \tag{8}$$

and

$$w_2(n) = k_w \delta(n - N_w), \tag{9}$$

where $N_g > 0$ and $N_w > 0$.

The conditions for Taylor series expansions are met by choosing $k_w$ and $k_g$ such that

$$|k_w| < \max_{\omega} \frac{1}{|W_o(e^{j\omega})|} \tag{10}$$

and

$$|k_g| < \frac{1}{|k_w|}. \tag{11}$$

Causality and non-overlapping of $w_o(n) * w_2(n)$ are achieved by choosing $N_w$ according to [7]

$$N_w \geq M + A_2 - 2A_1. \tag{12}$$

Specified the parameters of $g(n)$ and $w_2(n)$ according to the above discussion, the method starts by computing the real cepstra $c_e(n)$ and $c_y(n)$ from the signals $e(n)$ and $y(n)$, respectively, through the fast Fourier transform (FFT). Then, the method gets $\{w_2(n) * [g(n) - w_o(n)]\}\widehat{\phantom{x}}$, an estimate of $w_2(n) * [g(n) - w_o(n)]$ which is the resulting impulse response for $k = 1$ in (7), by selecting the first $M + A_2 + N_w - 1$ samples of $c_e(n) - c_y(n)$.

In the sequel, the method computes $\{g(n) - w_o(n)\}\widehat{\phantom{x}}$, an estimate of $g(n) - w_o(n)$, as follows [7]

$$\{g(n) - w_o(n)\}\widehat{\phantom{x}} = \{w_2(n) * [g(n) - w_o(n)]\}\widehat{\phantom{x}} * w_I(n), \tag{13}$$

where $w_I(n) = 1/k_w \delta(n + N_w)$ represents the impulse response of the inverse system to $w_2(n)$ and is known. Note that convolution with $w_I(n)$ consists of a sliding on the time axis and a multiplication.

Finally, the method computes $\hat{w}_o(n)$, an estimate of the stable optimal solution $w_o(n)$ for the feedforward occlusion cancellation, as [7]

$$\hat{w}_o(n) = - \{g(n) - w_o(n)\}\widehat{\phantom{x}} + g(n). \tag{14}$$

## 4    Simulation Configuration

This section describes the configuration of the experiment carried out in a simulated environment to evaluate the performance of the cepstral method in estimating $w_o(n)$ in the presence of noise.

## 4.1   Speech Database

The database consists of 20 recordings (12 male and 8 female) of the sustained vowel /a/ sampled at a frequency of 22050 Hz. The vowel /a/ was chosen due to its wide use in the acoustic analysis of voice. It was provided by the Medical Engineering Research Group of the National Council for Scientific and Technological Development (GPEM/CNPq). The recordings were performed at the Hospital das Clínicas, Faculty of Medicine, University of São Paulo (HC-FMUSP), approved by the Human Research Ethics Committee of the Federal University of São Carlos under the protocol number 256/2010.

In this work, the signals were resampled to 16 kHz and their active power levels were normalized to −26 dBov through the ITU-T Recommendation P.56 algorithm [10]. Longer speech signals were created by concatenating each signal with itself. Variable length segments of each resulting signal were used as $v(n)$ in order to evaluate the performance of the cepstral method as a function of the vowel length uttered by the hearing aid user.

The speech signals were additively contaminated with zero-mean white noise at the following signal-to-noise ratio levels SNR $= \{\infty, 50, 45, 40, 35, 30, 25, 20\}$ dB.

## 4.2   System Configuration

**Acoustic Path.** The acoustic path was represented in two ways. First, as in [4, 7,11], the acoustic path was a time delay defined as

$$a(n) = \delta(n - N_a), \tag{15}$$

where $N_a = 14$. This delay refers to a propagation length of 30.29 cm, assuming an average length of 15 cm between the glottis and the lips [12] and a distance of 15.29 cm between the lips and the external microphone of hearing aids [4]. In this case, the condition (10) becomes $|k_w| < 0.41$.

Closer to a real-world situation, the second acoustic path was modeled by a room impulse response available in [13]. The sampling frequency was reduced to 16 kHz and the first 17 samples were discarded to simulate the typical 14-sample delay from the lips to the hearing aid external microphone. Then it was truncated for computational cost reasons. The impulse response $a(n)$ and the frequency response magnitude of the second acoustic path are shown in Fig. 3. In this situation, the condition (10) becomes $|k_w| < 0.14$.
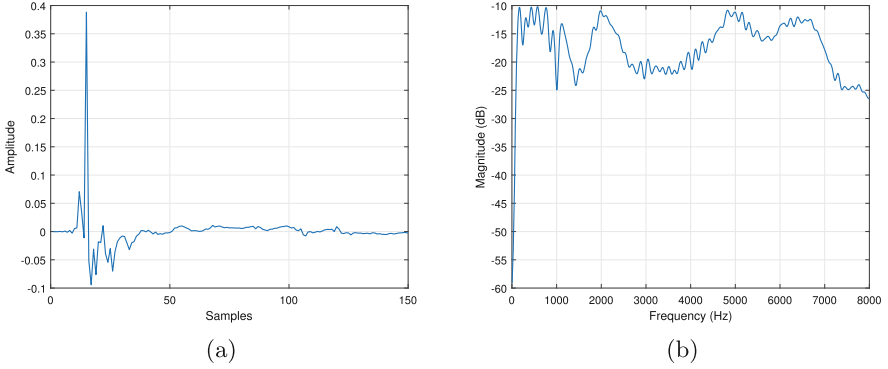
(a)                                    (b)

**Fig. 3.** Second acoustic path: (a) $a(n)$; (b) $|A(e^{j\omega})|$.
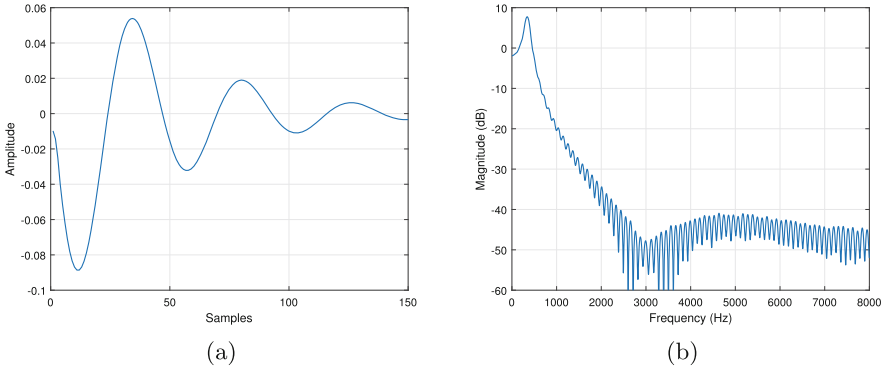




(a)                                    (b)

**Fig. 4.** Occlusion path: (a) $o(n)$; (b) $|O(e^{j\omega})|$.

**Occlusion Path.** The occlusion path was modeled by the impulse response available in [4], which was measured in a volunteer with a custom and non-ventilated ear mold and digitally recorded at a sampling rate of 16 kHz. The impulse response $o(n)$ and the frequency response magnitude of the occlusion path are shown in Fig. 4, where it is noted that $M = 150$.

**Forward Path.** As in [4,7,11], the forward path was modelled as

$$g(n) = \delta(n-1), \tag{16}$$

that is, $N_g = k_g = 1$. It is emphasized that, as discussed in Sect. 3, the forward path does not need to compensate the hearing aids user's loss during the calibration process. And, as discussed in Sect. 4, the optimal stable solution for feedforward occlusion cancellation is independent of $g(n)$.

### 4.3   Misalignment

The estimate of the stable optimal solution obtained by the cepstral method was evaluated through the normalized misalignment (MIS), which is defined as

$$\text{MIS} = \frac{\left\{\sum_n [w_o(n) - \hat{w}_o(n)]^2\right\}^{1/2}}{\left\{\sum_n w_o^2(n)\right\}^{1/2}}. \tag{17}$$

## 5   Simulation Results

The performance of the cepstral method was evaluated for several signal-to-noise ratios (SNR) in the hearing aid external microphone, namely, SNR = $\{\infty, 50, 45, 40, 35, 30, 25, 20\}$ dB, and for different sizes of the analysis window that is used to calculate the short-time cepstra. The Hann window was used.

### 5.1   Scenario 1

The impulse response $a(n)$ of the first acoustic path is defined in (15). Consequently, the impulse responses of its stable inverse system and the stable optimal solution are defined as $a_I(n) = \delta(n + 14)$ and $w_o(n) = o(n + 14)$, respectively, where the non-causality of the stable optimal solution is verified.

The cepstral method was configured as follows: $A_1 = -100$, $A_2 = 100$, $M = 150$, $k_w = 0{,}1$, $N_w = 450$, $k_g = 1$ and $N_g = 1$. The mean MIS obtained for various speech lengths and various SNRs are shown in Table 1. Examples of $\hat{w}_o(n)$, which have MIS close to the mean value, obtained by the method with windows of 2 and 4 s and SNRs of 30 and 40 dB are illustrated in Fig. 5.

The results show that the performance of the method improves both with the increase in the analysis window size and with the increase in the SNR. For a window of 2 s, a considerably short time for a person to sustain a vowel sound, SNRs equal to 30, 40 and 50 dB are needed so that, on average, MIS is lower than $-12$, $-20$ and $-25$ dB, respectively.

As initially demonstrated and discussed in [7], the performance improvement with window enlargement is due to the increase in the accuracy of the definition of $c_e(n)$ according to (7), caused by the reduction of the truncation effect of $e(n)$, which theoretically has infinite duration, necessary to compute $c_e(n)$. The performance improvement with the increase in SNR is also due to the increase in the accuracy of (7), but now caused by the reduction of the effect of the second term on the right-hand side of (5), which tends to zero as SNR tends to infinity.

### 5.2   Scenario 2

The impulse response $a(n)$ of the second acoustic path is represented in Fig. 3. The resulting $w_o(n)$ is shown in Fig. 6, where the non-causality of the stable

**Table 1.** Mean of the normalized misalignment in the first scenario.

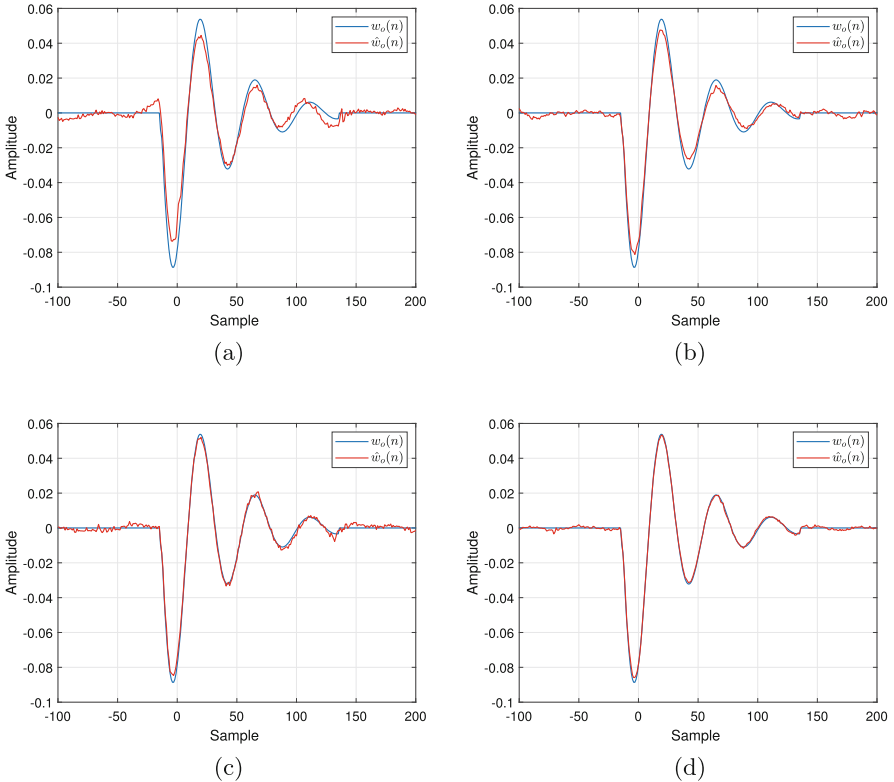| SNR | Window length (ms) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 500 | 1000 | 1500 | 2000 | 3000 | 4000 | 5000 |
| $\infty$ | -10.06 | $-18.21$ | $-20.69$ | $-25.99$ | $-27.89$ | $-29.51$ | $-35.31$ |
| 50 | $-10.00$ | $-17.47$ | $-20.65$ | $-25.23$ | $-27.31$ | $-29.82$ | $-33.89$ |
| 45 | $-9.69$ | $-16.38$ | $-20.36$ | $-24.15$ | $-26.61$ | $-29.20$ | $-32.21$ |
| 40 | $-8.94$ | $-13.57$ | $-18.96$ | $-21.96$ | $-24.28$ | $-27.04$ | $-29.03$ |
| 35 | $-7.24$ | $-10.32$ | $-15.90$ | $-17.54$ | $-18.83$ | $-22.25$ | $-24.01$ |
| 30 | $-4.97$ | $-6.71$ | $-11.05$ | $-12.21$ | $-13.45$ | $-16.64$ | $-17.78$ |
| 25 | $-3.67$ | $-3.93$ | $-6.99$ | $-7.62$ | $-7.76$ | $-10.48$ | $-11.30$ |
| 20 | $-2.23$ | $-2.45$ | $-3.76$ | $-4.32$ | $-4.32$ | $-6.22$ | $-7.06$ |



**Fig. 5.** Estimates of $w_o(n)$ in the first scenario for the following window sizes and SNRs: (a) 2 s, 30 dB; (b) 2 s, 40 dB; (c) 4 s, 30 dB; (d) 4 s, 40 dB.

**Table 2.** Mean of the normalized misalignment in the second scenario.

| SNR | Window length (ms) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 500 | 1000 | 1500 | 2000 | 3000 | 4000 | 5000 |
| $\infty$ | −5.90 | −16.33 | −19.16 | −25.42 | −26.38 | −28.04 | −33.96 |
| 50 | −5.65 | −14.07 | −17.22 | −22.07 | −23.59 | −24.67 | −26.99 |
| 45 | −5.32 | −11.86 | −15.67 | −18.46 | −19.47 | −20.36 | −20.99 |
| 40 | −4.89 | −9.29 | −12.59 | −14.08 | −15.86 | −16.89 | −17.02 |
| 35 | −3.75 | −6.70 | −9.34 | −10.14 | −10.80 | −11.55 | −11.98 |
| 30 | −2.74 | −4.09 | −6.04 | −6.61 | −6.78 | −7.83 | −8.23 |
| 25 | −1.84 | −2.73 | −3.91 | −4.20 | −4.16 | −5.25 | −5.70 |
| 20 | −1.04 | −1.58 | −2.17 | −2.36 | −2.39 | −3.13 | −3.43 |

optimal solution is evident. It is important to emphasize that, in this scenario, $w_o(n)$ is an infinite impulse response.

The cepstral method was configured as follows: $A_1 = -300$, $A_2 = 100$, $M = 150$, $k_w = 0,1$, $N_w = 850$, $k_g = 1$ and $N_g = 1$. The mean MIS obtained for various speech lengths and various SNRs are shown in Table 2. Examples of $\hat{w}_o(n)$ obtained by the method for speech signals with durations of 2 and 4 s and SNRs of 30 and 40 dB are illustrated in Fig. 6.

For the same reasons explained in the first scenario, the method performance improves with the increase in both the analysis window length and the SNR. However, it is observed that the average results are lower than those obtained in the first scenario. This is due to the combination of two factors: the infinite length of $w_o(n)$ inevitably causes overlap of the impulse responses present in $c_e(n) - c_u(n)$, thus impairing their estimation [7]; for the same size of the signal analysis window, the inevitable truncation of $e(n)$ to compute $c_e(n)$ can have an effect on the inaccuracy of (7) greater than in the first scenario since $w_o(n)$ is different.

The results presented in this work show that the performance differences become more significant with the decrease in SNR. For a window of 2 s, SNRs equal to 30, 40 and 50 dB are needed so that, on average, MIS is less than −6, −14 and −22 dB, respectively. In order to obtain mean MIS less than −10 and −20 dB, SNRs equal to or greater than 35 and 45 dB are required, respectively.
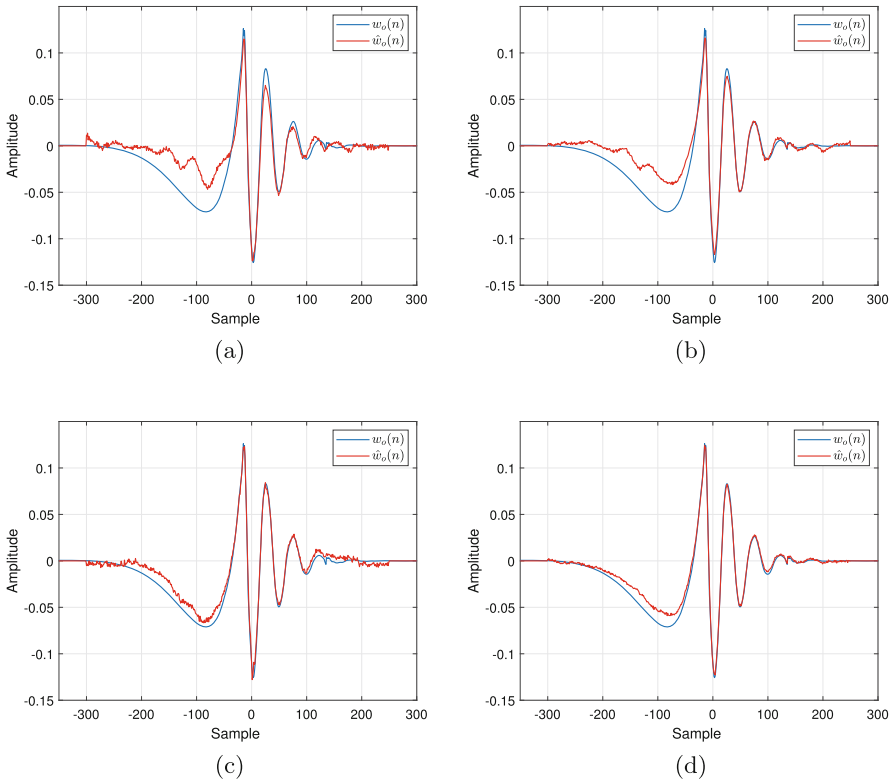
**Fig. 6.** Estimates of $w_o(n)$ in the second scenario for the following window sizes and SNRs: (a) 2 s, 30 dB; (b) 2 s, 40 dB; (c) 4 s, 30 dB; (d) 4 s, 40 dB.

## 6   Conclusions

This work evaluated the performance of the cepstral method to estimate, in the presence of noise, the stable optimal solution of feedforward occlusion cancellation in hearing aids. The estimation is performed in a calibration process carried out in a controlled environment. The method was originally evaluated considering the absence of ambient noise.

Simulations showed that the efficiency of the method reduces with the decrease in SNR at the hearing aid external microphone. For analysis windows of 2 s, estimates with mean MIS less than $-6$, $-14$ and $-22$ dB are obtained for SNRs of 30, 40 and 50 dB, respectively. SNRs equal to or greater than 35 and 45 dB are required to obtain mean MIS less than $-10$ and $-20$ dB, respectively. For windows of 4 s, estimates with mean MIS less than $-16$ and $-24$ dB are achieved for SNRs of 40 and 50 dB, respectively.

These results indicate that the calibration process, where the cepstral method is used, needs to be carried out in an acoustically controlled environment to

maximize the accuracy of the optimal solution estimate, which can be used in implementing a controller to reduce the occlusion effect.

# References

1. Mejia, J., Dillon, H., Fisher, M.: Active cancellation of occlusion: an electronic vent for hearing aids and hearing protectors. J. Acoust. Soc. Am. **124**(1), 235–240 (2008). https://doi.org/10.1121/1.2908279

2. Chung, K.: Challenges and recent developments in hearing aids: Part II. Feedback and occlusion effect reduction strategies, laser shell manufacturing processes, and other signal processing technologies. Trends Amplif. **8**(4), 125–164 (2004). https://doi.org/10.1177/108471380400800402

3. MacKenzie, D.J.: Open-canal fittings and the hearing aid occlusion effect. Hearing J. **59**(11), 50–52 (2006). https://doi.org/10.1097/01.HJ.0000286218.69092.dd

4. Borges, R.C., Costa, M.H.: A feed forward adaptive canceller to reduce the occlusion effect in hearing aids. Comput. Biol. Med. **79**, 266–275 (2016). https://doi.org/10.1016/j.compbiomed.2016.10.016

5. Borges, R.C., Costa, M.H., Cordioli, J.A., Assuiti, L.F.C.: An adaptive occlusion canceller for hearing aids. In: Proceedings of 21st European Signal Processing Conference, pp. 1–5. Marrakech, Morocco (2013)

6. Sunohara, M., Osawa, M., Hashiura, T., Tateno, M.: Occlusion reduction system for hearing aids with an improved transducer and an associated algorithm. In: Proceedings of 23rd European Signal Processing Conference, pp. 285–289 (2015). https://doi.org/10.1250/ast.35.318

7. Bispo, B.C., Borges, R.C.: A cepstral method to estimate the stable optimal solution for feedforward occlusion cancellation in hearing aids. J. Commun. Inf. Syst. **35**(1), 113–123 (2020). https://doi.org/10.14209/jcis.2020.12

8. Bispo, B.C., Rodrigues, P.M., Freitas, D.R.S.: Acoustic feedback cancellation based on cepstral analysis. In: Proceedings of 17th IEEE Conference on Signal Processing Algorithms, Architectures, Arrangements and Applications, pp. 205–209. Poznan, Poland (2013)

9. Bispo, B.C., Freitas, D.R.S.: On the use of cepstral analysis in acoustic feedback cancellation. Digit. Signal Proc. **44**, 88–101 (2015)

10. ITU-T P.56: P.56: Objective measurement of active speech level (2011)

11. Bispo, B.C., Borges, R.C., Costa, M.H.: Método cepstral para estimação da resposta ao impulso do efeito de oclusão em sistemas de auxílio á audição. In: Proceedings of the XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais. Petrópolis, Brazil (2019)

12. Fant, G.: A note on vocal tract size factors and non-uniform f-pattern scalings. Speech Transm. Lab. Q. Progr. Status Rep. **1**, 22–30 (1966)

13. ITU-T G.191: Software tools for speech and audio coding standardization. International Telecommunications Union (2016)

# Evaluating Semantic Segmentation of Tuberculosis Bacilli in Bright Field Microscopy Using Different Color Spaces Components and Mosaic Images

M. K. Serrão ⬛, I. M. Saldanha ⬛, M. G. F. Costa ⬛, and C. F. F. Costa Filho$^{(\boxtimes)}$ ⬛

Federal University of Amazonas/ Center for Research and Development
in Electronic and Information Technology – CETELI, Manaus, Brazil
`ccosta@ufam.edu.br`

**Abstract.** Tuberculosis is one of the infectious diseases that causes the most victims in the world. The early diagnosis of the disease is fundamental for the treatment to be carried out quickly, saving lives and decreasing the number of infected people. The tuberculosis diagnosis is a time-consuming process that involves the analysis of up to 100 fields of conventional microscopy, which causes fatigue for the technician involved in the process. In order to assist the technician in counting bacilli in tuberculosis bright field microscopy examinations, automatic methods using deep neural networks are proposed. In this work, we evaluate the use of a semantic segmentation network for bacilli detection, using as input several color components from the RGB, HSV, YCbCr, and Lab spaces. The model that presented the best performance uses only the components of RGB color space as input to the network, with values of accuracy, precision, sensitivity, specificity, and f1-score above 99%.

**Keywords:** Tuberculosis · Convolutional Neural Networks · Bacilli Detection · Color Spaces

## 1 Introduction

Tuberculosis (TB) is caused by the bacteria Mycobacterium tuberculosis and most commonly affects the lungs. According to the World Health Organization (WHO), about 9.9 million people got sick with TB in 2020, and before the covid-19 pandemic, TB was the leading cause of death from a single infectious agent worldwide [1].

For more than 100 years the bright field smear technique has been used in TB diagnosis and, because of its availability and simplicity, it remains the main technique in many settings with a high incidence of the disease [2]. In high-income countries, fluorescence microscopy is more widely used for diagnosing TB, but in middle- and low-income countries, where most TB cases occur, the standard diagnostic test is bright field smear microscopy, which uses a conventional microscope [3]. Bright field smear microscopy consists of preparing smears of sputum samples on slides stained with the

Ziehl-Neelsen (ZN) or Kinyoun staining methods. This process gives the TB bacilli a pink coloration and the other elements of the sample a blue coloration. A technical expert then uses a conventional microscope to view these slides and count the number of bacilli. The diagnosis is made ac-cording to the number of bacilli counted by the specialist.

Many works have been developed in the field of computer vision to identify bacilli in bright field sputum smear images. A pattern recognition technique that has been prominent in recent years for TB bacilli segmentation is the convolutional neural network (CNN) [4–9].

In [4], the authors used a CNN to identify malaria-causing pathogens in lung secretion smear samples, and intestinal parasites in stool samples. The input image of the network, in the RGB space, was a patch, i.e. a smaller image, with the size determined by the type of pathogen to be recognized in each case. In [5], the authors also use a CNN for bacilli segmentation. The input to the network was a 40x40 pixels patch and the network was evaluated with RGB images, gray-level images, and with the levels corresponding to R-G values, which corresponds to the difference between the R and G components of the RGB space, and intensity images, obtained from the RGB im-age. The networks with RGB and R-G input showed better results in comparison with the intensity image. Similarly, in [6] the authors used transfer learning, through a pre-trained model with the CIFAR-10 database, and trained the last dense layers for bacilli recognition. The network was trained with RGB patches of 32x32 pixels. In [9], the authors used a CNN to perform the bacilli segmentation task. The input RGB image has 400x400 pixels and is composed of 100 patches of 40 x 40 pixels, that can be positive or negative. Three CNN architectures were evaluated.

It can be observed that most of the cited works evaluate CNNs with the components of the RGB color space as inputs. This work proposes to evaluate a CNN with different inputs. The goal is to evaluate whether different combinations of the RGB, HSV, Lab, and YCbCr color spaces can improve the performance of the CNN in the detection of bacilli in bright field microscopy images.

The RGB color space was used in this work because it is the original space in which the images are represented, and was used in the other studies published in the literature. The HSV color space was chosen because it has characteristics similar to the human visual system, in which intensity information (H) is uncoupled from color information (S and H). To justify the choice of the YCbCr and Lab spaces, we will use the conclusion obtained in [5], about the importance of the R-G difference in the detection of bacilli. In the YCbCr space, the Cb component, where the R component is subtracted from the G component, would help the bacilli detection, while the Cr component would help the background detection. The inclusion of the Lab space was done to check the hypothesis established in [5], through the negation of the same, as in this space the R and G components are summed in the a component. Therefore, it is expected that this system does not present a good performance.

The following color space are combined in the CNN input: RGB; RGB and HSV; RGB, HSV, and Lab; RGB, HSV, and YCbCr; RGB, HSV, YCbCr, and Lab.

## 2   Materials and Methods

### 2.1   Materials

The database used in this work was provided by the Pattern Recognition and Optimization Research Group of UFAM [10]. This database consists of 8000 mosaic images of 400 × 400 pixels and their corresponding gold standard images. The mosaic images are formed from the combination of 100 positive and negative 40 × 40 pixels patches. Positive patches contain a centralized bacillus, and negative patches contain no bacillus. All these patches were extracted from sputum smear images stained with the Kinyoun technique and had the classification performed by an expert. Figure 1 shows an example of an RGB mosaic image and its gold standard image.
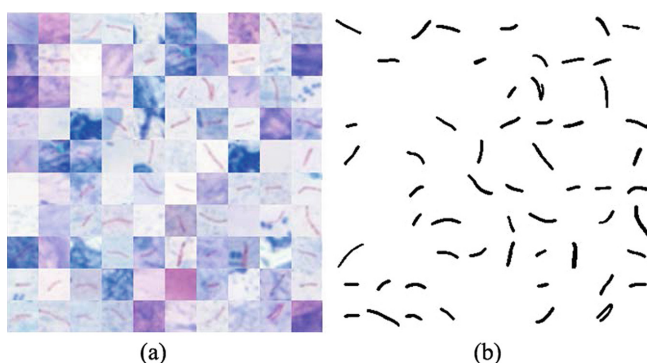


(a)                                              (b)

**Fig. 1.**  (a) Mosaico image in RGB space and (b) gold standard image.

The database was divided into training, validation, and test sets. For the training set, 6000 mosaic images were used, of which 3000 are composed of approximately 50% positive and 50% negative patches, 1500 are composed of approximately 2% positive and 98% negative patches, and 1500 are composed of only negative patches. For the validation set, 1000 mosaic images consisting of approximately 50% positive and 50% negative patches were used. For the test set, 1000 mosaic images were also used, composed of approximately 50% positive and 50% negative patches.

### 2.2   Semantic Segmentation Network Architecture

Convolutional neural networks are primarily used for image processing, performing both image classification and semantic segmentation tasks. The classification task aims to classify an image into categories, while semantic segmentation aims to extract regions of interest from the image. In [11], the authors investigated the ability of a convolutional neural network pre-trained with nonmedical images to classify pathologies in x-ray images. The network performs an important screening task by categorizing patients as healthy or having chest pathology that includes the classes: right pleural effusion, cardiomegaly, and abnormal mediastinum. In [12], the authors used a convolutional neural

network, through the U-net architecture, to perform the task of semantic segmentation of brain tumors in magnetic resonance images.

The main layers of a convolutional network are the convolution layer, which has the function of extracting features, and the subsampling layer, which aims to re-duce the dimensionality of the activation planes of the network [13]. The output of classification networks is a vector whose coordinate values indicate the probability that an image belongs to a particular class. In semantic image segmentation tasks, the network classifies each pixel of the image as belonging or not belonging to a region of interest. For such networks, the output is also an image with the same dimensions as the input and transposed convolution layers are responsible for recovering an image in the output with the same dimensions as the input image.

In [9] three different architectures were tested to per-form semantic segmentation of tuberculosis bacilli in mosaic images. The architecture that achieved the best performance was the one that used the largest number of layers. The architecture used in this work was based on the best-performing architecture presented in [9].

The architecture used in this work, shown in Fig. 2, is composed of an input layer, 4 subsampling stages, a dropout layer, 4 oversampling stages, and the output layers. The subsampling stages are composed of sequences of convolution layers with a 3x3 filter, batch normalization layer, ReLU activation function, and max-pooling subsampling layer with a $2 \times 2$ filter. The over-sampling steps are composed of transposed convolution layer sequences with a 4x4 filter, batch normalization layer, and ReLU activation function. Finally, there is a convolution layer with a 1x1 size kernel, softmax activation function, and a pixel classification layer, which classifies each pixel as belonging to a bacillus or to the background.

In order to evaluate the influence of different color space components on the performance of the semantic segmentation network of TB bacilli, input tensors with different color space components were used. Five models with different color spaces were evaluated to train the network. The list of these models are shown in Fig. 3.

Each evaluated model has an input tensor in the for-mat: rows x columns x channels. The first model was trained with the components of the RGB color space and has the input tensor with dimensions $400 \times 400 \times 3$. The second model was trained with a combination of the coordinates from the RGB and HSV color spaces and has the input tensor with dimensions $400 \times 400 \times 6$. The third model was formed by a combination of the coordinates of the RGB, HSV, and Lab color spaces, and has the in-put tensor with dimensions $400 \times 400 \times 9$. The fourth model was formed by a combination of the coordinates from the RGB, HSV, and YCbCr color spaces, and has the input tensor with dimensions $400 \times 400 \times 9$. The last model was trained was formed by a combination of the coordinates of the RGB, HSV, YCbCr, and Lab color spaces, and has the input tensor with dimensions $400 \times 400 \times 12$.

### 2.3   Training Parameters

The ADAM optimization method was used to train the network. Table 1 shows the parameters used for training. The training was performed with MATLAB version 2020a, a 3.2 GHz Intel i7-8700 processor with 16 GB of RAM, and a GeForce GTX 1070 GPU with 8 GB of RAM.

**Fig. 2.** CNN architecture used for semantic segmentation of TB bacilli.

### 2.4 Bacilli Classification

The CNN performs the classification of each pixel as bacillus or background. A set of 8-connected pixels, identified as bacilli, forms objects that can be bacilli or not. To classify these objects as belonging to bacillus or not, a post-processing step is performed on the binary images obtained from the output of the semantic segmentation network, as shown in the block diagram in Fig. 4. This figure shows a $40 \times 40$ pixels sliding window algorithm that runs through the entire image to check for 8-connected objects. Each iteration of the algorithm checks for 8-connected pixels in a sub-region of the image. Objects that have an area above the experimentally obtained threshold of 25 pixels are classified as bacilli. Otherwise, the object is disregarded. This process is carried out until the output image is fully evaluated.

### 2.5 Evaluation Metrics

The following metrics were used to evaluate the performance of the models: accuracy, precision, sensitivity, specificity, and F1- score. These metrics are shown in Eqs. (1) to (5). A true positive (TP) is when the semantic segmentation network detected a bacillus in a patch that contain a bacillus. A true negative (TN) is when the semantic segmentation network did not detect a bacillus in a patch that do not contain a bacillus. A false positive (FP) result is when the semantic segmentation network detected a bacillus in a patch that
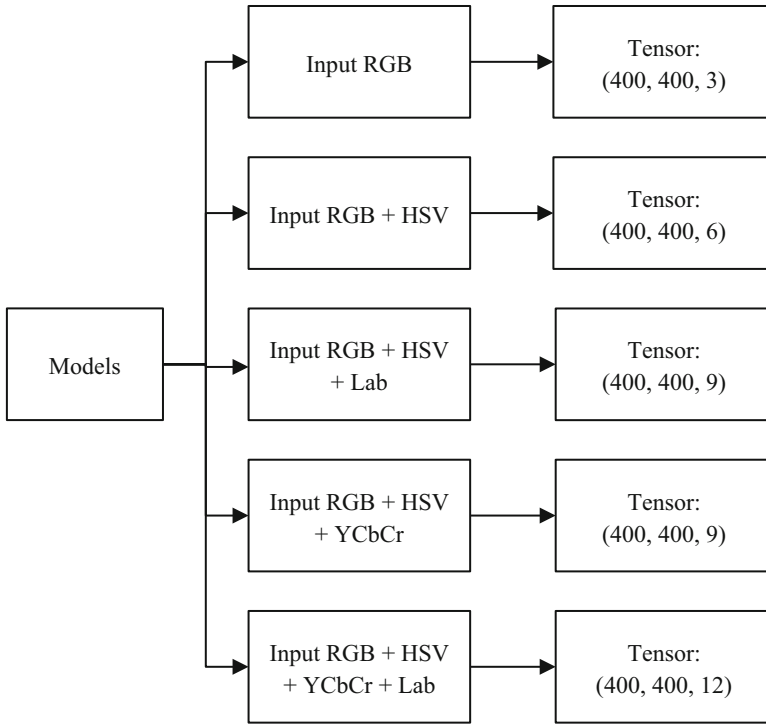
**Fig. 3.** 5 Inputs with different color spaces were used in the 5 semantic bacilli segmentation models.

**Table 1.** Parameters used in training the semantic segmentation network

| Parameters | Values |
|---|---|
| Learning rate | 0,001 |
| Learning drop factor | 0,05 |
| Maximum number of epochs | 20 |
| Batch size | 10 |
| Stop condition | 6000 |

do not contain a bacillus. A false negative (FN) result is when the semantic segmentation network did not detect a bacillus in a patch that contain a bacillus.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (1)$$

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$
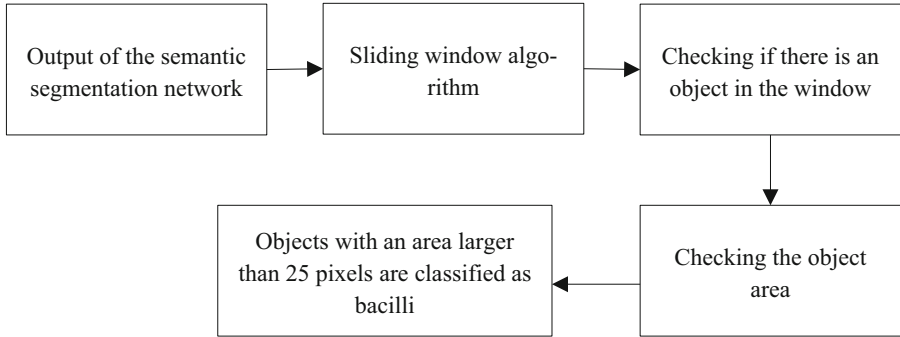
**Fig. 4.** Post-processing applied in binary images obtained in semantic segmentation network output, for bacilli detection.

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

$$F1 - score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \tag{5}$$

## 3   Results

Table 2 shows the results of the metrics for each of the models evaluated for the semantic segmentation of bacilli. From this table, one can conclude that, in general, the color plane that presented the best performance was the RGB, achieving an accuracy of 99.43%. This result is 2.53% better than the worst performing model, the one which used the color space components of RGB + HSV + YCbCr + Lab at inputs. The latter achieved an accuracy of 96.90%. The precision of the model that used the components of the color space RGB as inputs was 99.62%. This result is 5.11% better than the model that used the RGB + HSV + YCbCr + Lab color space com-ponents as inputs. Considering the sensitivity, the model with the components of the RGB + HSV + YCbCr + Lab color spaces as inputs showed a good performance, 99.54%. The specificity and F1-score values of the model with the components of RGB color space as inputs, were 99.63% and 99.43%, respectively.

Furthermore, it is possible to identify those combinations of the RGB, HSV, and YCbCr color spaces, in general, showed good results, with metric values above 99%.

Another observation is related to the Lab color space. In the models where this color space was used, the network presented a worse performance.

From these results, we conclude that the use of more color components at the network input does not contribute to the improvement of network performance. The model with the least amount of color components at the input, the one that used the RGB space, presented a much better performance than the models that used a larger amount of color

components. Using the components of the RGB color space as inputs of a convolutional network is sufficient to perform the TB bacilli seg-mentation task with high performance.

Figure 5 presents the confusion matrix of the test set for the RGB model. This model correctly predicted 49,631 bacilli, for a total of 50,015 bacilli.

The CNN network takes a few seconds to analyze a mosaic image, which is incomparably faster than the time taken by a technician to manually counting bacilli for a diagnosis. Thus, using the method proposed in this study greatly speeds up the microscopic examination of slides.

In [8], using the faster region-based convolutional neural network (RCNN) and the ZNSM-iDB database for bacilli detection, the authors achieved a recall of 98.4%, a precision of 85.1% and an F1-score of 91.2%. In [6], using CNN for patch classification from a database collected from the Department of Pathology, Peking University First Hospital, the authors achieved a sensitivity of 97.94% and a specificity of 83.65%. In the present work, using CNN for bacilli detection, an accuracy of 99.43%, a precision of 99.62%, a sensitivity of 99.23%, a specificity of 99.63% specificity, and an F1-score of 99.43%.

**Table 2.** Performance of the 5 models were evaluated for semantic segmentation of bacilli

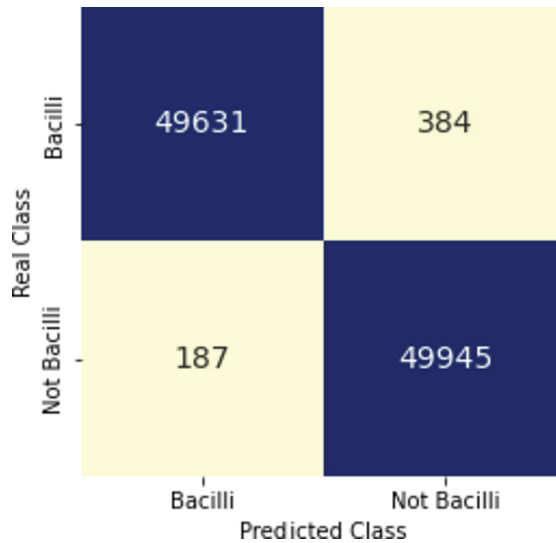| Models | Accuracy (%) | Precision (%) | Sensitivity (%) | Specificity (%) | F1-score (%) |
|---|---|---|---|---|---|
| RGB | 99.43 | 99.62 | 99.23 | 99.63 | 99.43 |
| RGB + HSV | 99.11 | 99.12 | 99.09 | 99.13 | 99.10 |
| RGB + HSV + Lab | 98.79 | 98.70 | 98.89 | 98.70 | 98.79 |
| RGB + HSV + YCbCr | 99.30 | 99.50 | 99.70 | 99.51 | 99.30 |
| RGB + HSV + YCbCr + Lab | 96.90 | 94.51 | 94.51 | 94.29 | 96.96 |

**Fig. 5.** Confusion matrix showing the performance on the test set of the model that uses the RGB color space components as input.

## 4 Conclusion

In this work, the performance of a convolutional neural network was evaluated to perform the semantic seg-mentation of TB bacilli, using different color spaces at the network input. The RGB model with input tensor $400 \times 400 \times 3$, was the one that presented the best performance for the considered metrics, reaching performance higher than 99% for all of them, while the model with components of the RGB + HSV + YCbCr + Lab color spaces as inputs, presented a lower performance compared to the other evaluated models.

The main conclusion is that the RGB color space alone is sufficient to perform the TB bacilli segmentation task with high performance. With a larger number of color space components in the input, satisfactory results are achieved, but lower than the results obtained using only the RGB space color. The use of the Lab color space compromises network performance. We believe that the main reason for this is because in the Lab color space, the R and G components are grouped into a single component, *a*.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. World Health Organization: Global Tuberculosis Report World Health Organization (2021)
2. Harries, A., Kumar, A.: Challenges and progress with diagnosing pulmonary tuberculosis in low- and middle-income countries. Diagnostics **8**(4), 78 (2018). https://doi.org/10.3390/diagnostics8040078
3. Steingart, K.R., Henry, M., et al.: Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. Lancet Infect. Dis. **6**(9), 570–581 (2006). https://doi.org/10.1016/S1473-3099(06)70578-3
4. Quinn, J.A., Nakasi, R., et al.: Deep convolutional neural networks for microscopy-based point of care diagnostics. In: International Conference on Machine Learning for Health Care Proceedings, vol. 56, pp 1–12 (2016). http://arxiv.org/abs/1608.02989
5. López, Y.P., Costa Filho, C.F.F., et al.: Automatic classification of light field smear microscopy patches using Convolutional Neural Networks for identifying Mycobacterium Tuber-culosis. In: IEEE CHILEAN CONFERENCE on Electrical, Electronics Engineering, Information and Communication Technologies, pp. 1–5 (2017). https://doi.org/10.1109/CHILECON.2017.8229512
6. Xiong, X.Y., Hou, B., et al.: Automatic detection of myco-bacterium tuberculosis using artificial intelligence. J. Thorac. Dis. **10**(3), 1936–1940 (2018). https://doi.org/10.21037/jtd.2018.01.91
7. Dinesh, R., Saumel, J., et al.: Tuberculosis (TB) detection system using deep neural networks. Neural Comput. Appl. **31**(5), 1533–1545 (2019). https://doi.org/10.1007/s00521-018-3564-4
8. El-Melegy, M., Mohamed, D., et al.: Identification of tuberculosis bacilli in ZN-stained sputum smear images: a deep learning approach. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2019-June, pp. 1131–1137 (2019). https://doi.org/10.1109/CVPRW.2019.00147
9. Serrão, M., Costa, M., et al.: Automatic Bacillus detection in light field microscopy images using convolutional neural networks and mosaic imaging approach. In: International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1903–1906 (2020). https://doi.org/10.1109/EMBC44109.2020.9176105
10. Pattern Recognition and Optimization Research Group (2008). http://dgp.cnpq.br/dgp/espelhogrupo/4186
11. Bar, Y., Diamant, I., et al.: Chest pathology detection using deep learning with non-medical training. In: Proceedings IEEE International Symposium on Biomedical Imaging, vol. 2015–July, pp. 294–297 (2015). https://doi.org/10.1109/ISBI.2015.7163871
12. El Kaitouni, S.E.I., Tairi, H.: Segmentation of medical im-ages for the extraction of brain tumors: a comparative study between the Hidden Markov and Deep Learning approaches. In: International Conference on Intelligent Systems and Computer Vision Proceedings, pp. 1–5 (2020). https://doi.org/10.1109/ISCV49265.2020.9204319
13. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, London (2017)

# Comparison Between Features Extracted in the Time and Frequency Domain with the Triangulation Method in the Recognition of Activities of Human Movements

L. C. Giacomossi[1,2(✉)] and S. F. Pichorim[2]

[1] Department of Informatics, Federal Institute of Paraná, Curitiba, Brazil
`luiz.giacomossi@ifpr.edu.br`
[2] Graduate School of Electrical Engineering and Computer Science (CPGEI-DAELN), Federal University of Technology Paraná (UTFPR), Curitiba, Brazil
`pichorim@utfpr.edu.br`

**Abstract.** In this research, classical features of the literature in the time and frequency domains were compared with the features based on the triangulation technique, which was applied to the input signals of human movements divided into categories (walking on a treadmill, running on a treadmill, walking in a circle, erasing a blackboard, going up and down stairs and hand tremors). Signals are captured with the internal accelerometer of a smartphone, using embedded software, which reads 3 orthogonal axes. The comparison was performed only at the highest percentage value in the rate, and the analysis of statistical significance was not performed. The triangulation method uses Euclidean concepts and basic statistics to extract features based on straight lines, angles, areas, perimeter, derivative and triangle counter in the composition of patterns. The results are promising, in view of the high accuracy rates in the final classification of the categories of movements object of study, with values similar to those obtained in the classification of patterns composed by the classic features used in the literature, with the MLP (Multilayer Perceptron) classifiers and KNN ($K$-nearest neighbors) for $k = 5$. The highest average hit rates obtained for the experiment with 500 points were 99% with the literature methods, 97.9% for the triangulation method and in the experiment with 1000 points windows, the highest average hit rates were achieved with the triangulation method 99.3% and 98.7% for the methods found in the literature, both with the KNN classifier for $k = 5$.

**Keywords:** Triangulation Method · Time and Frequency Domains Features · Pattern Recognition · Human Movements

# 1  Introduction

Human movements are objects of study in the medical field in the monitoring of daily activities with the elderly [1, 2], monitoring of pathological movements such as tremors in Parkinson's disease or epilepsy [3–5] in sports to assess the performance of athletes [6], activity recognition in unstructured environments [7], in robotics [8], among others.

Pattern recognition techniques are used in several studies aimed at the recognition of human movements, as they allow the automation of procedures [1–7]. The standards must be composed of representative features to facilitate the classification or separation between the studied classes. This procedure begins with the extraction of features in the time and/or frequency domains over input data windows (categories or object classes of study), for the composition of patterns, which are calculated separately by movements' class. In the literature, time domain features, such as statistical calculations (average, sum, minimum, maximum, RMS value, standard deviation, variance, among others) [9, 10] or frequency domain, such as the Fast Fourier Transform (FFT) [4, 11], Wavelet [5, 12], are widely used in several studies. Subsequently, after completing the pre-processing or feature extraction phase, from the set of generated patterns, one third will be submitted to the training or machine learning phase and the other two thirds, that is, those that were not part of the training, must be submitted to final classification, with the use of classifier software by neural networks MLP (Multilayer Perceptron), KNN (*K*-Nearest Neighbors), SVM (Support Vector Machine), decision tree, among others [10].

In this research, the irregular input data in amplitude and frequency, captured with a smartphone (internal accelerometer), of human movements, were divided into six categories of daily activities and one simulating a pathological tremor (walking on a treadmill, running on a treadmill, walking in a circle, erasing a blackboard, going up and down stairs and hand tremors). One of the advantages of the triangulation method is that the feature extraction takes place simultaneously in the time and frequency domain, that is, the calculations are based on sequences of triangles, a reduction of the input data, where straight lines (sides of triangles) are calculated between the maximum (peaks) and minimum (valleys) amplitude values, as well as between two minima or two maxima, depending on the sequence of points at the beginning of an input data window. The main objective and motivation was the comparison between the features extracted by the triangulation method and the classic features of the literature in the domain of time and frequencies. The individual classification indices by movement, as well as the average indices between the object classes of study, were used as a measure of comparison in the final classification of the generated patterns (6 classes) [13].

# 2  Materials and Method

## 2.1  Materials

A smartphone model M2006C3LG (Android) was used to acquire the data. A program (software), developed for this research, was used for the acquisition of data in the 3 orthogonal axes (x, y and z) of the internal accelerometer of the smartphone.

## 2.2   Method

To make the comparisons between the features extracted by the triangulation method and the main methods found in the literature, a methodology was applied, which was divided into three phases: 1) acquisition of the human movement signal; 2) feature extraction by triangulation and feature extraction found in the literature in the time domain and in the frequency domain; and 3) classification and testing [13].

In this study, two algorithms were compared: one applying the triangulation method and the statistical mean (Algorithm 1) and the other applying the classic features of the literature in the time domain and frequency domain (Algorithm 2).

## 2.3   Data Acquisition

In order to carry out the experiments and acquire the data in this research (phase 1), a project with a detailed description of all the procedures was approved for the permanent committee of ethics in research involving human beings of the University of Maringá, State of Paraná, Brazil, protocol number 1717907. Human movement data were collected from 24 normal volunteers (12 males and 12 females), aged between 18 and 45 years, in the categories of walking on treadmill, running on treadmill, walking in circle, erasing blackboard, going up and down stairs, tremors, in the three orthogonal axes ($x$, $y$ and $z$) and recorded in a separate text file.

## 2.4   Feature Extraction

In this second phase of the research, the features were extracted by the triangulation method, applying the main methods found in the literature.

Triangulation method: Twelve features were calculated per accelerometer axis ($x$, $y$ and $z$), totaling 36 features. They are based on ascending side, descending side, base side, area, perimeter, ascending angle, descending angle, base angle, ascending derivative, descending derivative, base derivative and triangle counter. These features were extracted based on the triangulation technique that applies Euclidean concepts and statistical mean (other calculations can be applied, such as sum, standard deviation or variance). The triangulation technique reduces the points of the input data windows, after finding the peak (maximum) and valley (minimum) points, forming a new signal with sequences of triangles [13], as shown in Fig. 1.

The triangulation method was compared with the main features of the literature in the domain of time and frequency:

- Features in the time domain - statistical calculations (mean, maximum, minimum, area, variance, standard deviation, RMS value) and
- Features in the frequency domain - the calculations were based on the Fast Fourier Transform (FFT frequency value, FFT frequency vector peak voltage value, total power, average frequency and average power).

The time domain features were merged with the frequency domain features, forming 12 features per accelerometer axis ($x$, $y$ and $z$), totaling 36 features.

A last processing was performed, merging the 36 features calculated by the triangulation method (Algorithm 1), with the 36 features calculated in the time-frequency domain (Algorithm 2), forming a set with 72 features.
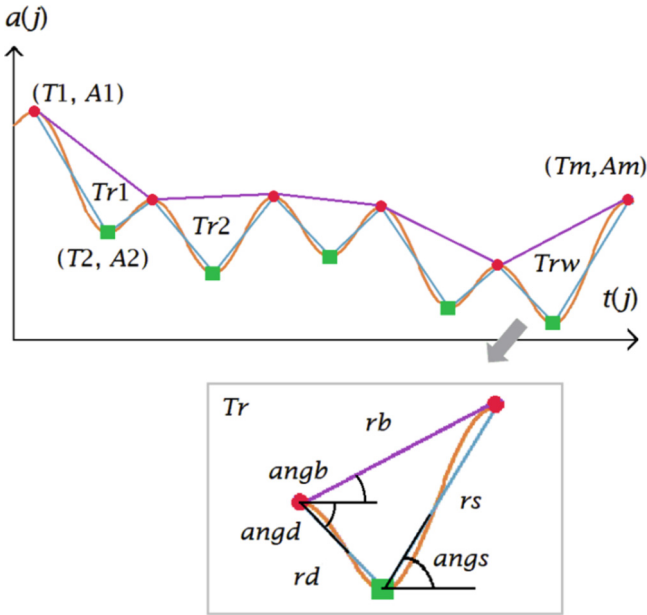


**Fig. 1.** A window of input data converted to a new signal generating a sequence of triangles. Original signal (in brown) with all peak (in red) and valley (in green) points. The calculated distances between the peak and valley points, generate the lines *rs*, *rd*, as well as the distances from the consecutive peak points *rb*, from the triangles *w* (*Tr1* to *Trw*).

## 3   Results

The third phase of this study corresponds to the final classification of the object of study movements. The individual and average classification indices using two classifiers, Multilayer Perceptron (MLP) and *K*-Nearest Neighbors for $k = 5$ (KNN), were used as a measure of comparison between algorithm 1 (triangulation method) and algorithm 2 (methods found in the literature). The patterns were calculated for the six input activities, using the two algorithms (algorithm 1 and algorithm 2), varying the size of the input window by 500 and 1000 points. Table 1 shows the number of discretized points read per axis of the accelerometer (*x*, *y* and *z*), according to the categories of movements object of study, as well as the number of patterns generated by applying algorithms 1 and algorithms 2.

Table 2 and Table 3 show the individual and average indices of final classification for the movements object of study, using MLP neural network and KNN for k = 5 classifier, respectively.

Figure 2 shows the results of the average final classification indexes for algorithm 1, algorithm 2, and merging of algorithms 1 and 2, using input data windows of 500 and 1000 points, with the MLP and KNN classifiers for k = 5. Considering the best average indices obtained with the KNN classifier, for windows of 500 points of input data, calculating the difference (99–97.9 = 1.1), it can be observed that the methods found in the literature in the time and frequency domains show a small advantage of 1.1%. Likewise, when merging the features of algorithms 1 and 2, the average hit rate obtained for windows of 500 points with algorithm 2 increased from 99% to 99.2%, improving by 0.2%. However, using the same classifier (where the best results were obtained), for windows of 1000 points, performing the same procedure (choosing the best average indices of each algorithm), calculating the difference (99.3–98.7 = 0.6), in this case, a small advantage of 0.6% was obtained for the triangulation method. By merging the features of algorithms 1 and 2, the average hit rate obtained for windows of 1000 points with algorithm 2 increased from 98.7% to 99.2%, improving by 0.5%, using the KNN classifier. When using the merge feature sets of algorithms 1 and 2, the best result was obtained with the MLP classifier for windows of 500 points, where the average hit rate of 98.2% obtained with algorithm 2 increased to 99.3%, that is, the set of features that apply triangulation (algorithm 1) contributed to improving the hit rate by 1.1%.
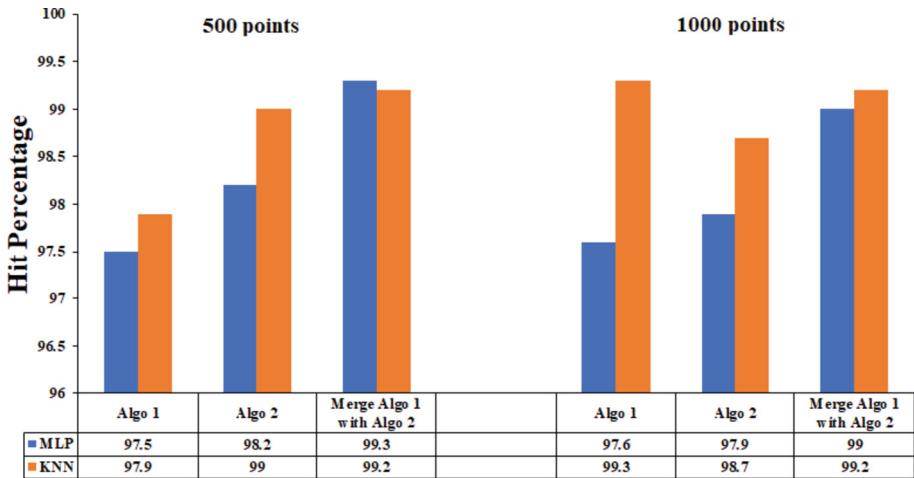


| | 500 points | | | | 1000 points | | |
|---|---|---|---|---|---|---|---|
| | Algo 1 | Algo 2 | Merge Algo 1 with Algo 2 | | Algo 1 | Algo 2 | Merge Algo 1 with Algo 2 |
| MLP | 97.5 | 98.2 | 99.3 | | 97.6 | 97.9 | 99 |
| KNN | 97.9 | 99 | 99.2 | | 99.3 | 98.7 | 99.2 |

**Fig. 2.** Hit rate in the final classification, comparing algorithm 1 (triangulation method) with algorithm 2 (methods found in the literature).

**Table 1.** Categories corresponding to the six movements object of study.

| Movements (class) | Number of points | Number of patterns per window | |
|---|---|---|---|
| | | 500 points | 1000 points |
| Walking on treadmill (1) | 160000 | 320 | 160 |
| Running on treadmill (2) | 147000 | 294 | 147 |
| Walking in circle (3) | 105000 | 210 | 105 |
| Erasing blackboard (4) | 105000 | 210 | 105 |
| Going up and down stairs (5) | 105000 | 210 | 105 |
| Tremors (6) | 120000 | 240 | 120 |
| Total | 742000 | 1484 | 742 |

**Table 2.** Hit rates in the final classification using MLP (Multilayer Perceptron) classifier.

| Window/ Algorithm | Class | | | | | | Avg |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 500 points | | | | | | | |
| Algo 1 | 98.4 | 97.3 | 92.9 | 100.0 | 96.2 | 100.0 | 97.5 |
| Algo 2 | 100.0 | 98.6 | 93.8 | 100.0 | 96.7 | 100.0 | 98.2 |
| Merging Algo 1 & 2 | 99.4 | 99.0 | 98.1 | 100.0 | 99.0 | 100.0 | 99.3 |
| 1000 points | | | | | | | |
| Algor 1 | 99.4 | 99.3 | 89.5 | 100.0 | 97.1 | 100.0 | 97.6 |
| Algo 2 | 100.0 | 100.0 | 93.3 | 100.0 | 94.3 | 100.0 | 97.9 |
| Merging Algo 1 & 2 | 99.4 | 100.0 | 96.2 | 100.0 | 98.1 | 100.0 | 99.0 |

Bland Altman's statistical analysis makes a comparison between two methods, that is, it is a simple way to assess the agreement between two different methods or instruments that must measure the same quantity. In Fig. 3, the individual indices per movement (objects of study) were used, for windows of 500 and 1000 points, with MLP and KNN classifiers for k = 5.

**Table 3.** Hit rates in the final classification using KNN ($K$-Nearest Neighbors) for $k = 5$ classifier.

| Window/ Algorithm | Class | | | | | | Avg |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 500 points | | | | | | | |
| Algo 1 | 99.7 | 98.0 | 97.1 | 99.0 | 93.3 | 100.0 | 97.9 |
| Algo 2 | 100.0 | 98.0 | 99.5 | 100.0 | 96.7 | 100.0 | 99.0 |
| Merging Algo 1 & 2 | 100.0 | 98.3 | 98.6 | 100.0 | 98.1 | 100.0 | 99.2 |
| 1000 points | | | | | | | |
| Algo 1 | 100.0 | 99.3 | 98.1 | 100.0 | 98.1 | 100.0 | 99.3 |
| Algo 2 | 100.0 | 100.0 | 99.0 | 100.0 | 93.3 | 100.0 | 98.7 |
| Merging Algo 1 & 2 | 100.0 | 100.0 | 98.1 | 100.0 | 97.1 | 100.0 | 99.2 |
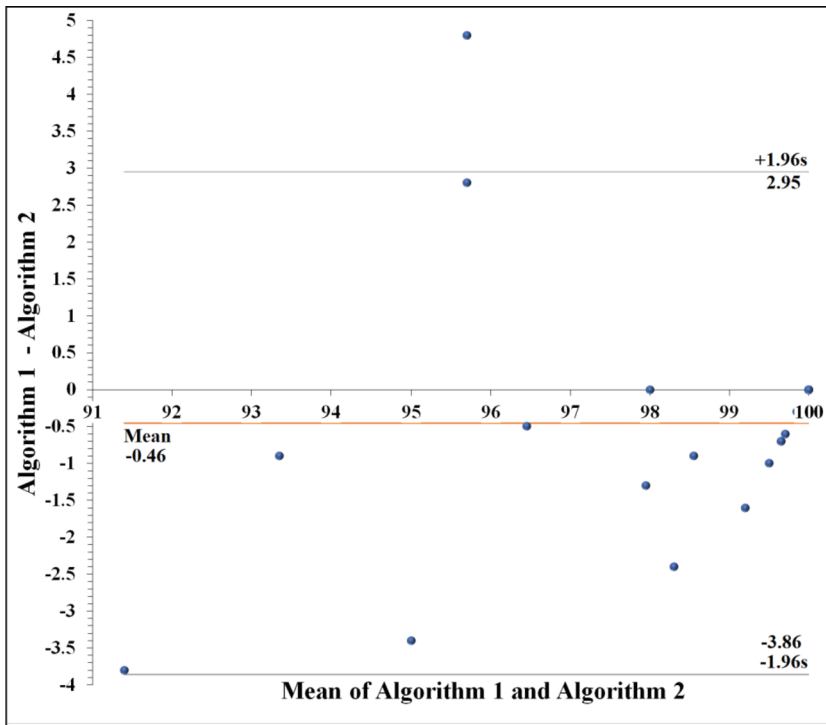


**Fig. 3.** Comparison between method 1 (triangles) and method 2 (literature).This analysis was positive, validating the triangulation method (algorithm 1), compared with the method that applies patterns composed of classic features from the literature extracted in the time and frequency domain (algorithm 2), considering that most points were within the bias (23 points or 95.83% agreement) and a single coordinate point (95.7; 4.8) was outside, that is, between the average differences and the agreement interval of 95% of the differences calculated between two methods, being eight coordinate points (100; 0) and two coordinate points (97.7; −0.7), coincident, from a total of 24 analyzed points.

## 4   Discussion

Analyzing the movement "walking on a treadmill" for windows of 500 input data points, a 99.7% hit rate was obtained with the triangulation method (algorithm 1) and 100% hit with the literature methods (algorithm 2), as well as for the movement "running on a treadmill", the same index of 98% was obtained with the two algorithms, triangulation method (algorithm 1) and methods found in the literature (algorithm 2), using the KNN classifier for k = 5, that is, in this case, there was a balance between the two methods. The same occurred for windows of 1000 input data points, with the movement "walking on a treadmill", where the two methods, algorithm 1 (triangles) and algorithm 2 (literature), reached a 100% success rate with KNN classifier for k = 5. Continuing the analysis for windows and 1000 points, was reached the hit rate of 100% in the classification of the movement "Running on the treadmill" for algorithm 2 (literature) and the index of 99.3% with algorithm 1 (triangles), using KNN classifier for k = 5, with small advantage for algorithm 2. These results are relevant, as the "walking" and "running" movements are objects of study in several studies, including the evaluation of athletes [6], in the medical area, in the monitoring of daily movements with the elderly [1, 2], among others.

As for the movement "erasing blackboard", both algorithms reached 100% of success rate, the same happening with the movement "tremors", both with KNN classifier for k = 5. It is important to note that in this study simulated hand tremors were used, even so, the results are relevant, as "tremors" in humans are studied in several studies, more specifically pathological tremors [3–5].

For the movement "walking in a circle" the best result was obtained with the methods that apply the features found in the literature (time and frequency domain) with 99% of success rate, while the triangulation method had an index slightly below with 98.1% hit rate. As for the "going up and down stairs" movement, the highest rate of 98.1% was obtained with the triangulation method, while the methods in the literature obtained a hit rate a little below 93.3%, in all cases using the classifier KNN for k = 5.

Making the asymptotic complexity analysis in a simplified way for the two algorithms, algorithm 1 (triangulation method), in the selection of maximum and minimum points, presents the degree of complexity is $O(N^2)$, where N corresponds to the size of the data windows and $O(f(N))$ is called asymptotic notation [14]. For the rest of the code, all calculations are based on straights, perimeter, angles, area, slope of the lines and statistical average. While for algorithm 2 (literature methods), the time domain features are based on statistical calculations such as mean, sum, maximum, minimum, RMS value, variance, standard deviation and the frequency domain features, all based on in the FFT, which has a degree of complexity $O(N.\log_2 N)$. In this case, an advantage for algorithm 2 (literature) was characterized, as the highest degree of complexity prevails in this analysis. However, the processing times of algorithm 1 and algorithm 2 were not measured and compared [14].

## 5   Conclusions

In this research, the classic features found in the literature in the time and frequency domains were compared with the features extracted by the triangulation method. The comparison was performed only at the highest percentage value in the rate, and the

analysis of statistical significance was not performed. One of the advantages of the triangulation method is that the feature extraction takes place simultaneously in the time and frequency domain, that is, the calculations are based on sequences of triangles, a reduction of the input data (data windows), where straight lines (sides of triangles) are calculated between the maximum (peaks) and minimum (valleys) amplitude values. The reduction of the signal to generate the minimum and maximum points is a necessary procedure to generate the triangle sequences, that is, it can be considered a disadvantage of the triangulation method (this processing time was not evaluated), since the methods found in the literature extract features directly from the input data windows. Features were extracted using two algorithms: triangulation method (algorithm 1) with 36 features and literature triangulation methods (algorithm 2) with 36 features, with 21 features in the time domain and 15 features in the frequency domain (merged features in the time and frequency domains). The discrete input signals captured with the internal accelerometer of a smartphone, irregular in amplitude and frequency, were used in two experiments, one with windows of 500 points and the other with windows of 1000 points, for the composition of patterns divided according to the classes of movements object of study. In the comparisons, individual and average indices were used in the final classification of some categories of daily human movements (walking and running on a treadmill, walking in a circle, erasing a blackboard, going up and down stairs and hand tremors), using two classifiers, MLP and KNN for k = 5. The results are promising with the use of the triangulation method, which uses Euclidean concepts and basic statistics to extract features based on straight lines, angles, areas, perimeter, derivative and triangle counter in the composition of patterns, in view of the high indices of correct answers in the final classification of the categories of movements object of study, with values similar to those obtained in the classification of patterns composed by the classic features used in the literature, with the MLP (Multilayer Perceptron) and KNN (K-nearest neighbors) for k = 5. The highest average hit rates obtained for the first experiment with 500 points were 99% with the literature methods, 97.9% for the triangulation method and in the second experiment with 1000 points windows, the highest average hit rates were achieved with the triangulation method 99.3% and 98.7% for the methods found in the literature, both with the KNN classifier for k = 5.

By merging the features of algorithms 1 and 2, the average hit rate obtained for windows of 1000 points with algorithm 2 increased from 98.7% to 99.2%, improving by 0.5%. When using the merge feature sets of algorithm 1 with algorithm 2, the best result was obtained with the MLP classifier for windows of 500 points, where the average hit rate of 98.2% obtained with algorithm 2 increased to 99.3%, that is, the set of features that apply triangulation (algorithm 1) contributed to improving the hit rate by 1.1%.

As for the proposed triangulation algorithm, in future studies, one should seek to reduce the degree of complexity in the procedure for obtaining the vector that stores the minimum and maximum points for the elaboration of sequences of triangles, from which all the features for the elaboration of the patterns. Likewise, the FFT, an optimized algorithm, derived from the Fourier Transform, which has been improved over time and has become one of the most effective algorithms, being used in several areas of science.

acquisition of the treadmill. This study was partially funded by the Brazilian research agency, CNPq.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. Khan, A.M., Lee, Y.K., Lee, S., et al.: Accelerometer's position independent physical activity recognition system for long-term activity monitoring in the elderly. Med. Biol. Eng. Comput. **48**(12), 1271–1279 (2010). https://doi.org/10.1007/s11517-010-0701-3
2. Straczkiewicz, M., James, P., Onnela, J.P.: A systematic review of smartphone-based human activity recognition methods for health research. Digital Med. **4**, 148 (2021). https://doi.org/10.1038/s41746-021-00514-4
3. Chronowski, M., Klaczynski, M., Dec-Cwiek, M., et al.: Speech and tremor tester – monitoring of neurodegenerative diseases using smartphone technology. Diagnostyca **21**(2), 31–39 (2020). https://doi.org/10.29354/diag/122335
4. Bazgir O., Frounchi J., Habibi S. A.: A neural network system for diagnosis and assessment of tremor in Parkinson disease patients. In: 22nd Iranian Conference on Biomedical Engineering, pp. 1–5 (2015). https://doi.org/10.1109/ICBME.2015.7404105
5. Bhattacharyya, A., Pachori, R.B.: A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform. IEEE Trans. Biomed. Eng. **64**(9), 2003–2015 (2017). https://doi.org/10.1109/TBME.2017.2650259
6. Howe S. T., Aughey R. J., Hopkins, W. G. et al.: Quantifying important differences in athlete movement during collision-based team sports: accelerometers outperform Global Positioning Systems. In: IEEE International Symposium on Inertial Sensors and Systems, pp. 1–4 (2017). https://doi.org/10.1109/ISISS.2017.7935655
7. Russell, B., McDaid, A., Toscano, W., et al.: Moving the lab into the mountains: a pilot study of human activity recognition in unstructured environments. Sensors **21**(2), 1–14 (2021). https://doi.org/10.3390/s21020654
8. Boyali, A., Hashimoto, N., Matsumoto, O.: A signal pattern recognition approach for mobile devices and its application to braking state classification on robotic mobility devices. Robot. Auton. Syst. **72**, 37–47 (2015). https://doi.org/10.1016/j.robot.2015.04.008
9. Sarcevic, P., Kincses, Z., Pletl, S.: Online human movement classification using wrist-worn wireless sensors. J. Ambient. Intell. Humaniz. Comput. **10**(1), 89–106 (2019). https://doi.org/10.1007/s12652-017-0606-1
10. Nurwulan, N.R., Selamaj, G.: Random forest for human daily activity recognition. J. Phys: Conf. Ser. **1655**, 12087 (2020). https://doi.org/10.1088/1742-6596/1655/1/012087
11. Gao, Y., Ku, M., Qiant, T., Wang, J.: FFT formulations of adaptive Fourier decomposition. J. Comput. Appl. Math. **3**, 204–215 (2017). https://doi.org/10.1016/j.cam.2017.04.029
12. Alemayoh, T.T., Lee, J.H., Okamoto, S.: New sensor data structuring for deeper feature extraction in human activity recognition. Sensors **21**(8), 2814 (2021). https://doi.org/10.3390/s21082814
13. Giacomossi, L.C., Pichorim, S.F.: Pattern recognition of human movements using features extracted by triangulation method—a comparison with features of time and frequency domains. Res. Biomed. Eng. **38**(2), 539–557 (2022). https://doi.org/10.1007/s42600-022-00204
14. Farmaga, I., Shmigelskyi, P., Spiewak, P. et al.: Evaluation of computational complexity of finite element analysis. In: IEEE 11th International Conference the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), pp. 213–214 (2011)

# Determination of Oxy and Deoxyhemoglobin Concentrations in Skin Tissue Using Spatial Frequency Domain Imaging and Artificial Neural Network

M. R. Gonçalves[1]([✉]) [ID], B. G. da Silva[1] [ID], G. H. S. Alves[2] [ID], A. F. G. do Monte[2] [ID], and D. M. da Cunha[2] [ID]

[1] School of Electrical Engineering, Federal University of Uberlandia, Uberlândia, Brazil
murillo.goncalves@ufu.br
[2] Physics Institute, Federal University of Uberlândia, Uberlândia, Brazil
murillo.goncalves@ufu.br

**Abstract.** Spatial frequency domain imaging is an emerging technology that enables rapid, wide-field, and non-invasive chromophore mapping. Essentially, in this technique, a large area of the tissue is illuminated with a spatially modulated light field. The light beam reflected by the tissue depends on its optical properties, so that it can provide information about tissue composition through different chromophore concentrations. In this work, we employed a combination of Principal Component Analysis and Artificial Neural Networks to directly determine oxyhemoglobin and deoxyhemoglobin concentrations in skin tissue from diffuse reflectance values obtained from spatial frequency domain imaging. The database consisted of 850500 samples computed from the Beer's law and Monte Carlo simulations, and it was divided into training, validation, and testing subsets in a 0.7:0.15:0.15 ratio. To reduce overfitting during the network training, Bayesian regularization, based on the Levenberg-Marquardt optimization, was employed. Results showed that the developed model predict values of oxy and deoxyhemoglobin concentrations with a correlation coefficient of 0.997 and 0.982, respectively. The average errors from the expected values were 0.98% and 0.99%, for oxy and deoxyhemoglobin, respectively, with most of the samples showing absolute errors lesser than 4%. The developed model was applied to an in vivo study to determine hemoglobin concentrations in the hand of a volunteer. Results indicate that the developed model provides good performance in determining the oxyhemoglobin and deoxyhemoglobin concentrations, and it can be easily applied to in vivo measurements, with the potential to aid in the diagnosis of vascular changes in skin tissue.

**Keywords:** Spatial frequency domain imaging · Chromophores · Artificial neural networks

# 1 Introduction

The ability of light to penetrate biological tissues, interact with them, and carry their physiological information is the fundamental principle of different medical diagnostic techniques [1]. The evaluation of the processes of light absorption and scattering could be used to define chromophore concentrations, yielding valuable information about tissue health. Hemoglobin is an important protein that transports oxygen throughout the human body and its concentration monitoring can not only predict the success of a flap incision or a burn wound outcome, but it could also be used to monitor the efficiency of some cancer treatments [2–4]. During the last years, the technology of Spatial Frequency Domain Imaging (SFDI) has been considered to effectively create functional chromophores maps in a rapid, non-invasive, and wide-field way.

The principle of SFDI is based on the projection of a spatially modulated light beam over the tissue, which interacts with it, and then is captured by a CCD camera [1]. After calibration and demodulation procedures, a set of pixel-by-pixel maps are created, so that they represent the correspondent diffuse reflectance ($R_d$) values of the tissue at the different spatial frequencies employed. Inversion algorithms are then used to correlate this quantity to the tissue optical properties of absorption and scattering. Finally, a set of linear equations, based on Beer's law, are solved to determine chromophore concentrations [5].

In the last years, machine learning methods have been used to effectively determine chromophore concentrations in a fast and accurate way [6–8]. In particular, it has been proposed in literature that Artificial Neural Networks (ANN) could be employed to correlate chromophore concentrations directly from diffuse reflectance measurements in SFDI, which could lead to less computational time for signal processing and minimization of errors [9]. An artificial neural network (ANN) is a machine learning method that can efficiently find relationships within a dataset. Its power relies on the capacity to establish non-linear relationships, being largely used for pattern recognition [10]. Overall, an ANN consists of fully connected layers which are divided between an input layer, an output layer, and hidden layers. Each layer has associated neurons with specific weights and biases that are adjusted at each iteration, according to the training function. After successive iterations, the model tends to converge to a more accurate result [11]. Regarding applications in SFDI, Tsui et al. [12] and Wang et al. [13] employed a feedforward network with two hidden layers to obtain chromophore information from diffuse reflectance spectroscopy. Zhao et al. [9] employed a deep residual network to directly mapping chromophore concentration from Rd values. Others ANN architectures were also employed by different authors in the literature [14, 15].

In this work, we propose a combination of a machine learning model based on Artificial Neural Networks and Principal Component Analysis, for pattern recognition that directly outputs oxyhemoglobin ($HbO_2$) and deoxyhemoglobin (Hb) concentrations from diffuse reflectance values in a fast and accurate manner. We also apply the developed model to an in vivo study, in order to evaluate its feasibility.

## 2   Methodology

### 2.1   Spatial Frequency Domain Imaging

The principle of SFDI is based on measuring the reflected light beam by a wide area of tissue when this latter is illuminated with a patterned light field. A simplified model of the SFDI equipment is shown in Fig. 1a, and the experimental setup is shown in Fig. 1b. A white light source (D555WH, Vivitek®) was used to project a spatially modulated light beam on the tissue. The light beam interacts with the tissue and carries its physiological information. A filter wheel (CFW6, Thorlabs Inc®) selected a given light wavelength, and a CCD camera (DCC3240C Digital Camera, Thorlabs Inc®) was used to capture the signal. In this work, two different wavelengths were employed: 488 and 650 nm, and five spatial frequencies for the incident beam were evaluated: 0, 0.05, 0.1, 0.2, and 0.4 mm$^{-1}$, and for each one, three different phases were used (0°, 120°, and 240°).
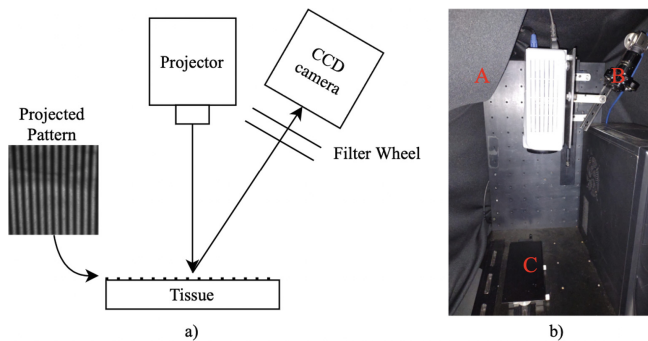


**Fig. 1.**  a) Schematic representation of the SFDI system. b) Experimental setup of the SFDI system from the Functional Imaging Lab (Federal University of Uberlandia): A - Projector; B - CCD camera with a filter wheel attached and C – Target location for positioning the sample.

For each light wavelength value, the image acquisition process consists in obtaining the diffuse reflectance ($R_d$) of the tissue at the spatial frequencies and phases considered. In this work, the determination of $R_d$ values was based on the methodology proposed by Cuccia et al. [5], which comprises the demodulation and calibration of the images captured by the camera, using an equivalent material simulator (phantom) [6]. At the end of this process, a set of five widefield images are created for each light wavelength, in which each pixel represents a diffuse reflectance value at a given spatial frequency.

### 2.2   Skin Model

The optical properties of skin tissue are directly related to its composition and can be defined quantitatively by the absorption and scattering coefficients [11]. These two parameters characterize the interactions of light photons with the tissue, and thus, they determine the behavior of the reflected beam [12]. In this work, the skin tissue was modeled considering oxy and deoxyhemoglobin as the major light absorbers [16]. In

particular, the absorption coefficient, $\mu_a$, measures how much the tissue absorbs light and is normally related to tissue metabolic functions [2]. As proposed by Tabassum et al. [2], the coefficient $\mu_a$ of the skin at a given light wavelength $\lambda$ can be computed as:

$$\mu_a(\lambda) = \ln(10)(C_{oxi}\varepsilon_{oxi}(\lambda) + C_{deoxi}\varepsilon_{deoxi}(\lambda)) \tag{1}$$

where $\varepsilon_{oxi}(\lambda)$ and $\varepsilon_{deoxi}(\lambda)$ are the extinction coefficients of oxyhemoglobin and deoxy-hemoglobin, respectively, in units of $cm^{-1}$/Mol, and $C_{oxi}$ and $C_{deoxi}$ are the respective concentrations of each chromophore, in units of Mol (M). The values of $\varepsilon_{oxi}(\lambda)$ and $\varepsilon_{deoxi}(\lambda)$ employed in this work were obtained from the data compiled by Jacques and Prahl [17].

To simulate real tissue parameters, we varied both hemoglobin and deoxyhemoglobin concentrations within a wide physiological range of 1–300 $\mu$M [11]. The reduced scattering coefficient, $\mu_s\prime$, which is related to the probability of a photon to be scattered, was also varied within the limits given in the literature for a human tissue, between 0.5–2.0 $mm^{-1}$ [9, 18]. The samples were organized in a full permutation manner, in order to model different skin compositions as possible [9].

## 2.3  Determination of $R_d$ $(\mu_a, \mu_s^{'})$ Database

The values of diffuse reflectance Rd for different pairs of coefficients $(\mu_a, \mu_s\prime)$ were obtained theoretically employing Monte Carlo (MC) simulations for light transport within the tissue. A MC code was developed, following the methodology proposed by Wang et al. [19]. The simulation consisted in projecting a light photon shower on a homogeneous medium characterized by four parameters: the absorption and reduced scattering coefficients $\mu_a$ and $\mu_s\prime$, respectively, the index of refraction $n$ and anisotropy factor $g$. The simulation returned the spatial distribution of the reflected beam, which can be employed for the computation of $R_d$ curves in the spatial frequency domain, through the application of a Fourier transform [6]. Values of $R_d$ at spatial frequencies of 0, 0.05, 0.1, 0.2 and 0.4 mm-1 were obtained, for different combinations of $\mu_a$ and $\mu_s\prime$ values in the interval $0.001 \leq \mu_a \leq 2.0$ mm-1 and $0.5 \leq \mu_s\prime \leq 2.0$ mm-1. In all cases, a value of $n = 1.4$ and $g = 0.7$ were considered, since they are representative of biological tissues [20]. The database generated with the simulations consisted of a total of 850500 samples, each one representing values of $R_d$ at different spatial frequencies at a given combination of $(\mu_a, \mu_s\prime)$ values. A surface plot of the values of $R_d$ at a spatial frequency of 0.1 $mm^{-1}$ is shown in Fig. 2, for different combinations of $\mu_a$ and $\mu_s\prime$.

## 2.4  Data Preprocessing

Due to the high dimensionality of the dataset, training an artificial neural network could be challenging and computationally demanding. To address this problem, we employed Principal Component Analysis (PCA), which is a multivariate statistic method for data dimensionality reduction [21]. This multivariate technique identifies new variables (principal components), which are linear combinations of the original variables, that maximize the data variance. In this work, each principal component $y_i$ at a given wavelength was computed as the linear combination of the five values of $R_d$ (one for each spatial

frequency), yielding five principal components. The first two components represented 99.9% of the total database variance, and they were kept in our model. The scores of each sample at the first two principal components were then obtained, so that the new database was composed by two components for each wavelength.

Additionally, a min-max normalization was implemented as a feature scaling method, so that sample values at each $y_i$ were rescaled to the interval [0–1], to improve machine learning accuracy [22, 23].
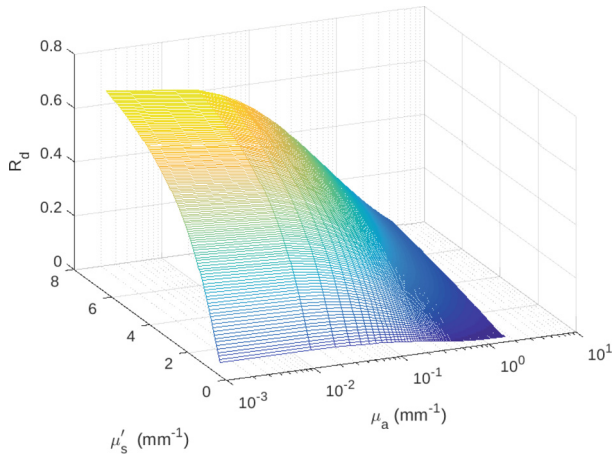


**Fig. 2.** Surface plot of $R_d$ values at a spatial frequency of 0.1 mm-1 for different combinations of optical coefficients $\mu_a$ and $\mu_s\prime$.

### 2.5   Artificial Neural Network Training and Testing

Once the database was created and preprocessed, it was used to train the ANN. Different ANN architectures were investigated, by varying the number of hidden layers and neurons in each layer. The ANN architecture adopted in this work consisted of a feedforward artificial neural network with 1 hidden layer and 25 neurons [7]. A model of the ANN is shown schematically in Fig. 3. The input layer receives four features: the first and second principal component scores at each wavelength. The output layer returns the concentrations $C_{oxy}$ and $C_{deoxy}$. Transfer functions of the hidden layer and output layer were hyperbolic tangent activation function and linear transfer function, respectively. The database was separated between training, validation, and testing at a 0.7, 0.15, and 0.15 ratio, respectively. To reduce overfitting, Bayesian regularization backpropagation was employed, which adjusts the weight and bias according to the Levenberg-Marquardt

optimization, and the error function considered was the mean squared error [24]. Additionally, for the testing and validation data, a 3% Gaussian noise was applied to the input values. As stopping criteria, 2000 training epochs or a mean squared error value of $10^{-6}$ were adopted to terminate the training process.
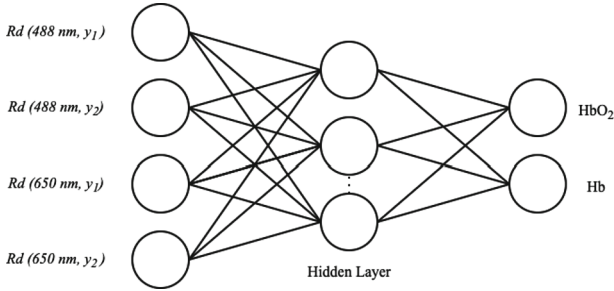


**Fig. 3.** Simplified model of the ANN. The first and second principal components from the PCA for each wavelength are inputs of the ANN, that directly outputs $HbO_2$ and Hb concentrations.

### 2.6 In-Vivo Study

In order to investigate the feasibility of the proposed model, an in vivo study was performed. The dorsal side of a volunteer's male hand was carefully positioned into the SFDI imaging system, and the ANN was used to directly calculate oxyhemoglobin and deoxyhemoglobin concentrations from the measured signal. The in vivo study in human beings has been approved by the Research Ethics Committee of the Federal University of Uberlândia (process n. 85363417.9.0000.5152).

## 3   Results

Figure 4 shows a scatter plot that compares the expected oxyhemoglobin concentration with the values estimated by the ANN. Results refer to the test data. Figure 5 shows the histogram of the relative percentual errors for the test data.

Figure 6 shows a scatter plot that compares the expected deoxyhemoglobin concentration with the values estimated by the ANN. Results refer to the test data. Figure 7 shows the histogram of error distribution of deoxyhemoglobin concentrations.

Figure 8(a) shows the in vivo study of the dorsal side of a patient hand, where the red rectangle represents the region of interest (ROI) captured by the CCD camera. Figure 8(b) shows the corresponding oxyhemoglobin concentration map, in units of µM, while Fig. 8(c) shows the deoxyhemoglobin concentration in µM.
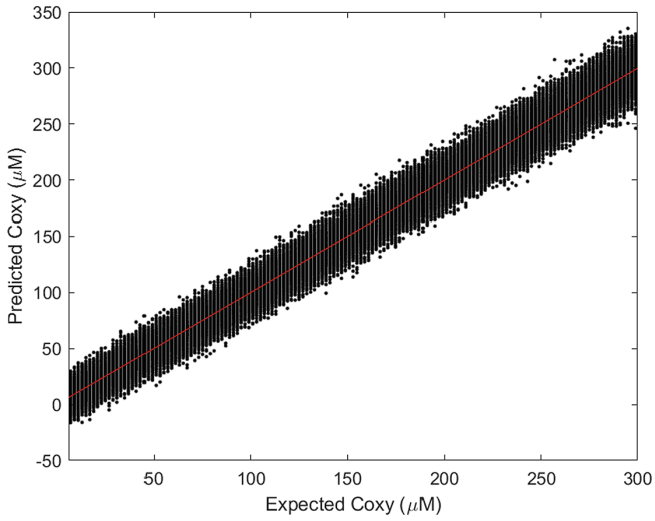
**Fig. 4.** Scatter plot of oxyhemoglobin concentration, comparing the expected values with the estimated values for the test data. The red line represents the expected values, and the black dots are the values estimated by the ANN.
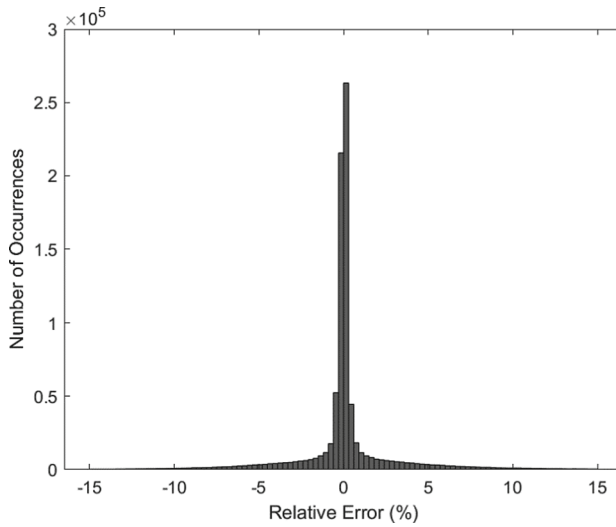


**Fig. 5.** Relative error distribution of oxyhemoglobin concentration for the test data.
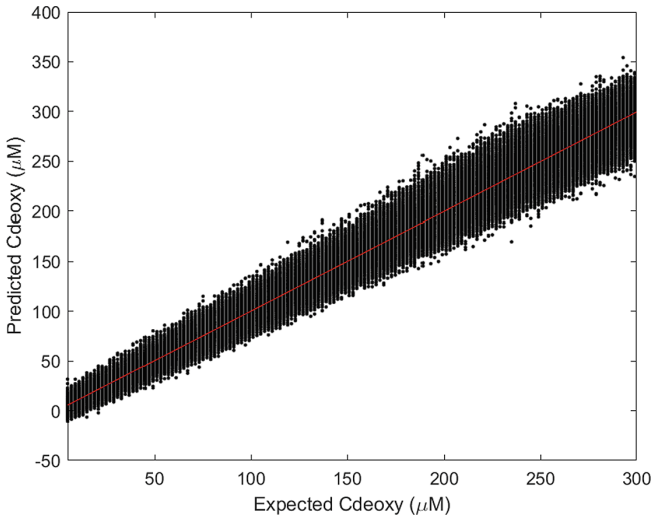
**Fig. 6.** Scatter plot of deoxyhemoglobin, comparing the expected values with the estimated values. The red line represents the expected values, and the black dots are the values estimated by the ANN.
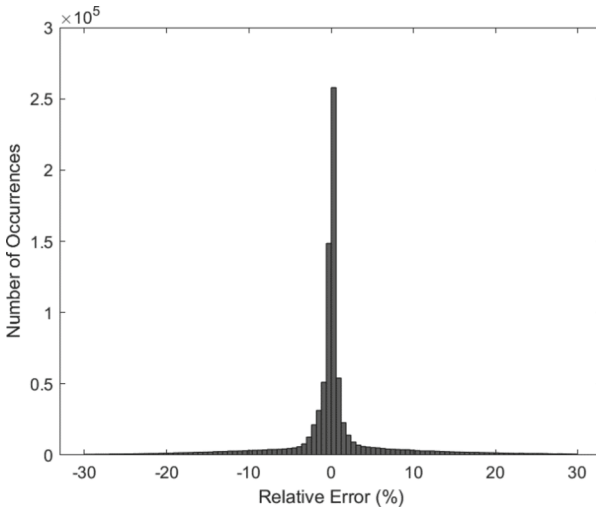


**Fig. 7.** Relative error distribution of deoxyhemoglobin concentration for the test data.
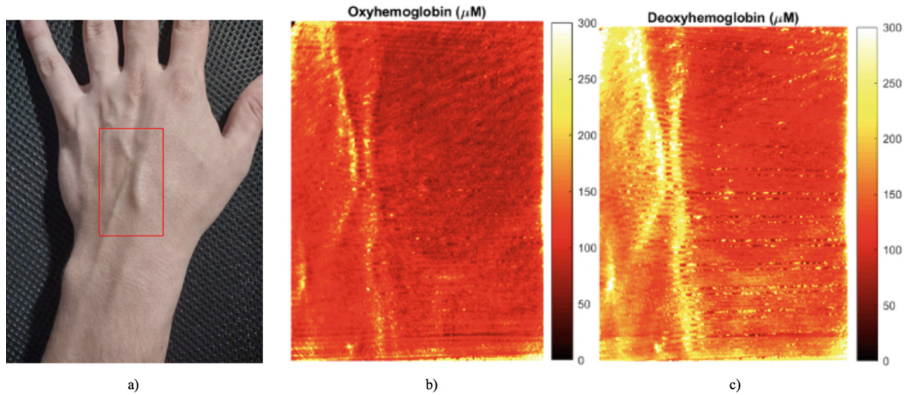
**Fig. 8.** a) Dorsal side of a patient hand used in the in vivo study. The red rectangle represents the imaged ROI. b) Oxyhemoglobin concentration map ($\mu M$) of the ROI. c) Deoxyhemoglobin concentration map ($\mu M$).

## 4 Discussion

Figures 4 and 6 show that the predicted values of oxyhemoglobin and deoxyhemoglobin are highly correlated with the expected values. In fact, the computation of $R^2$ factor, between predicted and expected values, yielded values of 0.997 and 0.982, for oxy- and deoxyhemoglobin, respectively, indicating a strong correlation between the quantities. From Fig. 5, it is observed that the average percentual error in predictions of oxyhemoglobin concentration was 0.99%, with 90% of samples lying within the error interval between -2 and 2%. For deoxyhemoglobin, the average error was also 0.98%, with most predicted values lying in the error margin interval between -4 and 4%. These results point to the excellent accuracy and precision of the developed model. They also indicate that simpler ANN architectures could be successfully employed in SFDI in order to obtain chromophore concentrations from diffuse reflectance data, in accordance with recent findings reported by Song et al. [25].

The in vivo study shows that the machine learning model can be applied to the SFDI technology, in order to determine hemoglobin concentrations from diffuse reflectance measurements. The concentration maps of oxy and deoxyhemoglobin of the hand of the volunteer, shown in Fig. 8, represent the vascularization of oxygenated and deoxygenated blood. The oxyhemoglobin concentration in the blood vessels were in the range [200, 250] $\mu$M, while values were approximately constant in the rest of the image, which lied in the interval [100, 150] $\mu$M. A similar behavior is observed in Fig. 8(c), for the deoxyhemoglobin concentration map, which shows values in the range [200, 300] $\mu$M for the blood vessels and [150, 200] $\mu$M in the remaining tissue. It is known that veins are rich in carbon dioxide and poor in oxygen, due to the conversion of arterial blood to venous blood in the hematosis process [26]. This can be seen in Fig. 8 as the blood deoxyhemoglobin concentration is higher than the oxyhemoglobin concentration, featuring the venous blood. These results points to the potential of the technique to provide diagnostic information about vascular changes in the skin, which could be related

to morphological or physiological tissue alterations, like those observed in different types of skin tumors [27].

## 5   Conclusion

In this work, we presented a combination of principal component analysis and an artificial neural network model for pattern recognition that accurately determines oxyhemoglobin and deoxyhemoglobin concentrations in skin tissue directly from the diffuse reflectance values, obtained from the SFDI system. To measure the effectiveness of our model, we used the $R^2$ correlation factor and a histogram of relative errors. For the oxyhemoglobin, a value of 0.997 was obtained for $R^2$ and the relative errors were mostly concentrated between the interval [-2, 2] % with a 0.99% mean. For the deoxyhemoglobin, the $R^2$ factor was 0.982 and the errors were concentrated in the [-4, 4] % range with 0.98% average. Additionally, we employed the developed model in an in vivo study, showing its feasibility in acquiring functional images of hemoglobin concentration. The obtained results indicate the good accuracy and usability of the model, since images are taken in a wide-field, non-contact, and rapid manner. These results particularly point to the potential of the technique in supporting physicians in the detection of morphological or physiological changes in vascularization of the skin, which could be related to different tissue abnormalities, for example, as those observed in skin cancer.

**Conflict of Interest.**   The authors declare that they have no conflict of interest.

## References

1. Gioux S., Mazhar A., Cuccia D.J.: Spatial frequency domain imaging in 2019: principles, applications, and perspectives. J Biomed Opt. **24**(7), 071613-1–071613-18 (2019)
2. Tabassum, S., Zhao, Y., Istfan, R., et al.: Feasibility of spatial frequency domain imaging (SFDI) for optically characterizing a preclinical oncology model. Biomed. Opt. Express **7**(10), 4154–4170 (2016)
3. Nguyen, J.T., Lin, S.J., Tobias, A.M., et al.: A novel pilot study using spatial frequency domain imaging to assess oxygenation of perforator flaps during reconstructive breast surgery. Ann. Plas. Surg. **71**, 308–315 (2013)
4. Kennedy, G.T., Stone, R., Kowalczewski, A.C., et al.: Spatial frequency domain imaging: a quantitative, noninvasive tool for in vivo monitoring of burn wound and skin graft healing. J. Biomed. Opt. **24**(7), 0716151–0716159 (2019)
5. Cuccia, D.J., Bevilacqua, F., Durkin, A.J., Ayers, F.R., Tromberg, B.J.: Quantitation and mapping of tissue optical properties using modulated imaging. J. Biomed. Opt. **14**(2), 024012 (2009). https://doi.org/10.1117/1.3088140

6. Barman, I., Dingari, N.C., Rajaram, N., et al.: Rapid and accurate determination of tissue optical properties using least-squares support vector machines. Biomed. Opt. Express **2**(3), 592–599 (2011)
7. Fredriksson, I., Larsson, M., Strömberg, T.: Machine learning for direct oxygen saturation and hemoglobin concentration assessment using diffuse reflectance spectroscopy. J. Biomed. Opt. **25**(11), 112905 (2020). https://doi.org/10.1117/1.JBO.25.11.112905
8. Smith, J.T., Ochoa, M., Faulkner, D., Haskins, G., Intes, X.: Deep learning in macroscopic diffuse optical imaging. J. Biomed. Opt. **27**(02), 020901 (2022). https://doi.org/10.1117/1.JBO.27.2.020901
9. Zhao, Y., Deng, Y., Yue, S., et al.: Direct mapping from diffuse reflectance to chromophore concentrations in multi-fx spatial frequency domain imaging (SFDI) with a deep residual network (DRN). Biomed. Opt. Express **12**(1), 433–443 (2021)
10. Zou, J., Han, Y., So, S.S.: Overview of artificial neural networks. Method Mol. Cell Biol. **458**, 15–23 (2008)
11. Jacques, S.L.: Optical properties of biological tissues: a review. Phys. Med. Biol. **58**(14), 5007–5008 (2013)
12. Tsui, S.Y., Wang, C.Y., Huang, T.H., et al.: Modelling spatially-resolved diffuse reflectance spectra of a multi-layered skin model by artificial neural networks trained with Monte Carlo simulations. Biomed. Opt. Express **9**(4), 1531–1544 (2018)
13. Wang, C.Y., Kao, T.C., Chen, Y.F., et al.: Validation of an inverse fitting method of diffuse reflectance spectroscopy to quantify multi-layered skin optical properties. Photonics Res. **6**(2), 1–17 (2019)
14. Aguénounon, E., Smith, J.T., Al-Taher, M., et al.: Real-time, wide-field and high-quality single snapshot imaging of optical properties with profile correction using deep learning. Biomed. Opt. Express **11**, 5701–5716 (2020)
15. Chen, M.T., Durr, N.J.: Rapid tissue oxygenation mapping from snapshot structured-light images with adversarial deep learning. J. Biomed. Opt. **25**(11), 112907 (2020). https://doi.org/10.1117/1.JBO.25.11.112907
16. Wirkert, S.J., Kenngott, H., Mayer, B., et al.: Robust near real-time estimation of physiological parameters from megapixel multispectral images with inverse Monte Carlo and random forest regression. Int. J. Comput. Ass Rad. **11**(6), 909–917 (2016)
17. OMLC Homepage. https://omlc.org/index.html. Accessed 20 Mar 2022
18. Lister, T., Wright, P.A., Chappell, P.H.: Optical properties of human skin. J. Biomed. Opt. **17**(9), 0909011 (2012). https://doi.org/10.1117/1.JBO.17.9.090901
19. Wang, L., Jacquesa, S.L., Zhengb, L.: CML-Monte Carlo modeling of light transport in multi-layered tissues. Comput. Method Prog. Bio. **47**, 131–146 (1995)
20. Yudovsky, D., Durkin, A.J.: Spatial frequency domain spectroscopy of two layer media. J. Biomed. Opt. **16**(10), 107005 (2011)
21. Abdi, H., Williams, L.J.: Principal component analysis. Wirel. Comput. Stat. **2**(4), 433–459 (2010)
22. Bishop, C.M.: Neural Networks for Pattern Recognition, 1st edn. Clarendon Press, Oxford (1995)
23. Shanker, M., Hu, M.Y., Hung, M.S.: Effect of data standarization on neural network training. Int. J. Manage. Sci. **24**(4), 385–397 (1996)
24. Burden, F., Winkler, D.: Bayesian regularization for neural networks. Method Mol Bio. **458**, 25–44 (2008)
25. Song, B., Jia, W., Zhao, Y., et al.: Ultracompact deep neural network for ultrafast optical property extraction in spatial frequency domain imaging (SFDI). Photon. Res. **9**(5), 1–12 (2022)

26. Rieser, T.M.: Arterial and venous blood gas analyses. Top Companion Anim. Med. **28**(3), 86–90 (2013)
27. Kreusch, J.F.: Vascular patterns in skin tumors. Clin. Dermatol. **20**(3), 248–254 (2002). https://doi.org/10.1016/S0738-081X(02)00227-4

# Single Trial P300 Detection Using Dimensionality Reduction and Extreme Learning Machine

C. F. Blanco-Díaz$^{(\boxtimes)}$ , C. D. Guerrero-Mendez , and T. F. Bastos-Filho

Postgraduate Program in Electrical Engineering, Federal University of Espírito Santo (UFES), Vitoria, Brazil
`cblanco88@uan.edu.co`

**Abstract.** Brain-Computer Interfaces (BCI) are systems that function as a communication channel between people and external devices through brain information. P300 is an Event-Related Potential (ERP) widely used to decode an individual's intent in BCI system applications, such as spellers in communication skill rehabilitation systems. However, the Signal-to-Noise Ratio (SNR), which quantifies a relation between the power of the transmitted signal and the power of the noise that corrupts it, for P300 is low so which makes it difficult to detect the using a single trial, which reduces the accuracy of the system. Additionally, in real-time BCI systems, it is necessary to maintain a good ratio of detection and execution time. In this work, three methods based on mean-based feature dimensionality reduction (PCA – Principal Component Analysis, FA – Factor Analysis, and MDS – Multi-Dimensional Scaling) as well as ELM – Extreme Learning Machine are presented, which allow efficient detection of a single trial P300 using 65.25% of the features normally used in the literature. PCA and MDS techniques reach an Area Under the Received Operation Curve (AUC) maximum of 0.90, and an average of 0.72, which are significant compared to the standard method based on the signal mean. Additionally, training times below 0.05 s are obtained, which are very important for real-time operation. The results allow us to conclude that the proposed methods are suitable for the detection of the single-trial P300, which can be used in BCI systems with real-time speller for rehabilitation engineering.

**Keywords:** Brain Computer Interfaces (BCI) · Single-trial P300 detection · Principal Component Analysis (PCA) · Multi-Dimensional Scaling (MDS) · Factor Analysis (FA)

## 1 Introduction

Nowadays, Brain-Computer Interface (BCI) systems have been widely used in rehabilitation engineering as they work as a communication channel between the human brain and the computer. These systems allow the interaction of an individual with the environment by decoding brain information (e.g. Electroencephalography (EEG)) to recognize intention, and subsequently execute a specific task [1]. However, to make this possible it is necessary to generate patterns

in the brain signals of the individual. The literature reports a large number of paradigms that generate these patterns, such as the Event-Related Potential (ERP) P300 evoked by the oddball paradigm [2]. The P300 signal consists of a positive deflection in the parieto-central lobe of the brain, which appears after a visual or auditory stimulus with a latency of approximately 300 ms. The P300 signal is often used with spellers as it can be adapted for the rehabilitation of individual's communication skills [2,3].

However, to this days there are still problems in identifying P300 signals, since ERPs have a very low SNR [3]. A common method is to synchronously average many trials, which effectively decreases the random noise, but this is not an effective method in BCI applications because it is quite slow, reducing the communication rate to a large extent. Many researchers have tried to solve this problem by reducing the number of averaged trials or going toward single-trial detection [4–7]. Nevertheless, the detection of single-trial P300 is still an open challenge for the scientific community. Another way to improve the use of the number of trials is to use robust techniques that allow a correct extraction of features, such as wavelets [4], component analysis [6], nMeans [8], but this can lead to an increase in the complexity and number of features and channels, which usually increases the computational cost.

In recent years, the development of new machine learning algorithms, such as Convolutional Neural Networks (CNN) [9,10] or ELM [11,12] has improved classification rates in single-trial P300 detection [6,13]. However, increasing the robustness of a classifier usually brings an increase in training time, which is not very practical in real-time. Therefore, a pre-classification phase reported in the literature consists of dimensionality reduction, where PCA is often used to reduce the size of the feature vector in a CNN or other types of classifiers [14]. Nevertheless, in ELM, dimensionality reduction techniques have not been deepened. On the other hand, FA is a mathematical technique commonly used in P300 for the neuroscience area, however, as a dimensionality reduction technique, it has not been used [15] as does have MDS [16,17].

It is also important in BCI systems to increase the generalization of the methods, avoiding overfitting of computational techniques, because the performance of a BCI system from one subject to another may change due to physical, mental, personal, and other conditions (intersubject variability). For this reason, the implementation of P300 decoding in large population samples is a major challenging issue [2,5,7,13,14].

In this paper, the objective is to improve the detection of P300 in a single trial using PCA, FA, and MDS algorithms to reduce the dimension of the feature vector nMeans, which not only reduces the calculation time, but also improves the SNR of the data [14]. Additionally, ELM is used for P300 signal classification. The methods were evaluated on a public database based on 54 individuals running the experimental protocol of a P300-based BCI speller system. The AUC metric was evaluated, obtaining a maximum value of approximately 0.9. PCA and MDS obtained better overall performance than the standard nMeans method ($p < 0.05$), with AUC between 0.69 and 0.70 with 53 features. The results allow

concluding that PCA and MDS methods allow maximizing the P300 recognition and, at the same time, a reduction of the initial feature size, obtaining significant results with approximately 66.25% of the feature vector, and with training time lower than 50 ms. These results allow evaluating an extraction, dimensionality reduction, and classification configuration that would allow an adequate recognition in an online BCI speller system based on P300.

## 2  Materials and Methods

The methodology was segmented into selecting a BCI speller system based on P300 dataset, signal processing and feature extraction, dimensionality reduction, classification, and statistical analysis. The summary of the methodology is represented in Fig. 1.
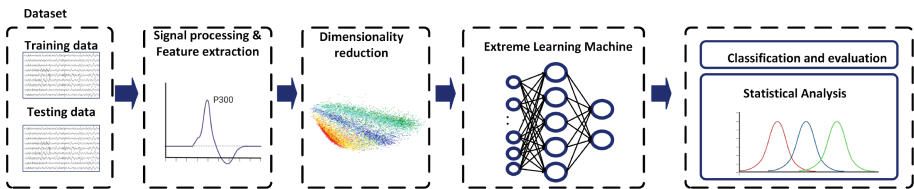


**Fig. 1.** Block diagram of the methodology for the identification of P300 using dimensionality reduction and ELM as classifier.

### 2.1  Dataset

The public database implemented by Lee *et al.*[8] was used. P300 signals were recorded from 54 individuals at a sampling rate of 1000 Hz with 62 EEG electrodes placed on the cerebral cortex, according to the 10–20 international system. The experimental protocol consisted of two sessions (one training and one validation), where individuals had to interact with a $6 \times 6$ speller, forming a sentence of 33 characters in the first session and 36 characters in the second session, with a stimulation time of 80 ms and inter-stimulus of 135 ms. Each session was repeated 10 times, generating 1980 (330 target and 1650 non-target signals) and 2160 trials (360 target and 1800 non-target signals) for the training and evaluation session, respectively. In this study, the data from the first session were used to train the models, and those from the second session to validate them.

### 2.2  Signal Processing and Feature Extraction

The EEG signals were filtered with a band-pass filter of zero-phase shift based on both Fast Fourier Transform (FFT) and Inverse FFT (IFFT), which firstly

transforms an epoch into the frequency domain, making zero those frequency components outside the band of interest (0.1 to 30 Hz) by multiplying a rectangular bandpass filter function. These were segmented between 0 and 800 ms, where a time window between –200 ms and 0 was used to remove the baseline. The channels were reduced of 62 to 8 more used channels in the literature for P300 ($F_Z$, $C_Z$, $P_Z$, $P_3$, $P_4$, $O_Z$, $PO_3$ and $PO_4$) [13]. The system outputs correspond to the target signals (P300) identified as positive outputs, and no-target signals (non-P300) as the negative outputs of the system. To validate the integration of the data, all the data segmented by target and non-target were averaged. For feature extraction, the method nMeans was used, which consists of extracting mean values of the signal at $n$ time-windows between 0 and 800 ms for each channel, and concatenating them together [8]. Considering that by increasing the number of $n$ a higher number of features is obtained (which cause overfitting) and with a low number underfitting can occur, a number of $n = 10$ was considered for obtaining P300 features [8].

## 2.3   Dimensionality Reduction

For dimensionality reduction, three methods were applied through the open MATLAB toolbox for dimensionality reduction proposed by Laurens van der Maaten, available in [17]. The proposed methods were compared with the original method for feature extraction, which consists of nMeans without dimensionality reduction.

**Principal Component Analysis (PCA).** PCA transforms the original data into a set of linearly independent data vectors in various dimensions through linear transformation, which can be used to extract the main feature components of the data. PCA is often used for dimensionality reduction of high-dimensional data. Summarizing, the PCA method is computed by Eq. 1. More information on this method is available in [14].

$$Y = PX, \tag{1}$$

where $X$ corresponds to the feature vector nMeans, $P$ corresponds to the matrix that represents the principal components, and generates the new feature vector $Y$. In this case, $Y$ has a number of $k$ features where $k = 1,2,3...80$.

**Factor Analysis (FA).** Factor analysis is a technique used to reduce a large number of factors to a smaller number. The technique is based on the extraction of maximum common variances that explain the correlations between observed variables in terms of a smaller number of unobserved variables called factors. Summarizing, the FA method can be computed by Eq. 2. More information on this method is available in [14].

$$Y = FX, \tag{2}$$

where $Y$ is the new feature vector, with $k$ number of features ($k$ =1,2,3...80), $X$ is the features vector obtained by the nMeans, and $F$ corresponds to the factor matrix. It is possible to see that the equation for FA and PCA are very similar, however, the way of computing $F$ and $P$ matrices is different.

**Multi-Dimensional Scaling (MDS).** Multi-Dimensional Scaling (MDS) starts from a notion of distance or affinity $K$ that is computed between each pair of training examples. The distances are converted to equivalent dot products using the double-centering formula given by the Eq. 3. More information on this method is available in [16].

$$\tilde{M}_{ij} = -\frac{1}{2}\left(M_{ij} - \frac{1}{n}S_i - \frac{1}{n}S_j + \frac{1}{n^2}\sum_k S_k\right),\tag{3}$$

where $M$ corresponds to the features matrix obtained from nMeans, $S_i$ is the *i-th* row sum of the affinity matrix $M$, $S_j$ is the *j-th* columns sum of the affinity matrix $M$, and $\tilde{M}$ corresponds to the final matrix whose eigenvalues are found to obtain the matrix with reduced features.

## 2.4   Extreme Learning Machine (ELM)

Extreme Learning Machine (ELM) is a neural network family of Single Hidden Layer Feedback Neural Networks (SHLFN), which can be used in classification problems, created by Huang *et al.* in 2006 [11]. It is composed of a single hidden layer with its respective activation function, where the objective is to train the network, modifying the weights of the neurons of the inner layer to such an extent that it allows the classification of a group of input variables [12]. ELM has advantages over other types of neural networks, since the training time is low, which makes it a suitable option for classification in BCI systems. The ELM output function is given by Eq. 4.

$$f_j = \sum_{i=1}^{L} \beta_i \varrho(\omega_i \cdot x_j + b_i),\tag{4}$$

where $\beta_i$ is a column vector of the weights relating the *i-th* hidden layer neurons and the output neurons, calculated analytically. $\varrho$ is a matrix denoting the activation function of each hidden layer neuron (this case is a sigmoid function), $W$ is the weights vector connecting the *i-th* hidden layer neuron and the input neuron, calculated randomly, $b$ is the threshold of each *i-th* hidden layer neuron, $x_j$ is the *j-th* feature given by the output of the dimensionality reduction methods after normalization, and $f$ corresponds to the network output, which, for this case, corresponds to the classification between P300 (positive class) and non-P300 (negative class). The hidden layer was adjusted to a total of 2*k neurons, where $n$ equals the number of actual features, according to the dimensionality reduction method ($k$=1,2,3...80).

### 2.5   Evaluation Metrics

According to the confusion matrices, one of the most commonly used metric reported in the literature for P300 classification is the Area Under the Curve (AUC) of the Received Operation Curve (ROC), as it is a metric that relates the true positive rate to the false positive rate. In this case, being a biased classification problem, the positive variable (P300) is of interest [5]. The AUC is defined by Eq. 5.

$$AUC = \int TPR \, d(FPR),\tag{5}$$

where $TPR$ and $FPR$ are the true positive rate and the false positive rate, respectively. Also, the neural network training time was used as a performance metric, which was estimated in MATLAB (Version 2021b, MathWorks Inc) on a computer with an 8th generation intel core i5 processor of 64 bits and 8 GB of RAM memory.

### 2.6   Statistical Analysis

An one-way ANOVA was implemented using MATLAB to determine significant differences between the performance of the proposed methods based on dimensionality reduction. ANOVA was applied with a significance $p-value$ of 0.05. The alternative hypothesis was that dimensionality reduction-based methods perform significantly better (according the AUC) than the reference nMeans method, depending on the number of features evaluated ($k$=1,2,3 ... 80) for the detection of single trial P300. The null hypothesis is the opposite. Additionally, according the training time, the alternative hypothesis is that methods based on dimensionality reduction have a significantly shorter training time than the reference method.

## 3   Results

Before performing the feature extraction, a validation of the data integrity was performed, where through the averaging of the signals it is possible or not to see the behavior of the P300 signal in the different channels for the 54 individuals. For this reason, it is possible to see in Fig. 2 the behavior of the EEG signal averaged between target and non-target for a test individual in the time segment between 0 and 800 ms. It is possible to notice the positive deflection at about 300 ms after the trigger (P300). Normally, the P300 signal is generated on the parieto-central cortex of the brain [3], however, stimulation, when visual, can generate evoked potentials in the occipital lobe of the brain, such as $O_Z$, $PO_3$ or $PO_4$ [2].

On the other hand, Fig. 3 shows the behavior of the average AUC variable of the 54 subjects for each number of features used after the reduction performed by PCA, FA, and MDS. Additionally, the behavior of the AUC performance of the nMeans method with 80 features functions as a reference with the aim of comparing the methods applied in this study with what is reported in the literature
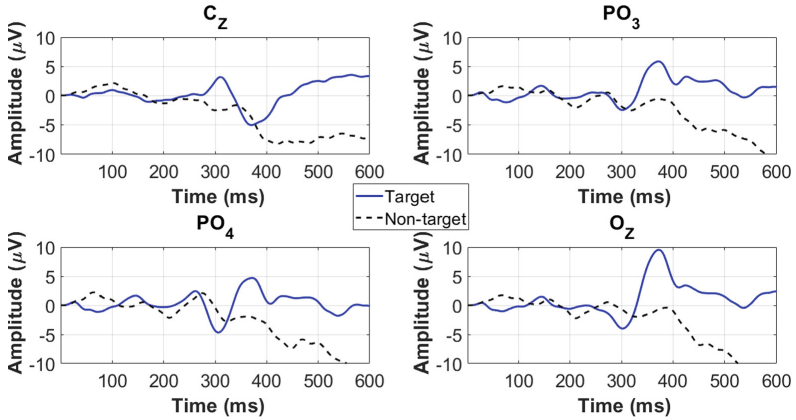
**Fig. 2.** Behavior of the averaged EEG signal in the $C_Z$, $PO_3$, $PO4$ and $O_Z$ channels for a test subject when a P300 is being evoked (target) and when it is not evoked (non-target).

[8]. The performance of the reference method was approximately 0.60±0.19, with a training time of 0.066±0.02 s, approximately, whereas PCA and MDS methods performed similarly using 40 features: both with a value close to 0.65±0.09 of AUC, but with a training time of approximately 0.0194±0.0064 s for PCA, and 0.0203±0.0065 s for MDS. The best average performance obtained by the methods is using 80 features, with a result of approximately 0.72±0.09, 0.60±0.07, and 0.72±0.08 of AUC for PCA, FA, and MDS, respectively. To obtain these values, an average training time of approximately 0.0449±0.015, 0.0448±0.0157, and 0.0163 s was necessary for PCA, FA, and MDS, respectively. The maximum P300 detection performance for one individual was AUC of approximately 0.81 with reference method, 0.90 with PCA, 0.77 with FA, and 0.87 with MDS. It should be noted that these values allow probabilistic determination of the difference between the P300 signal and the non-P300 signals in BCI systems in real time, using a single trial. This value can be increased by averaging more trials, however, this requires more stimulation time and computational cost.

In the evaluation of the statistical analysis for the AUC metric using a *for loop*, it is observed, for the number of features $k > 1$, that the alternative hypothesis is rejected up to the value $k = 53$, which is where the null hypothesis is rejected because, in terms of this metric, the PCA and MDS methods obtain a better performance, which is significant ($p < 0.05$). With the number of features $k=53$, the results 0.69±0.10 and 0.70±0.09 of AUC are obtained for the PCA and MDS methods. Finally, it is possible to observe that for the network training time, the times of the dimensionality reduction-based methods are significantly shorter than the reference method ($p < 0.05$).

(a) AUC                          (b) Training time
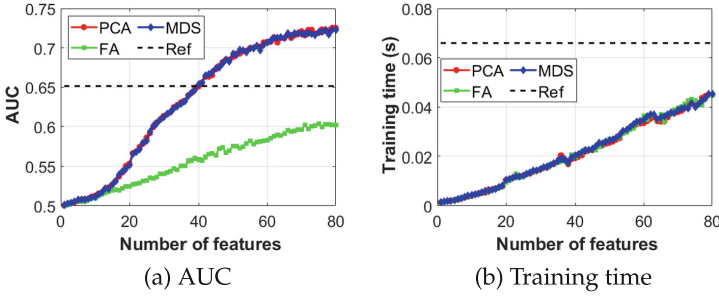
**Fig. 3.** Behavior of the evaluation metrics for the three dimensionality reduction methods evaluated: (a) AUC; (b) Training time, for each number of features evaluated. Reference line corresponds to the behavior of the nMeans method without dimensionality reduction.

## 4    Discussion

The results of the AUC performance metric are shown in Fig. 3a, where it is possible to observe an increasing performance as the number of features increases, nevertheless, when using the same number of features as the benchmark method proposed by Lee *et al.*, based on nMeans with 80 features [8], it is observed that the performance of the PCA and MDS methods is higher. This perhaps is because dimensionality reduction methods not only allow the feature vector size reduction, but also improve the SNR of the data, as stated by Li *et al.* in their study of PCA with CNN [14]. On the other hand, the AUC values for single-trial P300 detection are among the values reported in the literature, where average AUCs of 0.57, 0.62, and 0.65 have previously been reported using signal shape-based methods and linear classifiers in a population with 10 individuals [5].

For ELM, the extraction methods had similar performances where Kong *et al.* reported a classification rate of approximately 72% for a $6 \times 6$ speller with two test individuals for 10 test sequences using signal components as input features to the neural network [6]. Additionally, Xie *et al.* reported for 8 test subjects an average classification rate of 86% for a single trial using 2500 neurons in the hidden layer of the ELM testing different channels of the parieto-central-occipital lobe of the brain, and training times between 0.01 and 0.14 s, however, the latter does not use speller as a paradigm for ERP evocation [13].

Thus, the results allow determining that the methods based on dimensionality reduction and ELM as classifier improve the detection of the P300 single-trial both in terms of accuracy and training time. This is because PCA and MDS methods obtain a maximum average performance of approximately 0.72±0.09 and 0.72±0.08, respectively, although a significant improvement of 0.07 is found from the use of 53 features, which, compared to the initial nMeans vectors, corresponds to approximately 66.25% of the vector. According to the ELM training time, the methods based on dimensionality reduction propose an improvement

because the times were less than 0.05 s, which is fully significant compared to the reference method. For example, if a configuration of PCA and MDS methods with 53 features is used, it would be possible to obtain an acceptable classification of about 70%, with a response time of about 0.02 s. On the other hand, the FA-based reduction method did not obtain an improvement compared to the reference method. When comparing the performance of PCA and MDS with each other, no significant differences are obtained. This may be because both methods are based on linear configurations, so their behavior in some cases may be similar [16, 17]. Finally, the generalizability of the methods was demonstrated by evaluating the configurations in a sample of 54 subjects.

## 5 Conclusion

The results of this study allow concluding that the techniques based on dimensionality reduction using PCA and MDS to reduce the feature vectors based on average of EEG signals and ELM neural network improved the recognition of the single-trial P300 in BCI systems based on speller. The principal results were obtained with an AUC performance of 0.70, which is higher than the standard method ($p<0.05$) with 66.25% of the initial feature vector and maximum performance of 0.90 for one individual evaluated in a single trial. Additionally, ELM training times of less than 0.05 s were obtained, which would allow practical application in real-time BCI systems related with rehabilitation of communications disabilities. Future studies will focus on the application of the methods in a real-time BCI speller system, as well as the evaluation of more robust methods that allow improving the performance metrics and the reduction of computational cost at the same time.

## References

1. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain–computer interfaces for communication and control. Clin. Neurophysiol. **113**(6), 767–791 (2002). ISSN 1388-2457. https://doi.org/10.1016/S1388-2457(02)00057-3

2. Farwell, L.A. Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr. Clin. Neurophysiol. **70**(6), 510–523 (1988). ISSN 0013-4694. https://doi.org/10.1016/0013-4694(88)90149-6, https://www.sciencedirect.com/science/article/pii/0013469488901496

3. Picton, T.: The P300 wave of the human event-related potential. J. Clin. Neurophysiol. **9**, 456–479 (1992). https://doi.org/10.1097/00004691-199210000-00002

4. Farukh Hashmi, M., Kene, J.D., Deepali, M.: An efficient p300 detection algorithm based on kernel principal component analysis-support vector machine. Comput. Elect. Eng. **97**, 107608 (2022). ISSN 0045-7906. https://doi.org/10.1016/j.compeleceng.2021.107608

5. Díaz, C.F.B., Olaya, A.F.R.: A novel method based on regularized logistic regression and CCA for P300 detection using a reduced number of EEG trials. IEEE Lat. Am. Trans. **18**(12), 2147–2154 (2020). https://doi.org/10.1109/TLA.2020.9400443

6. Kong, W., et al.: Weighted extreme learning machine for p300 detection with application to brain computer interface. J. Amb. Intell. Human. Comput. 1–11 (2018)

7. Nashed, N.N., Eldawlatly, S., Aly, G.M.: A deep learning approach to single-trial classification for p300 spellers. In: 2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME), pp. 11–16 (2018). https://doi.org/10.1109/MECBME.2018.8402397

8. Lee, H., et al.: EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. Giga-Science **8**, 1–16 (2019). https://doi.org/10.1093/gigascience/giz002

9. Kundu, S., Ari, S.: Fusion of convolutional neural networks for p300 based character recognition. In: 2019 International Conference on Information Technology (ICIT), pp. 155–159 (2019). https://doi.org/10.1109/ICIT48102.2019.00034

10. Kundu, S., Ari, S.: P300 based character recognition using convolutional neural network and support vector machine. Biomed. Signal Process. Control. **55**, 101645 (2020). ISSN 1746-8094. https://doi.org/10.1016/j.bspc.2019.101645, https://www.sciencedirect.com/science/article/pii/S1746809419302265

11. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: Extreme learning machine: theory and applications. Neurocomputing **70**(1–3), 489–501 (2006)

12. Henrique, V., Ribeiro, A., Reynoso-Meza, G., Valadares Siqueira, H.: Multiobjective ensembles of echo state networks and extreme learning machines for streamflow series forecasting. Eng. Appl. Artif. Intell. **95**, 103910 (2020)

13. Xie, S., Wu, Y., Zhang, Y., Zhang, J., Liu, C.: Single channel single trial p300 detection using extreme learning machine: Compared with BPNN and SVM. In: 2014 International Joint Conference on Neural Networks (IJCNN), pp. 544–564 (2014). https://doi.org/10.1109/IJCNN.2014.6889400

14. Li, F., Li, X., Wang, F., Zhang, D., Xia, Y., He, F.: A novel p300 classification algorithm based on a principal component analysis-convolutional neural network. Appl. Sci. **10**(4), 1546 (2020)

15. Lytaev, S., Vatamaniuk, I.: Physiological and medico-social research trends of the wave p300 and more late components of visual event-related potentials. Brain Sci. **11**(1), 125 (2021)

16. Bengio, Y., Paiement, J., Vincent, P., Delalleau, O., Roux, N., Ouimet, M.: Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and spectral clustering. In: Advances in Neural Information Processing Systems, vol. 16 (2003)

17. Van Der Maaten, L., Postma, E., Van den Herik, J., et al.: Dimensionality reduction: a comparative. J. Mach. Learn. Res. **10**(66–71), 13 (2009)

# 2D Time-Difference Electrical Impedance Tomography Image Reconstruction in a Head Model with Regularization by Denoising

Roberto G. Beraldo[1(✉)] [iD], Leonardo A. Ferreira[1] [iD], Fernando S. Moura[2] [iD],
André K. Takahata[1] [iD], and Ricardo Suyama[1] [iD]

[1] Federal University of ABC, Avenida dos Estados, 5001, Santo André, SP
09210-580, Brazil
`roberto.gutierrez@ufabc.edu.br`
[2] Federal University of ABC, Alameda da Universidade, s/n, São Bernardo do
Campo, SP 09606-045, Brazil

**Abstract.** Time-Difference Electrical Impedance Tomography (TDEIT) is an imaging technique to visualize resistivity changes over time in a region of interest. Regularization is necessary because TDEIT is an ill-posed problem. In this work, we use Regularization by Denoising (RED) with four different denoisers to reconstruct brain images in a simplified 2D head model. We compared the RED results to two traditional reconstruction methods, generalized Tikhonov regularization and total variation regularization. In both the noiseless and the noisy scenarios, we achieved the best results using RED with non-local means as the denoiser in relation to figures of merit such as ringing, resolution and shape deformation.

**Keywords:** ill-posed problems · discrete inverse problems · image processing · white Gaussian noise

## 1 Introduction

Electrical Impedance Tomography (EIT) is a method to obtain resistivity values from a region of interest (ROI) [1]. To do so, electrodes inject electrical current into the boundary of the ROI. The resulting electrical potentials are measured in the same boundary and used in image reconstruction.

It is possible to use the finite element method (FEM) to calculate the forward problem, i.e. to determine all the nodal voltages given the resistivities of the elements and the nodal electrical current [1]. Let $p$ be the number of mesh nodes and $n$ the number of mesh elements. Then, the voltage $\mathbf{v}_{all} \in \mathbb{R}^p$ in all the mesh nodes is given by

$$\mathbf{v}_{all} = \mathbf{Y}(\boldsymbol{\rho})^{-1}\mathbf{c}, \tag{1}$$

which is a nonlinear function of the resistivities $\boldsymbol{\rho} \in \mathbb{R}^n$, where $\mathbf{Y}(\boldsymbol{\rho}) \in \mathbb{R}^{p \times p}$ is the global conductivity matrix and $\mathbf{c} \in \mathbb{R}^p$ is the nodal electrical current [2, p. 53].

The method is called Time-Difference Electrical Impedance Tomography (TDEIT) when the objective is to estimate resistivity changes $\Delta\boldsymbol{\rho} \in \mathbb{R}^n$ given the voltage changes at the electrodes $\Delta\mathbf{v}$ between two instants of time.

Considering a pair-wise current injection, all the electrodes measure the resulting voltages, including the ones used for the injection. Let $e$ be the number of electrodes and $\Delta\mathbf{v}_i \in \mathbb{R}^e$ for $i = 1, ..., e$ be the resulting voltage difference given the electrical current input from the $i$th pair of electrodes. After concatenating $\Delta\mathbf{v}_i$, for all $i$, into a single vector $\Delta\mathbf{v} \in \mathbb{R}^m$, where $m = e^2$, the linearized EIT problem [1] is given by

$$\mathbf{J}(\boldsymbol{\rho}_0)\Delta\boldsymbol{\rho} = \Delta\mathbf{v} + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is an additive noise and $\mathbf{J}(\boldsymbol{\rho}_0) \in \mathbb{R}^{m \times n}$ is the Jacobian matrix $\frac{\partial\left(\mathbf{Y}(\boldsymbol{\rho}_0)^{-1}\mathbf{c}\right)}{\partial\boldsymbol{\rho}}\Big|_{\boldsymbol{\rho}_0}$ resulting from the first-order Taylor expansion of (1) around an initial resistivity $\boldsymbol{\rho}_0$ and calculated from the voltages of the electrode nodes only. From now on, we denote $\mathbf{J}(\boldsymbol{\rho}_0) = \mathbf{J}$.

EIT is a severely ill-posed inverse problem, which means that the discrete linearized problem (2) is ill-conditioned. Thus, regularization is necessary to solve it [1]. Traditional regularization methods include the Tikhonov Regularization and the Total Variation (TV) regularization [3]. These methods may be understood as including prior information to solve the problem [4].

Recently, [5] proposed the Regularization by Denoising (RED), where general inverse problems, other than denoising, are solved with the prior information that the solution presents a small noise. Considering an additive white Gaussian noise $\boldsymbol{\epsilon}$, the resulting functional is given by

$$\Delta\hat{\boldsymbol{\rho}}_{RED} = \arg\min_{\Delta\boldsymbol{\rho}} \left\{ \frac{1}{2\sigma^2}||\Delta\mathbf{v} - \mathbf{J}\Delta\boldsymbol{\rho}||_2^2 + \lambda R(\Delta\boldsymbol{\rho}) \right\}, \tag{3}$$

for

$$R(\Delta\boldsymbol{\rho}) = \frac{1}{2}\Delta\boldsymbol{\rho}^T \left( \Delta\boldsymbol{\rho} - f(\Delta\boldsymbol{\rho}) \right), \tag{4}$$

where $\Delta\hat{\boldsymbol{\rho}}_{RED}$ is the estimated $\Delta\boldsymbol{\rho}$ with RED, $\sigma$ is the noise variance, $\lambda$ is a regularization parameter and $f(\Delta\boldsymbol{\rho})$ is a denoiser. Note that the additive noise is present in $\Delta\mathbf{v}$, but the denoiser is applied in the values of $\Delta\boldsymbol{\rho}$ to be reconstructed.

It is possible to optimize the RED framework functional with different algorithms, such as the steepest descend method, the alternating direction method of multipliers (ADMM), and fixed-point methods. Applications include deblurring, super-resolution [5], nonlinear traveltime tomography [6], and hyperspectral anomaly detection [7].

In this work, we use the RED framework to reconstruct TDEIT resistivity images from simulated data in the region of the head. We compare the reconstruction performance of different denoisers from both the noiseless and noisy $\Delta\mathbf{v}$ cases.

## 2   Materials and Methods

We performed no tests on animals or volunteers. We simulated $\Delta \mathbf{v}$ and implemented the TDEIT inversion in MATLAB, as described below.

### 2.1   2D Mesh Generation

Using MATLAB, we obtained the contours of three head tissues (scalp, skull, and brain) from a cross-section above the eye height of a young adult head model [8]. We generated the 2D mesh with the software GMSH[1] with triangular elements and added 32 electrodes around it. Table 1 presents the information about the two meshes, one to simulate $\Delta \mathbf{v}$ (finer mesh) and the other to solve the inverse problem (coarser mesh). To avoid the inverse crime [4], we used different meshes and resistivities between them, but the electrode positions are the same. A circular target represents an alteration in the ROI. The target properties are the same in all simulations.

**Table 1.** $\Delta \mathbf{v}$ simulation and inverse model properties.

| $\Delta \mathbf{v}$ simulation | Brain | Skull | Scalp | All |
|---|---|---|---|---|
| $\boldsymbol{\rho}[\Omega m]$ | 9.12 | 47.94 | 2.5 | - |
| $\mathrm{N}^o$ of Elements | 8067 | 1459 | 1086 | 10708 |
| $\mathrm{N}^o$ of nodes | - | - | - | 5420 |
| **Inverse problem** | **Brain** | **Skull** | **Scalp** | **All** |
| $\boldsymbol{\rho}_0[\Omega m]$ | 9.5 | 45 | 3 | - |
| $\mathrm{N}^o$ of Elements | 1222 | 226 | 524 | 2068 |
| $\mathrm{N}^o$ of nodes | - | - | - | 1100 |

| Target | Rd.[m] | Center[m] | $\boldsymbol{\rho}[\Omega m]$ |
|---|---|---|---|
| Circular | 0.015 | (0.078, 0.09) | 1.42 |

### 2.2   Forward Problem Solution ($\Delta \mathbf{v}$ Simulation)

Considering the complete electrode model [9], [2, pp. 207-210], we solved the forward problem with the finite element method in the finer mesh. Following [10], the current injection pattern was Skip-16 (a diametral, but with an even skip), with single-ended voltage measurements in relation to the electrical ground in the center of the mesh. Table 2 presents the information of the forward problem.

We simulated two scenarios, the noiseless and noisy signals, generating 10 different realizations of white Gaussian noise $\boldsymbol{\epsilon} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$, where $\boldsymbol{\mu} = \mathbf{0}[V]$ is the mean, $\sigma = 0.003[V]$ is the standard deviation and $\mathbf{I}$ is an identity matrix. To

---

[1] https://gmsh.info/.

**Table 2.** Finite element method equations and parameters.

| Domain | |
|---|---|
| Elements type | Triangular |
| Local conductivity matrix | Eq. A.17 [2] |
| h (Thickness) [m] | 0.03 |
| **Electrode** | |
| Elements type | Triangular |
| Model | Complete [9] |
| Local conductivity matrix | Eq. A.42 [2] |
| b (Thickness) [m] | 0.03 |
| a (Width) [m] | Variable |
| $\boldsymbol{\rho}_t$ (Parameter) [$\Omega m^2$] | 0.02 |
| **General** | |
| Global conductivity matrix | Eq. A.32 [2] |
| Current value [A] | 0.001 |
| Current injection pattern | Skip-16, Eq. 2.35 [2] |
| Jacobian matrix calculation | Eq. 2.45 [2] |

make each $\boldsymbol{\epsilon}$ repeatable for the different methods, we fixed the random number seeds[2].

Figure 1a) shows the forward problem mesh with the three head layers (brain, skull, and scalp) and electrodes around the scalp. Figure 1b) shows the target $\Delta\boldsymbol{\rho}$. Figure 1c) shows an example of the resulting $\Delta\mathbf{v}$ for the noiseless and noisy scenarios, with values for three different skip-16 electrode pairs (96 elements out of 1024) for better visualization.

## 2.3   Inverse Problem Solution

To solve the inverse problem, it is also necessary to solve the forward problem in the coarser mesh and then calculate the Jacobian matrix $\mathbf{J}$ as explained in [2]. We solved the forward problem considering all three tissues, but only the brain elements are updated in the inverse problem, making its dimensionality smaller. Information from Table 2 was also applied for solving the inverse problem.

We considered that the noise variance $\sigma$ was unknown. We set $\frac{1}{2\sigma^2} = 1$ in (4), which allows $\lambda$ to be the only responsible for controlling the bias-variance trade-off. Then, we visually chose the regularization parameter $\lambda$.

An implementation of RED by Choi and Poullet is available online[3], under the Apache License 2.0, and we adapted it to the EIT problem. To solve (3) when $R(\Delta\boldsymbol{\rho})$ is given by (4), we chose the fixed-point method [5] with 10 inner

---

[2] In MATLAB, the seeds 1 through 10 in the *rng* function.
[3] https://github.com/google/RED.

a) $\rho[\Omega m]$

b) $\Delta\rho[\Omega m]$

c) Resulting noiseless and noisy $\Delta\mathbf{v}$ [V]
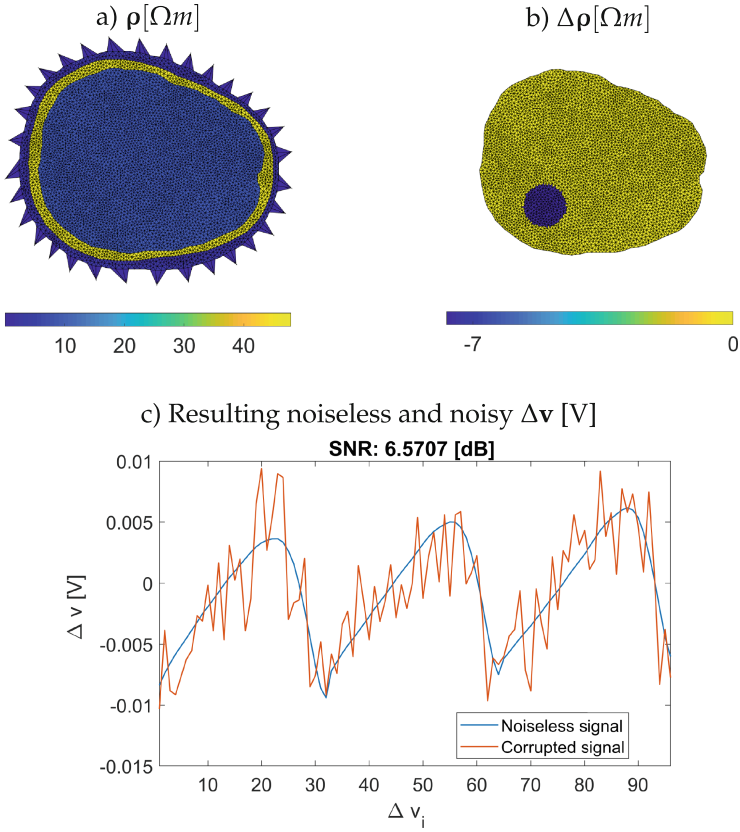
**Fig. 1.** Forward problem. a) Head mesh resistivities. b) Brain mesh with $\Delta\rho$ values, including the circular target. c) Noiseless and noisy $\Delta\mathbf{v}$ for a single realization of the white Gaussian noise $\boldsymbol{\epsilon}$. SNR is the signal-to-noise ratio of $\boldsymbol{\epsilon}$ in relation to the noiseless signal.

iterations and a zero vector as the initial estimate. The number of outer iterations depended on the following stopping criterion to evaluate the solution changes in the last three steps:

Let $i$ be the iteration number. For $i > 3$, if

$$\frac{1}{3}\left\{||\Delta\boldsymbol{\rho}_i - \Delta\boldsymbol{\rho}_{i-1}|| + ||\Delta\boldsymbol{\rho}_{i-1} - \Delta\boldsymbol{\rho}_{i-2}|| + ||\Delta\boldsymbol{\rho}_{i-2} - \Delta\boldsymbol{\rho}_{i-3}||\right\}$$

is smaller than a tolerance $5e - 08$, the iterations end.

Instead of denoising directly the mesh elements values, we defined all denoisers in the image domain by linearly interpolating the reconstructed data into a regular grid of a 2D image, with $35 \times 35$ pixels, considering null values outside the ROI. Thus, the number of pixels was in the same order as the number of the

brain mesh elements. Then, we applied the denoiser and linearly interpolated the values back to the mesh.

Each denoiser is a built-in MATLAB function. Table 3 shows the parameters of all the methods, such as the boundary conditions (BC), the size of the neighborhood (NHood) of the median filter, or the size of the Gaussian kernel, to name a few. Denoisers from $N^{o}1$ to 3 are edge-preserving denoisers (depending on the parameters), while method $N^{o}4$ is a smoothing operator.

## 2.4   Performance Comparison

We compared the RED results with the generalized Tikhonov regularization [1, Section 1.5] based on the functional given by

$$\Delta\hat{\boldsymbol{\rho}}_{Tikh} = \arg\min_{\Delta\boldsymbol{\rho}} \left\{ ||\Delta\mathbf{v} - \mathbf{J}\Delta\boldsymbol{\rho}||_2^2 + \lambda||\mathbf{L}\Delta\boldsymbol{\rho}||_2^2 \right\}, \tag{5}$$

with the closed-form solution

$$\Delta\hat{\boldsymbol{\rho}}_{Tikh} = \left( \mathbf{J}^T\mathbf{J} + \lambda\mathbf{L}^T\mathbf{L} \right)^{-1} \mathbf{J}^T\Delta\mathbf{v}, \tag{6}$$

where the regularization matrix $\mathbf{L} \in \mathbb{R}^{n \times n}$ is a Gaussian high-pass filter (HPF) [2, p. 115]. The filter standard deviation can be associated with prior information regarding the smoothness of the solution.

We also solved the problem with the isotropic total variation (TV) regularization [11,12] to reconstruct stepwise constant images by minimizing the following functional

$$\Delta\hat{\boldsymbol{\rho}}_{TV} = \arg\min_{\Delta\boldsymbol{\rho}} \left\{ ||\Delta\mathbf{v} - \mathbf{J}\Delta\boldsymbol{\rho}||_2^2 + \lambda||\mathbf{L}_d\Delta\boldsymbol{\rho}||_1 \right\}, \tag{7}$$

where the regularization matrix $\mathbf{L}_d$ is a discrete first derivative operator [12] and its size depends on the number of edges between the elements.

We approximated the $\ell_1$-norm using the iteratively reweighted least squares (IRLS) method [13, p. 197]. This method also depends on a tolerance value, denoted $tol_{IRLS}$ to differ from the tolerance of the stopping criterion. Table 3 also shows the relevant parameters of these methods.

## 2.5   Performance Evaluation

We compared the number of iterations necessary to respect the stopping criterion in the noiseless and noisy cases.

We also compared the resulting images with the numerical phantom of the forward problem using figures of merit considering the one-fourth amplitude set. We calculated the following figures of merit: the position error (PE), which compares the center of the target to that of the reconstructed image; Ringing (RNG), which measures the opposite sign (in this case, increasing) resistivity in the reconstruction around the target; Resolution (RES), that compares the size of the reconstructed target to the size of the entire ROI; and Shape deformation

**Table 3.** Denoiser and reconstruction parameters.

| N° | Regularization | Information |
|---|---|---|
| 1 | RED with Median filter [14] | Zero BC<br>$\lambda = 1e - 6$<br>NHood size: [5x5] |
| 2 | RED with Non-local means [15] | $\lambda = 1e - 6$<br>Degree of smoothing: 1 |
| 3 | RED with Bilateral filter [16] | Replicate BC<br>$\lambda = 1e - 6$<br>Degree of smoothing: 0.1<br>Spatial Sigma: 25<br>NHood size: [11x11] |
| 4 | RED with Gaussian filter [14] | Zero BC<br>$\lambda = 1e - 6$<br>Kernel size: [5x5]<br>Kernel std. deviation: 1 |
| 5 | Generalized Tikhonov (One step) | $\mathbf{L}$ = High Pass Filter<br>Filter std. deviation $= 5e - 3$<br>$\lambda = 1e - 6$ (noiseless)<br>$\lambda = 1e - 5$ (noisy) |
| 6 | Total Variation (Iterative) | $\mathbf{L}_d$ = First derivative<br>$tol_{IRLS} = 5e - 3$<br>$\lambda = 4e - 3$ (noiseless)<br>$\lambda = 8e - 3$ (noisy) |

(SD), which compares the original target shape (circular, in this study) to the target reconstruction shape.

For all figures of merit, a smaller absolute value indicates improved results. PE, RNG, and SD could result in zero values. RES is a positive value that depends on the target radius. For further information on calculation and interpretation, we ask the reader to refer to [17]. In the case of the reconstructions from the 10 different noise realizations, we calculated the mean and the standard deviation of each figure of merit.

## 3  Results

Figure 2a–f) shows the reconstructed images in the noiseless case. Table 4 presents their figures of merit. Both TV regularization (N°6) and RED with NLM (N°2) present the lowest RES and SD values. RED with NLM (N°2) also presents the lowest RNG, but the number of iterations, regarding the adopted stopping criterion, is a drawback.
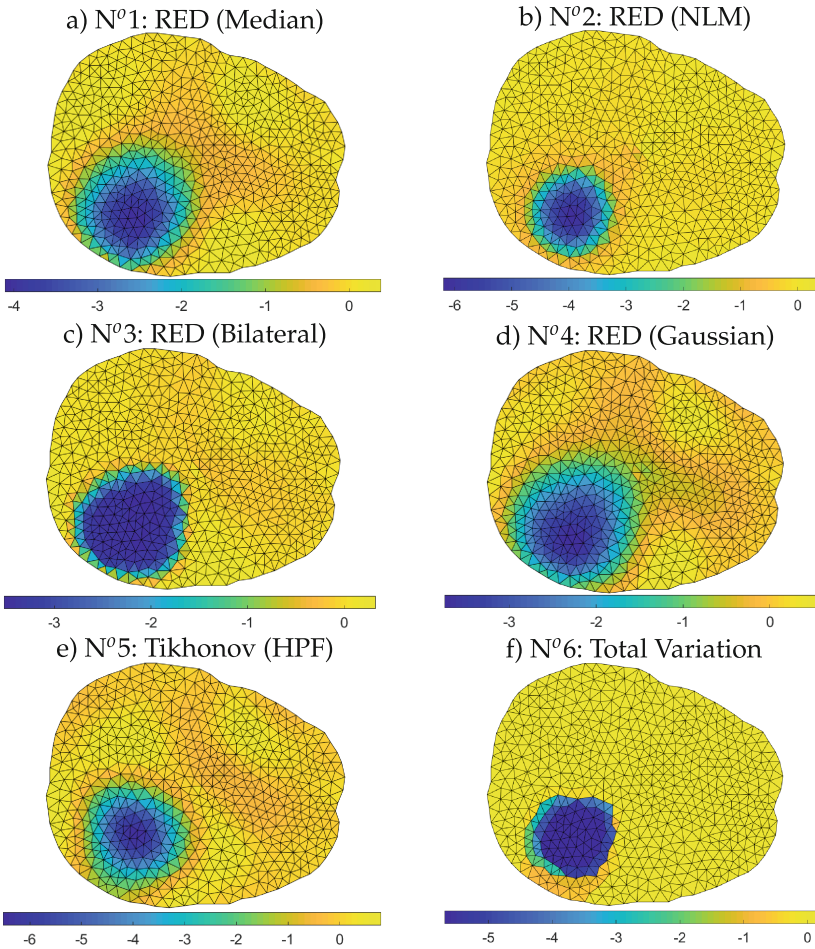
**Fig. 2.** Reconstructions $\Delta\boldsymbol{\rho}[\Omega m]$ obtained from the noiseless signal $\Delta\mathbf{v}$.

Figure 3a–f) shows the reconstructed images with the noisy signal. Table 5 presents their figures of merit and the number of iterations that results from this stopping criterion. It was necessary to use a higher value of $\lambda$ with methods $N^o5$ and $N^o6$, as shown in Table 3, to decrease artifacts. Nevertheless, the noise seemed to affect more traditionally reconstructed images than the RED images, as seen by the standard deviation of each figure of merit. RED with NLM ($N^o2$) presented the best PE-x, RNG, RES, and SD results, as the resulting image was quite similar to that of the noiseless case.
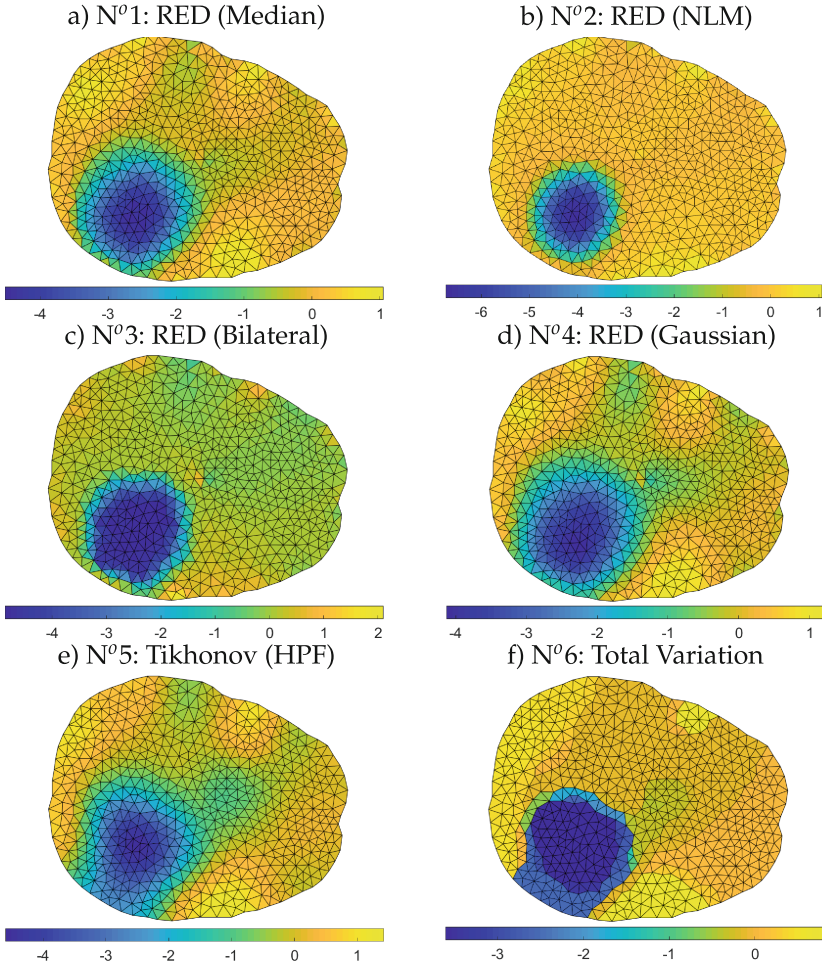
**Fig. 3.** Reconstructions $\Delta\boldsymbol{\rho}[\Omega m]$ obtained from the noisy signal $\Delta\mathbf{v}$.

## 4 Discussion

Different denoiser choices imply assuming different prior information. By visual inspection of Figs. 1 and 2, varying the methods can yield sharper solutions (especially N°6, as expected) or smoother solutions. To improve the results, a possibility is to find the optimal denoiser parameters. We can also choose other denoisers, such as those based on machine learning and deep learning. The authors in [5] discussed guarantees of convergence when the denoiser presents properties such as local homogeneity, strong passivity, and differentiability. These properties must be verified when using other denoisers so that there are theoretical guarantees about convergence.

**Table 4.** Noiseless signal: Figures of merit and number of iterations.

| N$^o$ | PE [mm] | RNG | RES | SD | Iter. |
|---|---|---|---|---|---|
| Ref. | (0, 0) | 0 | 0.21 | 0 | N/A |
| 1 | (0.6, 2.7) | -0.32 | 0.44 | 0.78 | 83 |
| 2 | **(-0.2, -1.2)** | **-0.035** | **0.33** | **0.59** | 234 |
| 3 | (1.3, 1.6) | -0.20 | 0.44 | 0.77 | 66 |
| 4 | (1.3, 3.3) | -0.48 | 0.47 | 0.80 | 45 |
| 5 | (1.1, 3.4) | -0.50 | 0.37 | 0.69 | 1 |
| 6 | (-0.8, 2.7) | -0.091 | **0.33** | **0.59** | 23 |

**Table 5.** Noisy signal: Figures of merit and number of iterations. Values are represented as mean ± standard deviation.

| N$^o$ | PE-x [mm] | PE-y [mm] | RNG |
|---|---|---|---|
| Ref. | 0 | 0 | 0 |
| 1 | 2.8 ± 2.1 | 3.9 ± 2.8 | -0.66 ± 0.18 |
| 2 | **-0.1 ± 1.8** | -2.6 ± 3.4 | **-0.22 ± 0.06** |
| 3 | 2.2 ± 1.3 | **1.9 ± 1.5** | -0.58 ± 0.19 |
| 4 | 3.9 ± 2.0 | 4.5 ± 2.0 | -0.86 ± 0.19 |
| 5 | 5.2 ± 3.7 | 6.0 ± 3.8 | -1.03 ± 0.32 |
| 6 | 3.0 ± 3.1 | 2.6 ± 5.0 | -0.76 ± 0.36 |
| N$^o$ | RES | SD | Iter. |
| Ref. | 0.21 | 0 | N/A |
| 1 | 0.46 ± 0.03 | 0.79 ± 0.03 | 75 ± 7 |
| 2 | **0.33 ± 0.03** | **0.60 ± 0.04** | 199 ± 52 |
| 3 | 0.43 ± 0.04 | 0.77 ± 0.04 | 93 ± 30 |
| 4 | 0.48 ± 0.03 | 0.81 ± 0.02 | 45 ± 5 |
| 5 | 0.50 ± 0.05 | 0.82 ± 0.03 | 1 |
| 6 | 0.47 ± 0.06 | 0.77 ± 0.06 | 59 ± 16 |

Regarding Tables 4 and 5, the number of iterations depends on the method, specifically on the denoiser. All methods achieved better performance with fewer iterations in the noiseless case. With the noisy signals, the performance of all methods was worse and the number of iterations tended to be higher, as expected. In both noise scenarios, while method N$^o$2 (RED with NLM) obtained the best results, it required more iterations to satisfy the stopping criterion than all the other methods, hindering real-time applications.

Concerning the RED framework, we projected the mesh results to an image of size $35 \times 35$ pixels to denoise with known denoisers from image processing, then projected back the results to the mesh. Other interpolation methods are

possible, such as the nearest neighbors. Another way is to denoise $\Delta\boldsymbol{\rho}$ directly in the mesh, without any intermediate interpolation step to an image grid.

Finally, in the context of TDEIT, we assumed the availability of the reference measure of the healthy patient, necessary to obtain $\Delta\boldsymbol{\rho}$. If the patient presents pre-existing conditions by the time the monitoring begins, TDEIT may not be suitable. Thus, a research direction is to investigate the use of RED to reconstruct absolute EIT images, that is, to reconstruct $\boldsymbol{\rho}$ directly, using the voltage measurements from a single instant of time.

## 5    Conclusion

In this work, we proposed to use the Regularization by Denoising method to reconstruct TDEIT brain images in a simplified 2D head mesh. By varying the denoiser and its parameters, it is possible to obtain smooth or sharp target reconstructions, providing flexibility to the method. In the noiseless case, RED with non-local means was the best reconstruction method, but both the Tikhonov and TV regularization were also satisfactory. In the noisy scenario, while the traditional reconstruction methods were sensitive to noise, RED with non-local means as the denoiser achieved better results in relation to the figures of merit we evaluated.

Future works include: solving the three-dimensional TDEIT problem using RED; denoising directly the mesh elements value, without projecting the problem to the image domain; automatic denoiser parameters selection; and using deep learning based denoisers.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Holder, D.: Electrical Impedance Tomography: Methods, History and Applications, 1 edn. CRC Press (2004). ISBN 978-0750309523
2. Moura, F.S.: Estimação não linear de estado através do unscented Kalman filter na tomografia por impedância elétrica. Ph.D. thesis, Polytechnic School of the University of São Paulo (2013)
3. Hansen, P.C.: Discrete Inverse Problems: Insight and Algorithms. Society for Industrial and Applied Mathematics, Philadelphia (2010). ISBN: 978-0-898716-96-2
4. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Springer, New York (2005). https://doi.org/10.1007/b138659
5. Romano, Y., Elad, M., Milanfar, P.: The little engine that could: regularization by denoising (RED). SIAM J. Imag. Sci. **10**(4), 1804–44 (2017). https://doi.org/10.1137/16m1102884

6. Vargas, A.: Regularization by denoising applied to non-linear traveltime tomography. Master's thesis, University of Alberta (2020)
7. Fu, X., Jia, S., Zhuang, L., Xu, M., Zhou, J., Li, Q.: Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization. IEEE Trans. Geosci. Remote Sens. **59**(11), 9553–68 (2021). https://doi.org/10.1109/TGRS.2021.3049224
8. Hammond, D., Price, N., Turovets, S.: Construction and segmentation of pediatric head tissue atlases for electrical head modeling. In: OHBM, Vancouver, Canada, June 2017
9. Cheng, K.-S., Isaacson, D., Newell, J.C., Gisser, D.G.: Electrode models for electric current computed tomography. IEEE Trans. Biomed. Eng. **36**(9), 918–24 (1989). https://doi.org/10.1109/10.35300
10. Silva, O.L., Lima, G., Martins, T.C., Moura, F.S., Tavares, R.S., Tsuzuki, M.S.G.: Influence of current injection pattern and electric potential measurement strategies in electrical impedance tomography. Control. Eng. Pract. **58**, 276–86 (2017). https://doi.org/10.1016/j.conengprac.2016.03.003
11. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–68 (1992). https://doi.org/10.1016/0167-2789(92)90242-f
12. Borsic, A., Graham, B.M., Adler, A., Lionheart, W.R.B.: In vivo impedance imaging with total variation regularization. IEEE Trans. Med. Imaging **29**(1), 44–54 (2010). https://doi.org/10.1109/TMI.2009.2022540
13. Aster, R.C., Borchers, B., Thurber, C.H.: Parameter Estimation and Inverse Problems, 3rd edn. Elsevier, Amsterdam (2019). ISBN: 9780128134238
14. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 4th edn. Pearson, New York, NY (2018). ISBN: 978-1-292-22304-9
15. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 2, pp. 60–65 (2005). https://doi.org/10.1109/CVPR.2005.38
16. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271), pp. 839–846 (1998). https://doi.org/10.1109/ICCV.1998.710815
17. Adler, A., et al.: GREIT: a unified approach to 2D linear EIT reconstruction of lung images. Physiol. Meas. **30**(6), S35 (2009). https://doi.org/10.1088/0967-3334/30/6/S03

# Design, Simulation and Analysis of a MATLAB/Simulink Based Delay-and-Sum Beamforming Model for Ultrasound Imaging

Gilson Maekawa Kanashiro[1,2]([✉]) [ID], Michel Andrey Freitas de Souza Kohler[1] [ID],
Ednilson de Souza Contieri[1] [ID], Rojelio de Bairro[1] [ID], Larissa Comar Neves[1] [ID],
Thiago Mathias Oliveira[1] [ID], Amauri Amorin Assef[1,3] [ID],
Joaquim Miguel Maia[1,4] [ID], and Eduardo Tavares Costa[5] [ID]

[1] Graduate Program in Electrical and Computer Engineering (CPGEI),
Federal University of Technology - Paraná (UTFPR), Curitiba, Brazil
`gilson.kanashiro@ifpr.edu.br`
[2] Electrical Engineering (Undergraduate Program), Federal Institute of Paraná
(IFPR), Paranavaí, Brazil
[3] Academic Department of Electrical Engineering (DAELT), UTFPR, Curitiba,
Brazil
[4] Biomedical Engineering Program (PPGEB) and Academic Department of
Electronics Engineering (DAELN), UTFPR, Curitiba, Brazil
[5] DEEB-FEEC and CEB, University of Campinas (UNICAMP), Campinas, Brazil

**Abstract.** In ultrasound research, an important subject is the quality of
image assessment. Despite a myriad of technological solutions, qualitative
and quantitative metrics are critical indicators to measure and compare
the quality of ultrasound images. This paper presents a simulation of a
model that implements step-by-step an image reconstruction based on
delay-and-sum (DAS) method from the Matlab/Simulink environment.
The general objective was to validate this model to compare the qual-
ity metrics of images against the computational one. Model results were
validated using the Field II as gold reference, a program for simulating
ultrasound transducer fields and ultrasound imaging using linear acous-
tics. The model implements DAS beamforming to process 65 scanlines
generated by a 128-element transducer. For quantitative analysis, the
following metrics have been used: contrast resolution (CR), contrast-to-
noise (CNR), signal-to-noise ratio (SNR), normalized root mean square
error (NRMSE), normalized residual sum of squares (NRSS), and full
width at half maximum (FWHM). According to those metrics, the pro-
posed model performance is strongly correlated comparatively to the ref-
erence, showing errors of 1.49%, 1.29%, 0.22%, 2.45%, 7.38% for the CR,
CNR, SNR, FWHM, and NRMSE, respectively. As indicated by met-
rics the Matlab/Simulink model images are similar to the computational
ones.

**Keywords:** Ultrasound · delay-and-sum · B-mode image · quality
indicators · image quantifiers

# 1    Introduction

Ultrasound (US) medical images are non-invasive, free of ionizing radiation, and relatively cost-effective compared with x-ray systems. For these reasons, US technology has become widely used in diagnosis medicine.

In US research, it is necessary to assess the quality of an image despite of a myriad of technological solutions. As these various solutions address image quality comparison problems, studies on qualitative and quantitative metrics are justified as critical indicators to measure and compare the quality of US images.

The objective of this work was to validate a simulated model by comparing its image quality metrics against the computational model. Simulated model results were validated using the Field II [7] as gold reference, a program for simulating ultrasound transducer fields and ultrasound imaging using linear acoustics.

First, a computational (Matlab/Field II) model implemented a step-by-step image reconstruction. Next, a simulated (Simulink) model do the same approach to be validated. Both implementations were based on delay-and-sum (DAS) method to reconstruct a US image [6,15]. This technique is considered well-established. However, to improve the image quality and accuracy of medical diagnosis, the raw radio-frequency (RF) signal needs to be manipulate directly [2], customizing the simulation of signal processing in real-time. More than that, it is necessary to measure the quality of images generated.

The functional block diagram Fig. 1 shows a general overview of the computational and simulated models. Next, the diagram and the US image processing steps will be explained briefly.
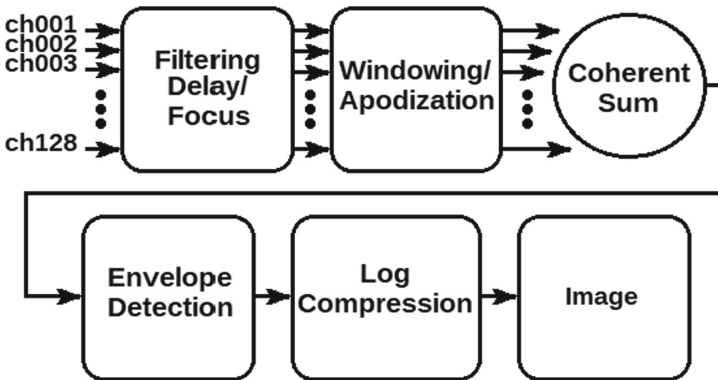


**Fig. 1.** Receive beamformer - functional steps to obtain DAS B-mode images

It illustrates the main steps of DAS beamforming signal processing for US images: filtering, delay/focalization, windowing/apodization, coherent summation, envelope detection, logarithmic compression, and image composition.

About that mentioned computational model, to simulate the transmission/reception, the aperture for emission/reception, the impulse response, and the excitation were configured by Field II functions.

Special attention was paid to the backscattered signal obtained by reception process. So with Matlab code and tools with Field II functions, it was implemented Hanning windowing focalization, filtering, coherent summation, envelope detection, logarithmic compression and image forming by scanlines juxtaposition.

## 2  Methods

Simulations have been done in Field II [7], a program for simulating ultrasound transducer fields and ultrasound imaging using linear acoustics, to reconstruct a 20-target axial image with focalization at 50 mm depth. For transmission, the parameters were set up considering a 128-element linear array transducer available in our laboratories. Table 1 shows usual units of measurement. The "un" symbol means quantities of units or elements.

**Table 1.** Transducer parameters and excitation data

| Parameter | Values | Units |
|---|---|---|
| Linear array transducer | 128 | $un$ |
| Active elements in the transducer | 64 | $un$ |
| Transducer center frequency | 3.2 | $MHz$ |
| Scanlines generated | 65 | $un$ |
| Sampling frequency | 40 | $MHz$ |
| Speed of sound (soft tissue) | 1540 | $m.s^{-1}$ |
| Wavelength ($\lambda$) | 481.25 | $\mu m$ |
| Width of element | 481.25 | $\mu m$ |
| Height of element | 5 | $mm$ |
| Kerf | 8 | $\mu m$ |
| Fixed focal point $(x, y, z)$ | [0,0,50] | $mm$ |
| Elements in $x$-direction | 1 | $un$ |
| Elements in $y$-direction | 5 | $un$ |
| Targets simulated with phantom | 20 | $un$ |
| Axial target separation | 10 | $mm$ |
| Apodization | Hanning | - |

A DAS model to generate the US image was built using Matlab/Simulink, whose quality was evaluated. There are two major tasks to reach the main goal of this work: the reconstruction of the B-mode image and the appliance of metrics to assess the image's quality.

## 2.1   B-Mode Image Reconstruction

A 128-element transducer, with an aperture of 64 elements, has been used to receive scattered echoes (raw RF signal) to reconstruct US images. The received signals are processed using a low-pass filter (LP_Filter, Fig. 2) and delayers ($Z^{-1}$, Fig. 2) to provide a more precise and the correct focalization. The same approach described in [3] has been used to implement the LPF.

Next, a signal processing windowing technique, the apodization (Apod, Fig. 2) eliminates distortions caused by the lateral channels of the aperture. Here, a symmetric aperture was applied using the delay and apodization coefficients calculated by Field II.

The coherent summation, an adding operation of the 64 resulting signals $x(:, 1)$ to $x(:, 64)$ was executed recursively to generate the 65 scanlines (128 transducer's elements - 64 aperture's elements + 1). Figure 2 illustrates these steps performed by 64 channels signal processing.
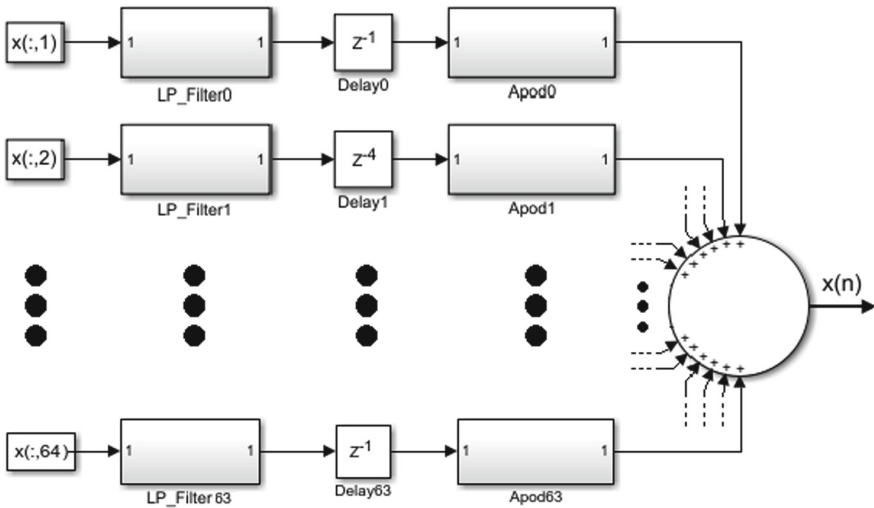


**Fig. 2.** Ultrasound DAS beamforming method. Block diagram illustrating the 64-channels of the first steps (acquisition, filtering, delay, apodization, and coherent summing)

The envelope detector will apply a Hilbert Transform (FIR Filter Hilbert Transformer, Fig. 3) to perform quadrature demodulation and determine an envelope curve ($E(n)$, Fig. 3) from an input RF signal ($x(n)$, Fig. 3). This detection corresponds to a tangent curve that involves each scanline $x(n)$.

As found in [10], envelope detection can be performed by modeling a Hilbert filter, using finite impulse response (FIR) techniques to obtain the phase $I(n)$ and quadrature $Q(n)$ components. A block diagram (Fig. 3) illustrates the concept behind these steps, whose details can be found in [1].
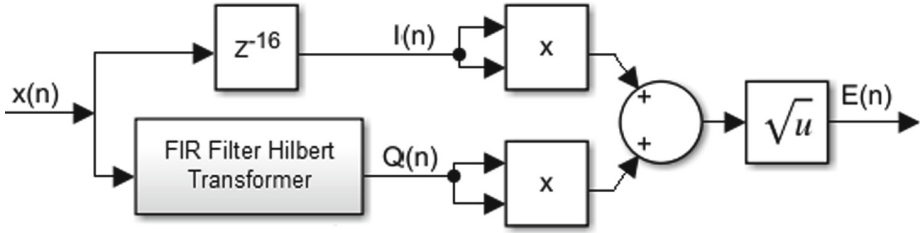
**Fig. 3.** Functional block diagram of envelope detector (quadrature demodulation)

Usually, the dynamic range of the signal is restricted, as in [2] and [4], because of a limitation to viewing grayscale. The dynamic range of human eyes can distinguish a 30 dB order. The dynamic range used in this work was restricted from $-60$ dB to 0 dB and the grayscale gradients from 0 (white) to 128 or 256 (black) with a so-called logarithmic compression. Additionally, this compression improves the contrast of the US image.

The rest of the process is related to conforming the generated image to the screen. A juxtapose operation is sufficient for the linear array transducer used to exhibit a B-mode image.

## 2.2 The B-Mode Image Assessment

US B-mode images can generated following the steps described in the last section and illustrated by Fig. 1. For quantitative analysis, following metrics have been used: contrast resolution (CR), contrast-to-noise (CNR), signal-to-noise ratio (SNR), normalized root mean square error (NRMSE), normalized residual sum of squares (NRSS), and full width at half maximum (FWHM). Those metrics were evaluated after logarithmic compression to assess the overall image processing. Moreover, the region of interest (ROI) was defined at the target located at 50 mm and its neighborhood where applicable. Only the central scanline was considered for metrics dependent on scanlines to calculate the image quality metrics.

A straightforward notion about these quality image quantifiers is described in the following lines.

**Contrast Metrics.** Contrast is a measure of the ability to distinguish luminance between different regions of the displayed images [6]. Logarithmic compression, image memory, dynamic range, and contrast agents are variables that influence this parameter. In this work, CR (1) was employed as defined in [16], and CNR (2) as defined in [11]. These two contrast quantifiers were applied like [1,18] using Eqs. 1 and 2:

$$CR = |\mu_{tgt} - \mu_{bck}|, \tag{1}$$

$$CNR = \frac{|\mu_{tgt} - \mu_{bck}|}{\sqrt{\sigma_{tgt}^2 + \sigma_{bck}^2}}, \tag{2}$$

where $\mu_{tgt}$ e $\mu_{bck}$ are mean intensities of target ($tgt$) and background ($bck$) regions, respectively. The $\sigma_{tgt}$ and $\sigma_{bck}$ are standard-deviations of the signal intensities in $tgt$ and $bck$ regions.

**Signal-to-Noise Ratio.** The signal-to-noise ratio (SNR) compares the signal and the noise levels at a determined ROI. Its general goal is to quantify the image quality through speckle measurement [13]. Speckles are a kind of artifact (alteration) that accentuates the granular aspect, making the image resultant not correspond to the target's actual image.

A well-known approach to measuring SNR is through the mean intensity ($\mu_s$) and the standard deviation ($\sigma_s$) of the signal [9,14], defined as:

$$SNR = \frac{\mu_s}{\sigma_s} = \frac{\frac{1}{mn} \cdot \sum_{j=1}^{m} \sum_{i=1}^{n} E(i,j)}{\sqrt{\frac{1}{mn-1} \cdot \sum_{j=1}^{m} \sum_{i=1}^{n} \left( E(i,j) - \frac{1}{mn} \cdot \sum_{j=1}^{m} \sum_{i=1}^{n} E(i,j) \right)^2}} \tag{3}$$

where $E(i,j)$ is the intensities of signals, $i$ and $j$ are a row and a column within an ROI, respectively.

**Evaluating the Goodness of Fit.** In regression analysis, it is common to measure the strength of the relationship between data. In our experiments, data are relative to numerical computation and simulated model one. The normalization data operation is to guarantee the same proportion.

In this work, the normalized residual sum of squares (NRSS) was used as a technique that quantifies the amount of error variance in the dataset. So, as residual summing, the nearer to zero is the NRSS metric, the best fit it is [5].

$$NRSS = \frac{\sum_{k=0}^{M-1} |E(k) - ht(k)|^2}{\sum_{k=0}^{M-1} |ht(k)|^2} \tag{4}$$

The normalized root mean square error (NRMSE) or normalized root mean square deviation (NRMSD) is the standard deviation of residuals (error prediction) and measures the average squared error between data. Although smaller NRMSE reflects greater accuracy, it is essential to remark that there is no best value for it [12]. The comparison of the computational and modeled system, the NRSS [17], and the NRMSE [8] were employed with algorithms similar to [1],

$$NRMSE = 100 \cdot \sqrt{\frac{\sum_{k=0}^{M-1} |E(k) - ht(k)|^2}{\sum_{k=0}^{M-1} \left|ht(k) - \overline{ht}\right|^2}} \tag{5}$$

where $ht$ and $\overline{ht}$ are the simulated envelope and the average value of it, respectively, $E(k)$ is the golden standard and $M$ is the number of samples.

**Geometric Distortion.** The quantifier called full width at half maximum (FWHM) measures the geometric distortion. In this work, the FWHM reports the width between two points belonging to the central scanline curve, whose line intercepts half of its maximum value, or better said, at $-6$ dB from the maximum of the main lobe [4]. Figure 4 shows the concept of FWHM. Two segment points are located at half amplitude $A/2$, in dB, $x_1$ and $x_2$, whose length is FWHM.
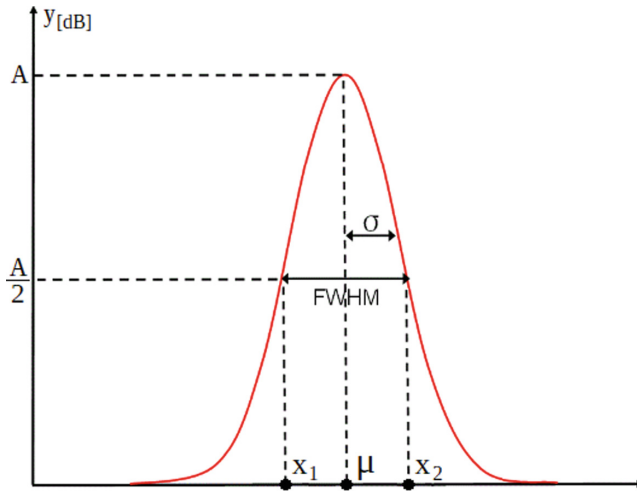


**Fig. 4.** Standard deviation $\sigma$ and full width at half maximum $FWHM$

## 3   Results and Discussion

The proposed method (modeled system) was implemented to generate the image through all the steps illustrated in Fig. 1.

This work integrates image quality metrics found in different articles, each of them assessing some property of US image. Because of that, observations about this set of metrics were parted by subjects. CR and CNR are employed according to [11, 16], respectively, and their application are according to [1, 18]. SNR is widely applied, and the current approach is like [9, 14]. The goodness of fit, NRSS and NRMSE was applied like [1]. Frequently, geometric distortion is considered. So FWHM was applied following [4] approach.

Considering the qualitative analysis, the envelope and the amplitude plot of the central scanline (of number 32) at the 50 mm target (Fig. 5a) was evaluated.
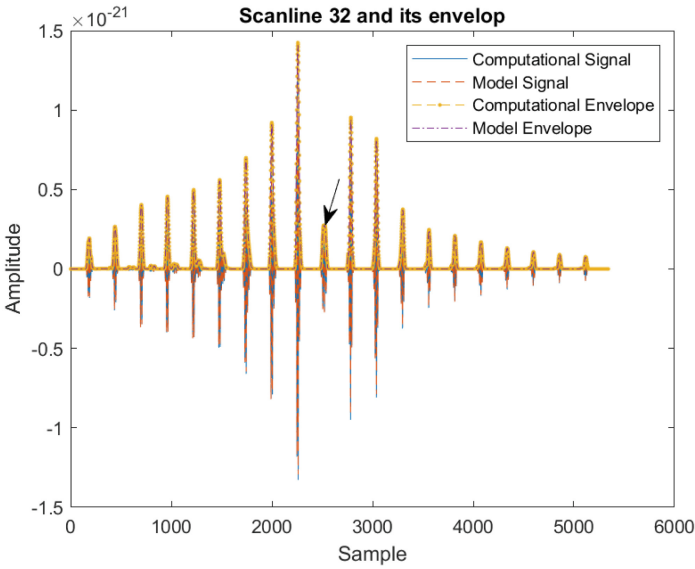
(a) The 20-targets scanline signal and its envelope
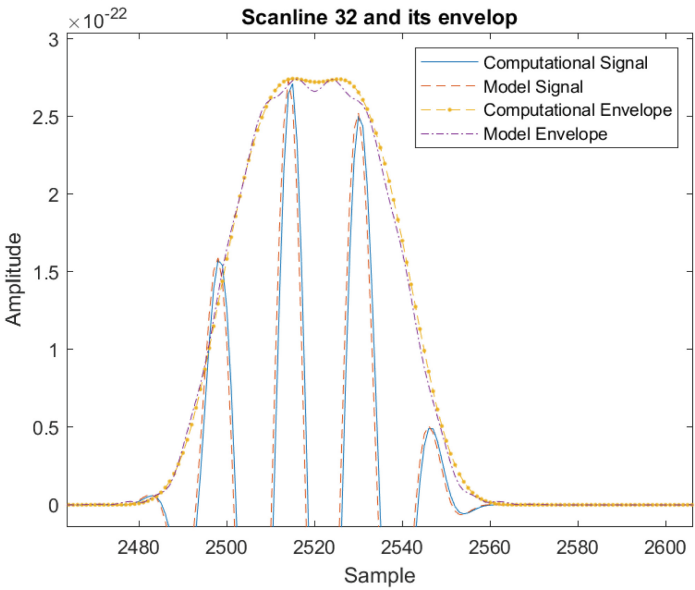


(b) Zoomed view of the signals indicated by the arrow

**Fig. 5. a** Scanline and envelope **b** Zoomed signal and envelope details of computational and modeled system showing similarity

Signals and their envelopes from the computational and modeled system show an excellent fit adjustment, as seen in Fig. 5b, at the location pointed by an arrow in Fig. 5a.

Another approach is comparing both generated US images. Here, the similarity of both images is remarkable, as shown in the resulting image of Fig. 6, where the accuracy of the model is quite evident even if considering the 20-targets. The squares with yellow (left) and blue (right) borders highlight the target and background areas, respectively. The target area and neighborhood were zoomed in for careful qualitative analysis. In Fig. 7, there is an arrow to indicate the slight difference related to the golden standard.
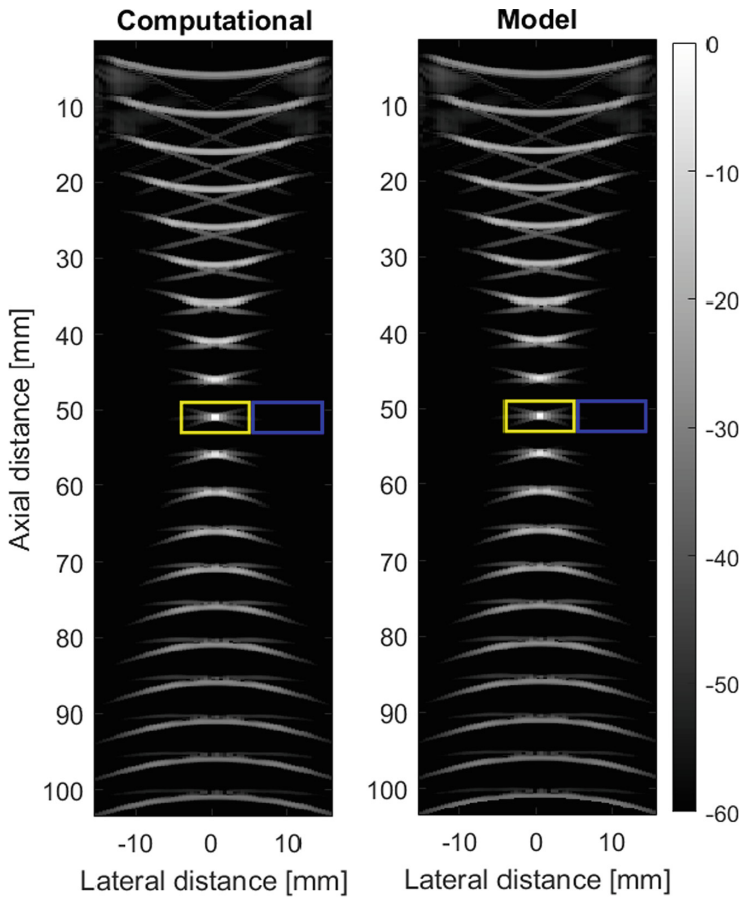


**Fig. 6.** Image qualitative evaluation. The model is similar to the computational one inside the corresponding rectangles

The proposed method results were compared to the computational Field II solution. After applying filtering in the raw signal, delays, apodization, and coherent summation, the envelope curve of the signal was generated.
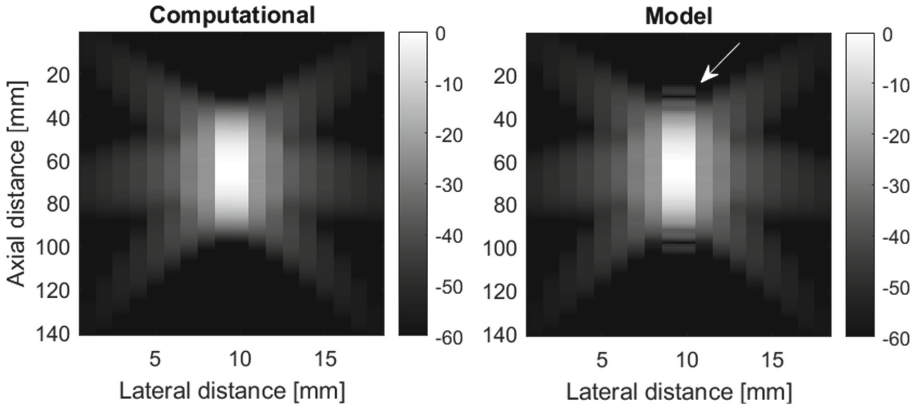


**Fig. 7.** Qualitative analysis of accuracy. Detail (arrow) of the model target showing slight difference

For the qualitative analysis, the goodness of fit was exhibited by Fig. 5a and 5b. This accuracy characteristic can be seen in Fig. 6, whose quality is detailed in a zoomed Fig. 7. For the quantitative analysis, relative to Table 2, it is noticeable that CR, CNR, SNR, and FWHM have very similar values between computational data and the model. The CR and CNR contrast metrics are better in the model as they are the greatest. The NRMSE shows a low value (less than 8%), which is considered an excellent approximation, and in particular, the near-zero NRSS refers to a good fit. Both computational and modeled systems show an excellent agreement.

**Table 2.** Image quality evaluation metrics

| Metrics | Methods | | |
|---|---|---|---|
| | Computational | Modeled | Error (%) |
| CR | 10.01 dB | 10.16 dB | 1.49% |
| CNR | 0.77 dB | 0.78 dB | 1.29% |
| SNR | $-6.20$ | $-6.19$ | 0.22% |
| FWHM | 0.55 mm | 0.56 mm | 2.45% |
| Goodness of fit metrics | | | Values |
| NRMSE (Computational versus Modeled) | | | 7.38% |
| NRSS (Computational versus Modeled) | | | $3.36 \times 10^{-4}$ |

Table 2 summarizes the results to be analyzed. The first column shows the image quality metrics. The second and third columns show values relative to the computational simulation and the modeled system from Matlab/Simulink. The last column shows the error percentage between the assessment percentual values of the systems. The two last lines show the goodness of fit metrics, calculated for both systems. Note that Table 2 groups apart NRMSE and NRSS because these metrics include both computational and modeled system parameters in their formulation.

## 4    Conclusion

This work compared two simulations: a computational implemented with Field II add-on for Matlab and a modeled one made in Matlab/Simulink. A DAS beamforming model generates US scanlines and images in both systems. Some image quality metrics were adopted to evaluate the signals and images. For qualitative and quantitative assessment, the results of the modeled system were compared to the computational simulation (reference). Both computational and modeled systems show an excellent agreement.

This fact indicates that the modeling method in Matlab/Simulink could shorten the image quality assessment. Moreover, it was possible to model an entire DAS beamforming image with a minimum quality degradation of US images.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Assef, A.A., de Oliveira, J., Maia, J.M., Costa, E.T.: FPGA implementation and evaluation of an approximate Hilbert transform-based envelope detector for ultrasound imaging using the DSP builder development tool. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2813–2816 (2019). https://doi.org/10.1109/EMBC.2019.8857671
2. Assef, A.A., Maia, J.M., Costa, E.T.: Initial experiments of a 128-channel FPGA and PC-based ultrasound imaging system for teaching and research activities. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5172–5175 (2016). https://doi.org/10.1109/EMBC.2016.7591892
3. Assef, A.A., et al.: Modeling of a simple and efficient cascaded FPGA-based digital band-pass FIR filter for raw ultrasound data. In: Costa-Felix, R., Machado, J.C., Alvarenga, A.V. (eds.) XXVI Brazilian Congress on Biomedical Engineering. IP, vol. 70/1, pp. 501–505. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-2119-1_77

4. Da Silva Ruzyk, M.J., Martinez, A.C., Assef, A.A., De Oliveira, L.R., Maia, J.M., Costa, E.T.: Comparison of three simulation methods for RF ultrasound signal envelope detection. In: 2021 IEEE UFFC Latin America Ultrasonics Symposium (LAUS), pp. 1–4 (2021). https://doi.org/10.1109/LAUS53676.2021.9639124
5. Draper, N.R., Smith, H.: Applied Regression Analysis. Wiley (1998)
6. Hedrick, W.R., Hykes, D.L., Starchman, D.E.: Ultrasound Physics and Instrumentation. Mosby Inc., Missouri (1995)
7. Jensen, J.A.: FIELD: a program for simulating ultrasound systems. In: 10th Nordic-Baltic Conference on Biomedical Imaging, vol. 4, Supplement 1, Part 1, pp. 351–353 (1996)
8. Levesque, P., Sawan, M.: Real-time hand-held ultrasound medical-imaging device based on a new digital quadrature demodulation processor. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **56**(8), 1654–1665 (2009). https://doi.org/10.1109/TUFFC.2009.1230
9. Omari, E.A., Varghese, T.: Signal to noise ratio comparisons for ultrasound attenuation slope estimation algorithms. Med. Phys. **41**(3), 032, 902 (2014)
10. Oppenheim, A.V.: Discrete-Time Signal Processing. Pearson Education India (1999)
11. Rindal, O.M.H., Åsen, J.P., Holm, S., Austeng, A.: Understanding contrast improvements from capon beamforming. In: 2014 IEEE International Ultrasonics Symposium, pp. 1694–1697 (2014). https://doi.org/10.1109/ULTSYM.2014.0420
12. Salkind, N.J.: Encyclopedia of Research Design, vol. 1. Sage (2010). https://books.google.com.br/books?id=HVmsxuaQl2oC
13. Schiefler, N., Maia, J., Schneider, F., Zimbico, A., Assef, A., Costa, E.: Generation and analysis of ultrasound images using plane wave and sparse arrays techniques. Sensors **18**(11), 3660 (2018). https://doi.org/10.3390/s18113660
14. Tasnim, T., Shuvo, M.M.H., Hasan, S.: Study of speckle noise reduction from ultrasound B-mode images using different filtering techniques. In: 2017 4th International Conference on Advances in Electrical Engineering (ICAEE), pp. 229–234 (2017). https://doi.org/10.1109/ICAEE.2017.8255358
15. Thomenius, K.E.: Evolution of ultrasound beamformers. In: IEEE Ultrasonics Symposium. Proceedings, vol. 2, pp. 1615–1622 (1996)
16. Xu, M., Yang, X., Ding, M., Yuchi, M.: Spatio-temporally smoothed coherence factor for ultrasound imaging [correspondence]. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **61**(1), 182–190 (2014). https://doi.org/10.1109/TUFFC.2014.6689786
17. Zhou, H., Zheng, Y.F.: An efficient quadrature demodulator for medical ultrasound imaging. Front. Inf. Technol. Eletron. **16**(4), 301–310 (2015)
18. Zimbico, A.J.: Bayesian-based beamformer with wiener post-filter for adaptive processing of ultrasond image using coherent plane wave compounding. Ph.D. thesis, Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba. PhD Thesis (CPGEI). (2019)

# Factors Affecting the Performance of FastICA Algorithm for Decomposition of High-Density Surface Electromyogram

Mateus Augusto Schneider Castilhos[1,2] and Leonardo Abdala Elias[1,2(✉)]

[1] Department of Electronics and Biomedical Engineering, School of Electrical and Computer Engineering, University of Campinas, Campinas, SP, Brazil
leoelias@unicamp.br

[2] Neural Engineering Research Laboratory, Center for Biomedical Engineering, University of Campinas, Campinas, SP, Brazil

**Abstract.** The motor unit (MU) physiology and the neural control of human movement can be studied via surface electromyogram (sEMG) decomposition techniques, which estimate the discharge times of MUs during a given contraction. The recent decomposition methods are based on the blind source separation (BSS) methods, and some approaches such as independent component analysis (ICA) can be applied to solve the problem. This study aims to investigate the influence of several parameters of the fast independent component analysis (FastICA) algorithm in the decomposition process of high-density sEMG. The extension factor, the number of iterations, and the method of initialization of the separation vector were explored for three evaluation metrics, namely, the execution time of decomposition, the absolute number of unique decomposed MUs, and the ratio of the number of unique MUs to the total number of extracted MUs. The number of iterations increased the execution time of decomposition and the absolute number of unique decomposed MUs. Besides, the method of initialization of the separation vector had little influence on the execution time of decomposition, an optimal value can be achieved for the absolute number of unique decomposed MUs. Moreover, the repeated convergence of the FastICA algorithm to the same source was mainly affected by the number of iterations. The study has reinforced the importance of evaluating the parameter combinations to achieve a better FastICA decomposition performance based on the chosen evaluation metric.

**Keywords:** Blind Source Separation (BSS) · Decomposition · Independent Component Analysis (ICA) · Surface Electromyogram (sEMG)

## 1 Introduction

The study of the motor unit (MU) physiology and neural control of movement can be accomplished by recording the electromyogram (EMG), either invasive or non-invasive, and using decomposition algorithms capable of identifying individual MU spike trains (MUSTs) [1, 2]. The EMG acquisition technology has evolved from recording systems with few bipolar channels to modern systems with hundreds of channels, thereby

enabling the acquisition of high-density EMG signals [3]. Likewise, new decomposition methods and algorithms have been employed to decompose the high-density EMG into its constituent MUSTs [4].

The basic concept of modern algorithms for decomposing high-density surface EMG (HD sEMG) signals refers to the blind source separation (BSS) problem [4], which consists of estimating the signal sources given a set of observations. Several techniques have been employed to solve the BSS problem, e.g., the so-called independent component analysis (ICA), which assumes that the sources are statistically independent and aims to maximize the nongaussianity of the estimated sources during the computation of the inverse problem [5]. FastICA is the most efficient and widely used ICA algorithm [5]. It is a fixed-point algorithm of fast convergence, which adopts the approximated negentropy as a measure of nongaussianity using new contrast functions [5].

In the context of HD sEMG decomposition, the FastICA can be applied to estimate MUSTs. For instance, Chen and Zhou [6] used FastICA associated with the "peel-off" approach to mitigate the influence of the already identified MU action potential on the surface EMG signals during the decomposition. The rationale of this approach is to increase the number of extracted MUs. In [7], another study employing the FastICA, the optimization criterion measures the sparsity of estimated sources instead of the nongaussianity as in [5]; however, the same contrast functions employed in the classical algorithm may be used due to the intrinsic features of the MUSTs. Nonetheless, in [7], the algorithm initialization is different than proposed by [5], and after the FastICA a second procedure based on [4] was suggested to improve the estimated MUSTs by minimizing the coefficient of variation of the interspike intervals ($CoV_{ISI}$).

When using algorithms to solve a problem, their performance needs to be evaluated. For decomposition algorithms, as proposed by [8], some performance metrics can be considered when varying a set of parameters, such as the computation time and the number of extracted MUs. Therefore, this study aims to investigate the influence of three main parameters of the FastICA algorithm (the extension factor, the number of iterations, and the method of initialization of the separation vector) on three decomposition output metrics (the execution time, the absolute number of unique MUs, and the ratio of the number of unique MUs to the total number of extracted MUs).

## 2 Methods

### 2.1 High-Density EMG Data

The sEMG signals used in the present study were recorded from a single participant without known neurological disorders. The study was conducted in accordance with the Declaration of Helsinki. All procedures were approved by the Ethics Committee of the Imperial College (no. N 18IC4685).

The HD sEMG signals were acquired using a surface grid of 64 ($5 \times 13$) electrodes from the first dorsal interosseous muscle. An isometric contraction at 5% of maximum voluntary contraction (MVC) was performed during a force control task [9]. The sEMG signals were recorded during 45 s, band-pass filtered (10–500 Hz), and A/D converted at 2,048 Hz sampling rate and 16-bit resolution. A signal from a single trial was used as

representative to test the performance of the algorithm by changing its parameters (see below). For detailed information about the experimental setup and protocol see [9].

## 2.2 Decomposition Algorithm

The implemented HD sEMG decomposition algorithm is based on [7] and can be summarized in the pseudo-code below. For this study, only the FastICA (steps one to seven) approach was analyzed. The HD sEMG signals are represented as the observation matrix $X = [x_1, \ldots, x_m]^T$, where $x_i$ with $i = 1, \ldots, m$ is the observation of the $i$-th sEMG signal.

```
Extend the observation matrix X by a factor R to obtain
the extended observation matrix X̃;
Subtract the mean from each new extended observation in
X̃;
Whiten X̃ to obtain Z;
Initialize the separation matrix B to empty matrix;
for j=1,…,M do:
  Initialize vectors wj;0 and wj;−1 randomly or based on the
  activity index γ(k);
  while |wᵀj;n wj;n−1 − 1| > Tol or n < 30 do:
    wj;n ← E{Zg[wᵀj;n−1 Z]} − Awj;n−1 with A = E{g [wᵀj;n−1 Z]}, where
    E{·} is the mathematical expectation;
    Orthogonalization: wj;n ← wj;n − BBᵀwj;n;
    Normalization: wj;n ← wj;n/‖wj;n‖;
    n←n+1;
  end
  Add wj;n as a column of B;
  Estimate the j-th source ŝj = wᵀj;nZ;
  Use peak detection and K-means classification method on
  ŝj to estimate the MUSTj and compute the SIL;
  if SIL > 0.9 then:
    Store the current iteration number j and the MUSTj;
  end
  if wj;n initialization is based on γ(k) then:
    Remove the discharge times in MUSTj from γ(k);
  end
end
Update B = [w1;n, …, wM;n] removing columns that are not equal
to the js stored in step 6;
Remove duplicated stored MUSTs;
```

The pre-processing procedure is composed of an extension of the observations (EMG signals) with the removal of the mean values and whitening. The extension is responsible

to create $R$ delayed versions of each sEMG signal, with delays ranging from 1 to $R$, setting up $\widetilde{X}$ with $m(R+1)$ observation signals. This is commonly done in HD sEMG decomposition as an attempt to increase the ratio of the number of sEMG signals to the number of estimated MUSTs (sources) [7]. The removal of mean values is used to center the signals at zero as a mean to simplify the computation of ICA. The whitening process is applied to spatially decorrelate the extended observations, maintaining the original matrix dimensions [4, 5, 7].

The separation vectors $w_{j;n}$ are responsible for estimating the MU discharge times from the whitened extended HD sEMG signals, and a suitable initialization of these vectors provides a fast convergence to the FastICA algorithm. The study by Negro et al. [7] suggests an initialization approach based on the activity index $\gamma(k)$ introduced in [4], as a global indicator of MU discharge activity.

The FastICA is an optimization algorithm that uses the first, $g(x)$, and second, $g'(x)$, derivatives of a contrast function $G(x)$ that measures the sparsity to find the optimal separation vector. The convergence criterion is satisfied if the inner product between two successive estimated vectors ($w_{j;n-1}$ and $w_{j;n}$) can differ from one less than a factor equals to *Tol*. Also, the maximum number of iterations for FastICA was defined to avoid an infinite loop. Although the Gram-Schmidt orthogonalization of $w_{j;n}$ tries to reduce the repeated convergence [5, 7], it is not enough because of the influence of extension procedure. A method for removing duplicated MUs is needed at the end of the algorithm [8].

The peak finding and K-means classification method are applied to the estimated source to cluster the peaks into two classes. The class with the highest centroid represents the MU discharge times, and the class with the lowest centroid represents the noise. The silhouette measure (SIL) is computed using these classified peaks and centroids, and represents how well the estimated MU discharges can be distinguished from the baseline noise [7]. The SIL was used to classify if an estimated source should be considered extracted or not.

Another way to mitigate the repeated convergence of FastICA can be based on a specific part of the method proposed in [4], where the $\gamma(k)$ becomes null at the time instants corresponding to the discharge times of the $j$-th MUST at each iteration. This method updates the activity index by removing part of the influence of the estimated MU activity [7].

## 2.3 Assessing the Performance of the Algorithm

Two parameters and one procedure have been examined, namely the number of decomposition iterations ($M$), the extension factor ($R$), and the method of initialization of the separation vector ($w_{init}$). Table 1 contains the range of values for these parameters and the initialization approach types. The values chosen for $R$ were based on the number of observed channels, as proposed in [7], and the values chosen for $M$ were based in [8]. Moreover, the methods for initialization of separation vectors were defined as: 1) both, $w_{j;n}$ and $w_{j;n-1}$, have been initialized with random numbers [5, 8]; 2) $w_{j;n}$ has been initialized as the column of $Z$ that corresponds to the time instant of the maximum value of $\gamma(k)$ [7] and $w_{j;n-1}$ as the column of $Z$ which corresponds to $Rdiv2$ samples

before $w_{j;n}$, where *div* is the integer division; and 3) similar to (2) but with the median of $\gamma(k)$ used instead of the maximum [4]. All possible combinations between the three algorithm parameters described above were accomplished providing 27 executions.

**Table 1.** Algorithm parameters explored in the present study. Number of decomposition iterations ($M$), extension factor ($R$), and method of initialization of the separation vector ($w_{init}$).

| $M$ | $R$ | Initialization Method |
|---|---|---|
| 100; 200; 300 | 8; 12; 16 | *random*; $\underset{arg}{max}(\gamma(k))$; $\underset{arg}{median}(\gamma(k))$ |

As suggested by Negro et al. [7], the *Tol* parameter was set to $10^{-4}$, the $G(x) = x^3/3$, and the silhouette threshold was 0.90. The rationale for restricting the maximum number of iterations of FastICA to 30 was based on previous decompositions (unpublished) using the same data, but this value can range from 20 to 80 [8].

The identification and removal of duplicate sources (MUSTs) were semi-automatic. Initially, the extracted MUSTs were sorted according to the number of spikes in ascending order. After, the sorted MUSTs were clustered, with each group containing 25% of the total number of extracted MUSTs. The evaluation of duplicates was carried out using pairs of MUSTs in the same cluster and a 50% overlap between clusters was considered. The spikes in both compared clusters was considered. The spikes in both compared MUSTs were considered the same spike if they occur in an interval of $(Rdiv2) \times 1$ ms. The MUSTs were considered the same if the ratio of the number of common spikes ($c_{spikes}$) to the total number of spikes ($n_{spikes}$) satisfies $c_{spikes}/n_{spikes} \geq 39;45;53\%$, for $R = 8;12;16$ respectively. These percentages were empirically determined after several decompositions using the same data. The MU with the greatest SIL was chosen as unique (i.e., not duplicated), but if two MUs with equal SIL were classified as equal, the one with the smallest $CoV_{ISI}$ was chosen at the final step of the algorithm. An expert has verified the duplicates and was able to confirm or refuse the algorithm result.

The algorithm evaluation was based on some metrics used in [8], such as the decomposition total time, which corresponds to the execution time duration of the *for* loop in the pseudo-code, and the total number of extracted MUs $N_{tot}$ (with duplicates). All possible combinations of parameters were analyzed considering, as cost functions, the normalized execution time ($T_{norm}$), the absolute number of unique MUs ($N_{uni}$), and the ratio $N_{uni} : N_{tot}$. The $T_{norm}$ computation was based on [8] with $T_{norm} = \frac{t_{exec} - min(t_{exec})}{max(t_{exec}) - min(t_{exec})}$, where $t_{exec}$ is the execution time of the parameter combinations.

The best parameter configurations for each cost function was suggested. For $N_{uni} : N_{tot}$ cost function, some MU metrics were analyzed, such as SIL, $CoV_{ISI}$, mean discharge rate ($DR_{mean}$), skewness (Fisher-Pearson definition) of ISI distribution ($Sk_{ISI}$), and kurtosis (Fisher's definition) of ISI distribution ($Kurt_{ISI}$). It was excluded from the computation of these metrics the MUs which presented $ISIs > 250$ ms.

The decomposition algorithm was coded in Python programming language (version 3.10.3). Decomposition and performance analyses were performed on a personal computer with the following specifications: Intel® Core™ i7-8550U processor (64 bits and from 1.80 to 1.99 GHz), 16 GB RAM, and Windows 10™ Home (version 21H2).

## 3   Results

Figure 1, panels A to C, shows a direct influence of the number of iterations ($M$) and the normalized execution time ($T_{norm}$), so that increasing the iteration number of the decomposition algorithm results in a longer execution time irrespective of the extension factor ($R$) and the initialization method ($w_{init}$). Analysing the influence of $w_{init}$ on $T_{norm}$, we did not notice any impact of this parameter since the execution time did not change for the different choices of vector initialization (Fig. 1, panels D to F). Conversely, by increasing the extension factor $R$ the $T_{norm}$ increased (Fig. 1, panels G to I). Similar results were obtained in [8]. The worst scenario (the highest execution time) was for the *maximum* initialization approach, $R = 16$ and $M = 300$, and the best (i.e., the shortest execution time) was for the *random* initialization approach, $R = 8$, and $M = 100$.
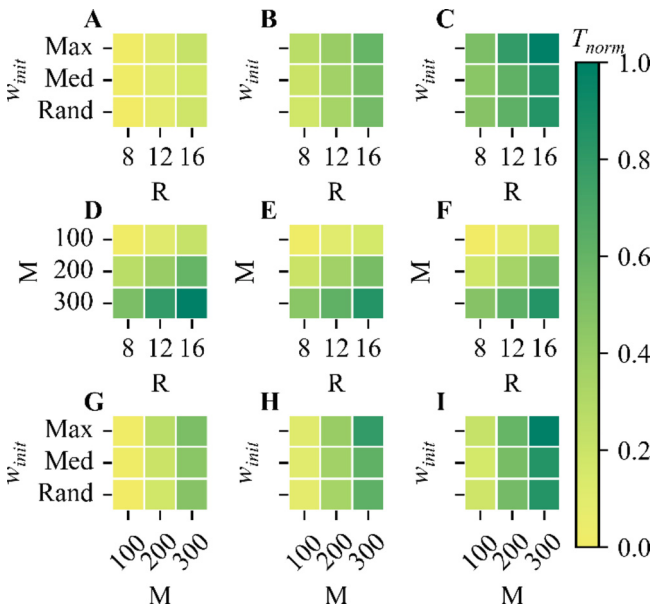


**Fig. 1.** Influence of parameters of decomposition algorithm on the normalized execution time ($T_{norm}$). Maps for all possible combinations of pairs of parameters are shown in colors. Max, Med, and Rand represent the methods for initialization of the separation vector ($w_{init}$) based on the *maximum* of the activity index, the *median* of the activity index, and *random* initialization, respectively. Combinations for $M = 100$ (**A**), 200 (**B**) and 300 (**C**); $w_{init} = $ Max (**D**), Med (**E**), and Rand (**F**); and $R = 8$ (**G**), 12 (**H**) and 16 (**I**).

The analysis of the absolute number of unique MUs shows that increasing the iteration number ($M$) there was an increased number of unique decomposed MUs ($N_{uni}$) (Fig. 2, panels A to C), as also reported in [8]. In general, a maximum in $N_{uni}$ is achieved if the initialization method is based on the *maximum* of the activity index, as compared to the other methods adopted in the present study (*random* and *median*), irrespective of the other two parameters evaluated here (Fig. 2, panels D to F). The extension factor ($R$) influences the $N_{uni}$ depending on the choice of $M$ and $w_{init}$ (Fig. 2, panels D to F). For the *maximum* initialization method $R = 16$ provides more unique MUs than $R = 8$ and 12, but for the *random* initialization $R = 8$ provides more $N_{uni}$. Moreover, there were three occurrences for both, the greatest and the smallest $N_{uni}$ (26 and 15, respectively). The choices of the best (the greatest $N_{uni}$) and worst case (the smallest $N_{uni}$) was *maximum* initialization method, $R = 16$, and $M = 200$; and *random* initialization method, $M = 200$, and $R = 16$, respectively. The choices were done assuming the execution time as the second selection criterion.

The $N_{uni} : N_{tot}$ ratio decreases (Fig. 3, panels A to C) by increasing the number of iterations ($M$). The method of initialization of the separation vector ($w_{init}$) and the extension factor ($R$) had little influence on the $N_{uni} : N_{tot}$ ratio compared to $M$ (Fig. 3, panels D to I). However, $w_{init}$ presents some dependency on the number of iterations (Fig. 3, panels D to F). The $N_{uni} : N_{tot}$ ratio also represents an implicit proportion
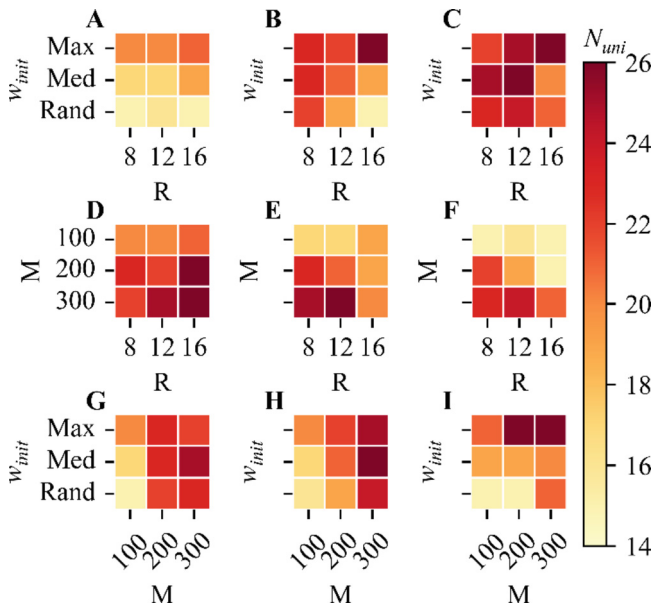


**Fig. 2.** Influence of parameters of decomposition algorithm on the number of unique decomposed motor units ($N_{uni}$). Maps for all possible combinations of pairs of parameters are shown in colors. Max, Med, and Rand represent the methods for initialization of the separation vector ($w_{init}$) based on the *maximum* of the activity index, the *median* of the activity index, and *random* initialization, respectively. Combinations for $M = 100$ (**A**), 200 (**B**) and 300 (**C**); $w_{init} =$ Max (**D**), Med (**E**), and Rand (**F**); and $R = 8$ (**G**), 12 (**H**) and 16 (**I**).

of between the amount of unique MUs to the number of repeated (duplicated) MUs ($N_{dup} = N_{tot} - N_{uni}$), and hence if $N_{dup}$ increases faster than $N_{uni}$, the $N_{uni} : N_{tot}$ will decrease for a specific parameter combination. The best parameter combination was for the *maximum* initialization method, $R = 8$, and $M = 100$, which has extracted 21.28% unique MUs of $N_{tot}$. The worst combination was for the *median* initialization method, $R = 16$, and $M = 300$, providing 7.04% unique MUs.

For the data used in the present exploratory study, the suggested combination of parameters ($[w_{init}; R; M]$) of the decomposition algorithm based on the FastICA is: 1) to minimize the $T_{norm}$ [*random*; ; 8; 100], 2) to maximize the $N_{uni}$, , [*maximum*; ; 16; 200], and 3) to maximize the $N_{uni} : N_{tot}$ ratio, [*maximum*; 8; 100]. For this last combination of parameters, the metrics SIL, $CoV_{ISI}$, $DR_{mean}$, $Sk_{ISI}$ and $Kurt_{ISI}$ of the unique extracted MUs ($ISI < 250$ ms) are summarized in Table 2.
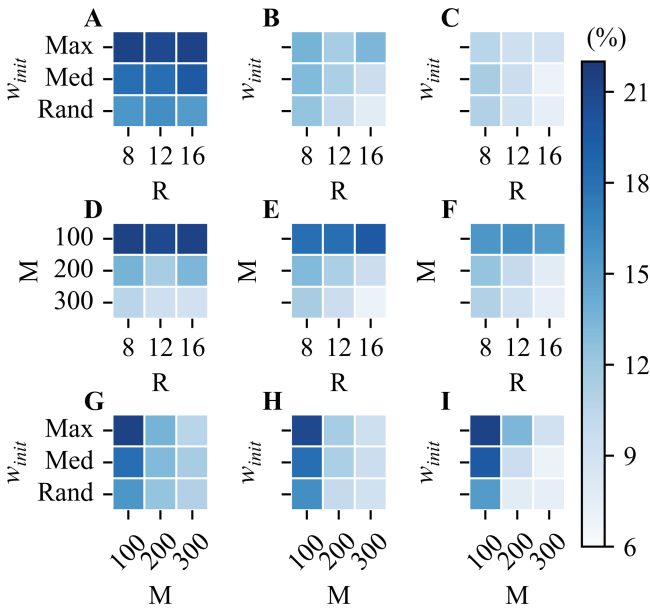


**Fig. 3.** Influence of parameters of decomposition algorithm on the ratio of the number of unique MUs to the total number of extracted MUs ($N_{uni} : N_{tot}$). Maps for all possible combinations of pairs of parameters are shown in colors. Max, Med, and Rand represent the methods for initialization of the separation vector ($w_{init}$) based on the *maximum* of the activity index, the *median* of the activity index, and *random* initialization, respectively. Combinations for $M = 100$ (**A**), 200 (**B**) and 300 (**C**); $w_{init} = $ Max (**D**), Med (**E**), and Rand (**F**); and $R = 8$ (**G**), 12 (**H**) and 16 (**I**).

## 4    Discussion

The results showed that by varying the iteration number ($M$) of the decomposition algorithm, the normalized execution time (Fig. 1) and the number of unique extracted MUs (Fig. 2) varied in the same direction as observed by [8]. Moreover, by varying $M$

**Table 2.** Metrics from the extracted motor unit spike trains (MUSTs) when the parameters of the decomposition algorithm was set to [$R = 8$; $M = 100$; $w_{init} = maximum$]. The metrics were: silhouette (SIL), coefficient of variation of interspike intervals ($CoV_{ISI}$), mean discharge rate ($DR_{mean}$), skewness ($Sk_{ISI}$), and kurtosis ($Kurt_{ISI}$) of the distributions of interspike intervals. For this set of parameters $N_{uni} = 20$ MUs, but seven was excluded from the analysis ($ISI > 250$ ms). Measures from single MUs are shown in the rows. Averages and the 95% confidence intervals are shown at the bottom row.

| MU | SIL (%) | $CoV_{ISI}$ (%) | $DR_{mean}$ (Hz) | $Sk_{ISI}$ | $Kurt_{ISI}$ |
|---|---|---|---|---|---|
| 1 | 95.96 | 23.52 | 8.28 | 2.09 | 11.77 |
| 2 | 97.41 | 25.65 | 9.24 | 1.50 | 5.07 |
| 3 | 94.16 | 23.56 | 9.38 | 4.04 | 27.85 |
| 4 | 94.12 | 27.58 | 9.39 | 1.80 | 5.80 |
| 5 | 96.33 | 27.51 | 10.19 | 1.54 | 7.78 |
| 6 | 95.55 | 20.48 | 10.65 | 1.55 | 9.46 |
| 7 | 95.28 | 25.14 | 12.03 | 1.49 | 4.80 |
| 8 | 91.82 | 22.51 | 12.24 | 2.59 | 18.37 |
| 9 | 95.56 | 19.07 | 12.40 | 1.25 | 5.18 |
| 10 | 95.24 | 20.65 | 13.18 | 1.66 | 9.40 |
| 11 | 90.60 | 28.84 | 13.33 | 2.51 | 10.56 |
| 12 | 95.81 | 18.82 | 13.82 | 1.09 | 4.61 |
| 13 | 94.58 | 20.25 | 14.50 | 1.13 | 5.07 |
|  | $94.80 \pm 1.00$ | $23.35 \pm 1.85$ | $11.43 \pm 1.10$ | $1.86 \pm 0.44$ | $9.67 \pm 3.65$ |

the $N_{uni} : N_{tot}$ ratio (Fig. 3) changes in the opposite direction. This implies that in our current data by increasing the value of $M$, the $N_{dup}$ increased more than the $N_{uni}$.

The extension factor $R$ and the $T_{norm}$ presented a direct relationship (Fig. 1) as observed by [8], but, it had little influence on $N_{uni} : N_{tot}$ (Fig. 3), as well as on $N_{uni}$ for $M = 100$ (Fig. 2). However, different from [8], we observed that for the $N_{uni}$, the extension factor is dependent on the other algorithm parameters. Figure 2, panels G to I, shows that for the *maximum* initialization approach and $M = 200$ and 300, in general, by increasing the extension factor the same occurred to $N_{uni}$. Otherwise, considering the *median* and *random* initialization approaches, $R = 8$ and 12 provide more unique MUs than $R = 16$.

The method for the initialization of the separation vector $w_{init}$ caused distinct effects on the output measures of the decomposition algorithm. By altering this parameter, the normalized execution time (Fig. 1) was little affected. Conversely, the $N_{uni}$ (Fig. 2) changed in the opposite direction tended to reach a maximum considering the *maximum* initialization approach. In general the $w_{init}$ had little influence on $N_{uni} : N_{tot}$ ratio (Fig. 3), but considering $M = 100$ the $N_{uni} : N_{tot}$ ratio (Fig. 3, panels D to F) is more affected than considering the other values of $M$, resulting in a higher $N_{uni} : N_{tot}$ ratio value

to the *maximum* initialization approach. This implies that for $M = 100$, the $N_{dup}$ is proportionally lower by choosing the *maximum* initialization approach than the $N_{dup}$ by choosing the other methods. Furthermore, considering the *random* initialization method for the separation vectors proposed by [5] as a reference and analyzing the ratio of the number of unique extracted MUs to the total number of MUs (Fig. 3), it was noticeable that the *maximum* [7] and *median* [4] initialization methods did mitigate the effect of repeated convergence of FastICA algorithm for $M = 100$. However, the same result was not achieved considering the other values of $M$.

The data for the best parameter combination (*maximum* initialization method, $R = 8$, and $M = 100$) that maximizes the $N_{uni} : N_{tot}$ cost function (Table 2) shows, on average, that the spikes of the estimated MUSTs are well distinguishable from the baseline noise with a silhouette score equal to $94.80 \pm 1.00$, which is compatible with the results from [7]. The average skewness of the ISI distribution ($1.86 \pm 0.44$) indicates that the probability distribution of the interspike intervals is asymmetrical, with a right-hand skew [10]. The kurtosis indicates that the probability distributions of all ISIs are super-gaussian (values higher than zero), likewise other studies [7]. The mean discharge rate was $\sim$12 Hz, as expected for the first dorsal interosseous muscle in a low-intensity contraction [10, 11]. The ISI CoV were $\sim$25%, suggesting that the MUs are discharging near their recruitment thresholds [11].

One should take into consideration that the present study was conducted using a single trial from a single participant in a very specific motor task (see Methods and [9]). For the purposes of the present study this was necessary, because we were interested in testing the changes in the decomposition parameters, and hence including a large dataset would increase the complexity of our analysis. However, the present study should be expanded to include a larger dataset to confirm the current interpretations regarding the influences of $M$, $R$, and the initialization method on the HD sEMG decomposition outcomes. Nonetheless, we do not expect to find a universal parameter set for all sEMG signals. We consider the experimenters and data analysts should explore the decomposition parameters to find those that better fit their needs and produce more consistent results based on physiological constraints.

## 5  Conclusion

We showed that the parameter choice of the FastICA algorithm would influence the outcomes of sEMG decomposition into constituent MUSTs. We are aware that a large dataset is needed to confirm our findings; however, our study suggests that for the first dorsal interosseous muscle during an isometric contraction force at 5% MVC, the number of decomposed MUs may be maximized by setting the initialization of the separation vector based on the maximum of the activity index. Moreover, a great number of MUs can be achieved with minimal increase in the time of execution by preferentially increasing the extension factor instead of the number of iterations. In addition, by maintaining the number of iterations equal to 100, not only the execution time is optimized, but also the ratio of the number of unique MUs to the total number of extracted MUs.

**Conflict of Interest.**   The authors declare that they have no conflict of interest.

# References

1. De Luca, C.J., Adam, A., Wotiz, R., et al.: Decomposition of surface EMG signals. J. Neurophysiol. **96**(3), 1646–1657 (2006). https://doi.org/10.1152/jn.00009.2006
2. McGill, K.C., Lateva, Z.C., Marateb, H.R.: EMGLAB: an interactive EMG decomposition program. J. Neurosci. Methods **149**(2), 121–133 (2005). https://doi.org/10.1016/j.jneumeth.2005.05.015
3. Farina, D., Negro, F., Muceli, S., et al.: Principles of motor unit physiology evolve with advances in technology. Physiology **31**(2), 83–94 (2016). https://doi.org/10.1152/physiol.00040.2015
4. Holobar, A., Zazula, D.: Multichannel blind source separation using convolution Kernel compensation. IEEE Trans. Signal Process. **55**(9), 4487–4496 (2007). https://doi.org/10.1109/TSP.2007.896108
5. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Networks **10**(3), 626–634 (1999). https://doi.org/10.1109/72.761722
6. Chen, M., Zhou, P.: A novel framework based on FastICA for high density surface EMG decomposition. IEEE Trans. Neural Syst. Rehabil. Eng. **24**(1), 117–127 (2016). https://doi.org/10.1109/TNSRE.2015.2412038
7. Negro, F., Muceli, S., Castronovo, A.M., et al.: Multi-channel intramuscular and surface EMG decomposition by convolutive blind source separation. J. Neural Eng. **13**(2), 026027 (2016). https://doi.org/10.1088/1741-2560/13/2/026027
8. Meng, L., Chen, Q., Jiang, X., et al.: Evaluation of decomposition parameters for high-density surface electromyogram using fast independent component analysis algorithm. Biomed. Signal Process. Control **75**, 103615–103623 (2022). https://doi.org/10.1016/j.bspc.2022.103615
9. Germer, C.M., Del Vecchio, A., Negro, F., et al.: Neurophysiological correlates of force control improvement induced by sinusoidal vibrotactile stimulation. J. Neural Eng. **17**(1), 016043 (2020). https://doi.org/10.1088/1741-2552/ab5e08
10. Barry, B.K., Pascoe, M.A., Jesunathadas, M., et al.: Rate coding is compressed but variability is unaltered for motor units in a hand muscle of old adults. J. Neurophysiol. **97**(5), 3206–3218 (2007). https://doi.org/10.1152/jn.01280.2006
11. Moritz, C.T., Barry, B.K., Pascoe, M.A., et al.: Discharge rate variability influences the variation in force fluctuations across the working range of a hand muscle. J. Neurophysiol. **93**(5), 2449–2459 (2005). https://doi.org/10.1152/jn.01122.2004

# Heartbeat Classification Using MLP and Random Forest Techniques

Carlos A. L. Ortiz$^{(\boxtimes)}$ and Jurandir Nadal

Biomedical Engineering Program, University Federal of Rio de Janeiro,
Rio de Janeiro, Brazil
`carlos.ortiz@peb.ufrj.br`

**Abstract.** This work aims at presenting a method for heartbeat classification based on multi-layer perceptron (MLP) and random forest (RF) techniques applied to the first difference of ECG signals. From the MIT-BIH Arrhythmia Database, each annotated P-QRS segment was extracted, low-pass filtered, and the first-order difference was used as input for the neural networks. The MLP and Random Forest were used to obtain a model for classifying the heartbeats. The results were compared with other algorithms existing in the literature, and the model developed produced good results and noticed improvements when using the first difference.

**Keywords:** Heartbeat Classification · Multi Layer Perceptron · Random Forest · Signal Processing

## 1 Introduction

The electrocardiogram (ECG) is being used for the non-invasive assessment of the bioelectric activity of the heart for more than 100 years. In intensive care units, it is widely used to continuously monitor the patients, either by the instantaneous heart rate or by the automatic detection of life-threatening arrhythmias [1–5]. The ECG is composed of five wave patterns that repeat beat-to-beat and represents the cardiac cycle. These waves are named P, Q, R, S, and T (Fig. 1). Figure 2 presents the distribution of electrodes and clamps for detecting the ECG signal. Each of these waves has a standard range of amplitude and duration.

Dozens of automatic arrhythmia monitoring algorithms based on ECG are being proposed in the last fifty years. Although automatic monitors showed to be helpful to prevent continuous human monitoring and have helped to save uncountable lives, they are also responsible for preventable noise in ICU, due to false alarms. Thus, the investigation of more accurate classifiers is still in progress in a wide number of research groups around the world. To standardize the classification methods evaluation and inter-algorithm performance comparison, an AAMI/ANSI [8] standard was established, which recommends grouping the MIT-BIH heartbeat classes into five groups: (1) Non-ectopic or Normal beats (N); (2) Supraventricular ectopic beats (S); (3) Ventricular Ectopic Beats (V);
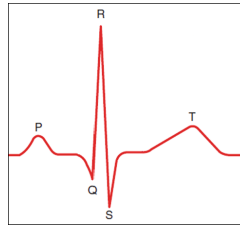
**Fig. 1.** P, Q, R, S and, T waves of an ECG signal.  Adapted from [6].
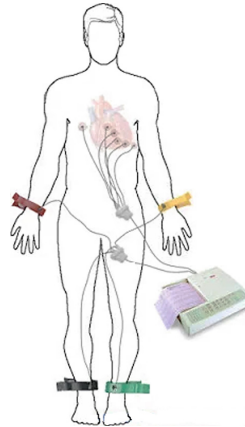


**Fig. 2.** Distribution of electrodes and clamps for detecting the ECG signal.  Adapted from [7].

(4) Fusion Beats (F); and (5) Unknown Beats (Q). Each group contains one or more type of heartbeat, as shown in (Table 1).

The first derivative of ECG data is widely used in QRS complex detection algorithms for improving the signal-to-noise ratio, however, the original data samples are used for the classification. This study aims to test the applicability of the first derivative for classification purposes. For this aim, two neural network topologies were employed: Multi-Layer Perceptron (MLP) due to its wide application in pattern selection and classification and Random Forest (RF) to verify the benefits of flexibility, reduction of overfitting, and its applicability in the classification of heartbeats.

## 2   Materials and Methods

### 2.1   MIT-BIH Arrhythmia Database

This database [9] is formed by 48 30-min records of ambulatorial ECG from 47 different patients, taken in two leads and digitally stored in 360 samples per second. Only the first lead (thoracic modified DII lead) was analyzed. The data files

**Table 1.** Principal classes of the MIT-BIH database and AAMI standard classes

| AAMI class | Symbol | MIT-BIH class |
|---|---|---|
| N | N | Normal beat |
| | L | Left bundle branch block beat |
| | R | Right bundle branch block beat |
| | e | Atrial escape beat |
| | j | Nodal (junctional) escape beat |
| S | A | Atrial premature beat |
| | a | Aberrated atrial premature beat |
| | J | Nodal (junctional) premature beat |
| | S | Supraventricular premature beat |
| V | V | Premature ventricular contraction |
| | E | Ventricular escape beat |
| F | F | Fusion of ventricular and normal beat |
| Q | p or / | Paced beat |
| | f | Fusion of paced and normal beat |
| | U | Unclassifiable beat |

were organized into four groups (Table 2), for the k-fold cross-validation approach with k = 4. These files were also grouped as determined by the AAMI Standard [8], for comparison purposes. The DII derivation was used, representing the thoracic bipolar derivation, tantalizing 24 h of ECG registers. Although containing 24 h of ECG data, with more than 107000 heartbeats, the database has a small number of cases in non-normal classes, requiring the parsimonious use of this data for classification assessment. Therefore, the use of k-fold cross-validation was adopted. This approach uses three of four datasets for classification development and validation, while the fourth dataset was reserved for independent testing. Thus, after four different classification training, each one preserving a different dataset Table 3, the classification performance may be assessed with the complete database. Each one of the four training datasets was divided into two sub-sets, where 70% of the data was reserved for training (training dataset) and 30% for validation (validation dataset), as usually adopted in the literature.

**Table 2.** Groups For Train, Test, and Validation

| datasets | MIT-BIH Patients |
|----------|------------------|
| dataset 1 | 100, 101, 102, 109, 116, 122, 124, 200, 201, 220, 221, 228 |
| dataset 2 | 105, 106, 117, 118, 119, 121, 203, 213, 207, 217, 222, 234 |
| dataset 3 | 104, 111, 113, 115, 123, 202, 208, 209, 212, 215, 219, 223 |
| dataset 4 | 103, 107, 108, 112, 114, 205, 210, 214, 230, 231, 232, 233 |

**Table 3.** Training and testing datasets for k-fold cross-validation

| Training/Validation | Testing |
|---------------------|---------|
| Group 1: dataset 2, dataset 3, dataset 4 | dataset 1 |
| Group 2: dataset 1, dataset 3, dataset 4 | dataset 2 |
| Group 3: dataset 1, dataset 2, dataset 4 | dataset 3 |
| Group 4: dataset 1, dataset 2, dataset 3 | dataset 4 |

## 2.2   ECG Signal Preprocessing

The ECG signals in the MIT-BIH Arrhythmia Database contain the noise of different frequencies, so the first step in signal preprocessing is filtering the ECG signals. For this, we designed a two-sample moving average filter. In this way, it was possible to attenuate the high-frequency noises of the signals. We used the notes made by two Massachusetts Institute of Technology cardiologists available in the MIT-BIH Arrhythmia Database, which presented the occurrence of each beat (represented by the R wave of the QRS complex) and also selected 108 samples around each heartbeat.

The MIH-BIH Arrhythmia Database also presents heartbeats classes with imbalanced data, which is a problem for the beats classification. The typical classification algorithms assume that the data are uniformly distributed in the class and that classes are balanced. When these conditions change, the algorithm's performance measurement can be skewed. For this reason, we resampled the data using a random oversampling approach [10]. In random oversampling, the minority class samples are increased by random sampling replication of minority class representatives. Furthermore, the principal component analysis was used to reduce the data dimensionality of the original database signals after filtering and selecting samples from the ECG signals.

## 2.3   ECG Signal Preprocessing

A segment of 110 samples containing the P-wave and the QRS-Complex were extracted from each annotated heartbeat of the Database, and the first difference was calculated (Fig. 3). As shown, the first difference attains a prominent wave in the QRS location, and solve the baseline problem in each ECG signal.
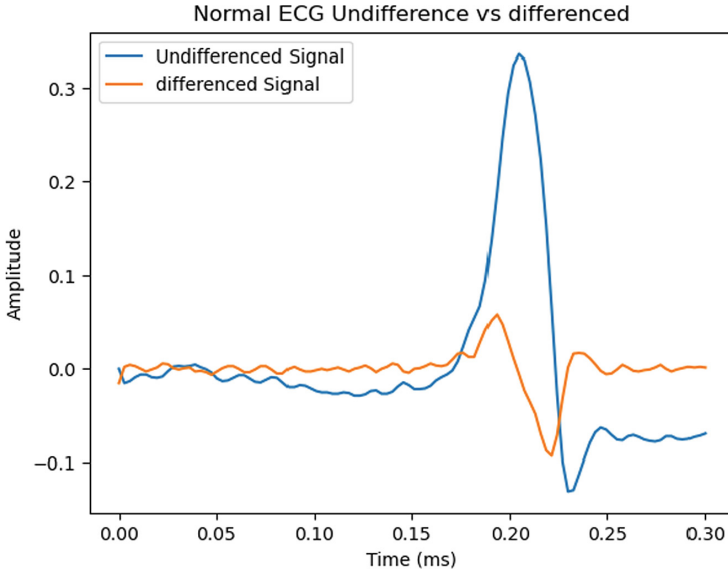
**Fig. 3.** Normal ECG and Differentiated Normal ECG. (Author)

After pre-processing, the signal segments were grouped in four different tables, according to the groups shown in Table 3, one heartbeat per column, followed by the corresponding class according to Table 1. To ensure the quality of signals, and reduce noise and artifact effects on signals, it was also applied a low pass moving average filter to each data segment. After these, the information of about 300 ms was retained around the R peak, which was represented by 72 samples before the R peak and 37 samples after the R peak. The position of the R peak was taken from the annotation file of the MIT-BIH database. Due to the unbalance in the database in the number of beats in each class, which could affect the classifier training, two different forms of results were verified: one for the unbalanced database and one for the balanced one. To perform the balanced database, the data for each beat were up-sampled to 10000 samples in each type.

## 2.4   Classifiers

Two classifier topologies were tested: MLP and RF. The results of those classifiers were validated using Accuracy (individual, and macro), Specificity, and sensibility. In the end, the high accuracy and the best results in the confusion matrix were used to determine the best classifier. The implemented classifiers, MLP and RF, followed the above structures: MLP: Randomstate = 1, max iter = 200, Hiden layer size: 100, Activation: 'relu'(rectified linear unit function), Solver = 'Adam', Learning rate = constant. RF: max depth = 8, estimators = 100, Criterion = Gini, Min samples split = 2, Min samples leaf = 1, Max

**Table 4.** Real Signal VS. Differentiated Signal
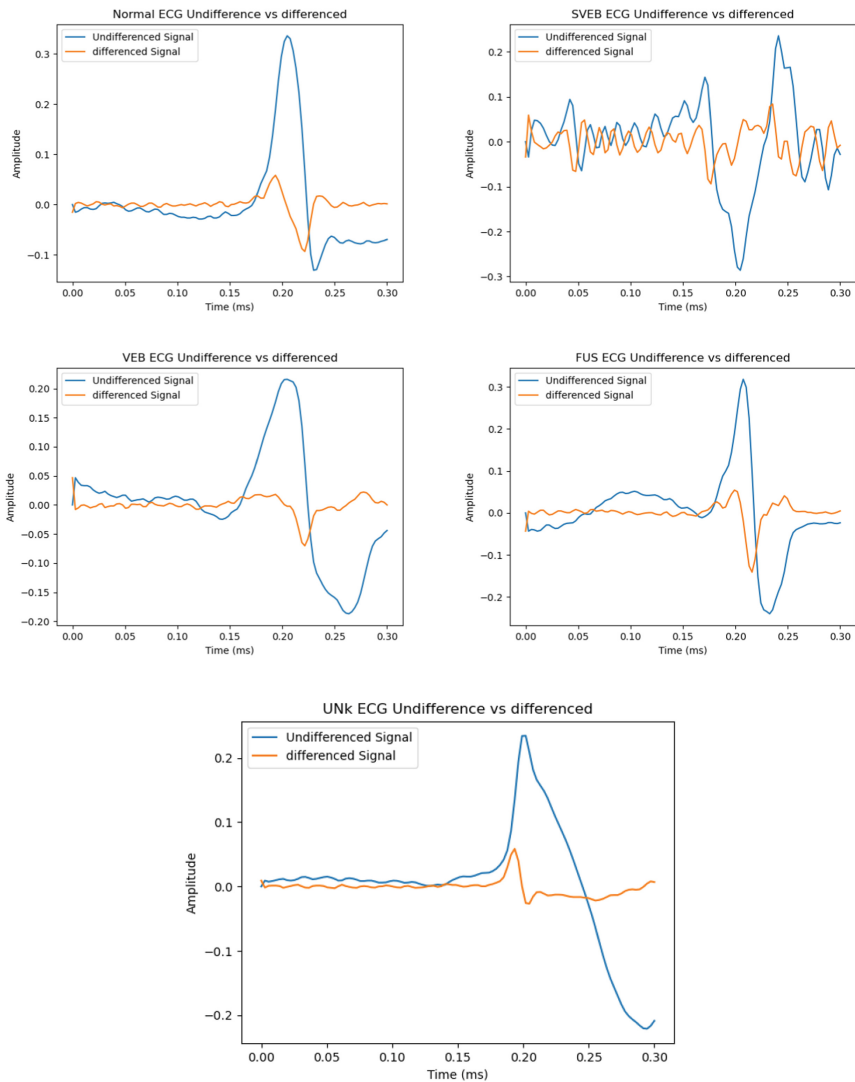


features = auto. These hyperparameters were subjectively selected in the search for better results in the present paper.

## 3    Results

This section presents the results obtained considering the training and testing datasets studied in this work. Comparisons between the method developed and

other methodologies found in the literature that also use the NN's are performed. Some measures, such as accuracy, precision, sensitivity, and specificity, are calculated to evaluate the models and are described in Equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative. Some results for these are shown in Table 4. As shown, the real signal shapes are widely variable, which could affect the classifier results. They are also corrupted by the mains noise (50 or 60 Hz), electromyographic noise from breath muscles, and movement artifacts. As seen in this figure, there is also a drift in the signal that could be solved easily with the first difference of the signal. The results for each classifier were conducted in two forms, first with an unbalanced database and second with a balanced database, where all beat classes have 10,000 beats, obtained by randomly replication of beats in the classes with few examples. The results for MLP (Table 5), illustrate the effect of data balancing. Although presenting the same accuracy, the classification with unbalanced samples was not effective to detect the beat class under-represented (F). The confusion matrix shows the classification results of the proposed MLP for the five heartbeats classes, exhibiting the quantity of correct and incorrect predictions (Table 6). The results for RF are presented in (Table 7) and the confusion matrix in Table 8, shows the classification results of the proposed RF for the five heartbeats classes, exhibiting the quantity of correct and incorrect predictions. In this case, the classifier did not detect the class F with unbalanced data. After balancing data, the overall accuracy reduced, but all classes were recognized.

## 4  Discussion

The methodology studied in this article presents satisfactory results, when compared with other in the literature (Table 9), whose algorithms also used methods of computational intelligence. Both in training/validation and in the test of the proposed MLP and RF network, the classifiers reach high accuracy values considering all classes studied. For almost all datasets used in the cross-validation, precision was higher than 97% in all classes, however, for RF with unbalanced data a Fusion class was not achieved due to the limited of amount o beats that present this class, without affecting the classification accuracy. When using RF

**Table 5.** Precision, sensitivity, and accuracy for each testing dataset considering AAMI standard classes. NN: MLP

| Unbalanced dataset, Train: 1,2,3; Test: 4 | | | |
|---|---|---|---|
| Class | Precision (%) | Sensitivity (%) | Specificity (%) |
| N | 100.00 | 100.00 | 97.76 |
| S | 97.00 | 97.00 | 100.00 |
| V | 99.00 | 98.00 | 99.42 |
| F | 60.00 | 50.00 | 98.44 |
| Q | 100.00 | 100.00 | 99.87 |
| Accuracy (%) | | 99.00 | |

| Balanced dataset, Train: 1,2,4; Test: 3 | | | |
|---|---|---|---|
| Class | Precision (%) | Sensitivity (%) | Specificity (%) |
| N | 100.00 | 100.00 | 99.99 |
| S | 100.00 | 100.00 | 100.00 |
| V | 97.00 | 100.00 | 99.99 |
| F | 100.00 | 95.00 | 98.67 |
| Q | 98.00 | 100.00 | 99.97 |
| Accuracy (%) | | 99.00 | |

**Table 6.** Confusion matrix of the developed model

| MLP with Unbalanced Data | | | | | |
|---|---|---|---|---|---|
| Class | N | S | V | F | Q |
| N | 5422 | 0 | 0 | 0 | 0 |
| S | 21 | 1386 | 9 | 0 | 0 |
| V | 0 | 42 | 1406 | 8 | 0 |
| F | 1 | 1 | 11 | 12 | 3 |
| Q | 0 | 0 | 0 | 0 | 2075 |
| | | Target class | | | |

Output class

| MLP with Balanced Data | | | | | |
|---|---|---|---|---|---|
| Class | N | S | V | F | Q |
| N | 9997 | 3 | 0 | 0 | 0 |
| S | 0 | 10000 | 0 | 0 | 0 |
| V | 0 | 0 | 9996 | 4 | 0 |
| F | 0 | 0 | 361 | 9464 | 175 |
| Q | 0 | 0 | 0 | 12 | 9988 |
| | | Target class | | | |

Output class

**Table 7.** Precision, sensitivity, and accuracy for each testing dataset considering AAMI standard classes. NN: RF

| Unbalanced dataset, Train: 1,2,3; Test: 4 | | | |
|---|---|---|---|
| Class | Precision (%) | Sensitivity (%) | Specificity (%) |
| N | 98.00 | 100.00 | 99.95 |
| S | 98.00 | 93.00 | 99.05 |
| V | 97.00 | 100.00 | 99.94 |
| F | 00.00 | 00.00 | 99.73 |
| Q | 100.00 | 100.00 | 99.89 |
| Accuracy (%) | | 88.00 | |

| Balanced dataset, Train: 2,3,4; Test: 1 | | | |
|---|---|---|---|
| Class | Precision (%) | Sensitivity (%) | Specificity (%) |
| N | 97.00 | 79.00 | 94.88 |
| S | 98.00 | 61.00 | 99.19 |
| V | 63.00 | 99.00 | 99.58 |
| F | 99.00 | 100.00 | 100.00 |
| Q | 100.00 | 100.00 | 100.00 |
| Accuracy (%) | | 99.00 | |

**Table 8.** Confusion matrix of the developed model

| RF with Unbalanced Data | | | | | |
|---|---|---|---|---|---|
| Class | N | S | V | F | Q |
| N | 5420 | 0 | 2 | 0 | 0 |
| S | 86 | 1312 | 18 | 0 | 0 |
| V | 3 | 2 | 1451 | 0 | 0 |
| F | 3 | 8 | 17 | 0 | 0 |
| Q | 0 | 0 | 9 | 0 | 2066 |
| | | | Target class | | |

Output class

| RF with Balanced Data | | | | | |
|---|---|---|---|---|---|
| Class | N | S | V | F | Q |
| N | 7856 | 7 | 2137 | 0 | 0 |
| S | 256 | 6089 | 3587 | 68 | 0 |
| V | 11 | 129 | 9856 | 4 | 0 |
| F | 0 | 0 | 0 | 10000 | 0 |
| Q | 0 | 0 | 0 | 0 | 10000 |
| | | | Target class | | |

Output class

with balanced data, this problem was solved, however the precision for S class decreased, affecting the classification performance. When comparing the MLP results, all classes were achieved correctly, excepting Fusion class with 60% precision. In this case, the data balancing allowed obtain the best result. The sensitivity also reached high values, above 97%, excepting for F and S class in RF classifier. The study reported in this paper is a work in progress, and the results are the best achieved at the present stage. Various parameters of the algorithm are to be adjusted prior to the final neural network architectures. In conclusion, the proposed model stands out compared to other classification algorithms found in the literature, presenting the best accuracy. The global results of the model demonstrate the robustness of the developed automatic heartbeats classifier. Furthermore, it is worth noting that the stratified division of the datasets and the balancing of the database contributed to improving the network performance.

**Table 9.** Comparison of the proposed method and other algorithms published in the literature

| Paper | Class | Preprocessing | Feature Extract | Accuracy (%) |
|---|---|---|---|---|
| Dalvi et al. [5] | N, L, R, PB, V, A | With and Without PCA | ANN | 96.97/97.41 |
| Rajesh et al. [11] | N, S, V, F, Q | Mean separation **from noisy ECG,** Moving Average filter of order five | AdaBoost EC | 99.1 |
| Sanino & Pietro.[12] | DS1, DS2 | Denoising, Peak Detection, Signal Segmentation Temporal Feature Extraction | DNN | 99.52 |
| **Proposed Method** | **N, S, V, F, Q** | **Moving average filter + Resample** | **MLP / RF** | **99.0 / 99.0** |

## 5    Conclusions

This paper presented a method for automatic heartbeat classification based on principal component analysis and a convolutional neural network on ECG signals. We created a database containing the first ten principal components and the relative RR intervals of P-QRS complexes from the MIT-BIH Arrhythmia Database patients, which were used as input to MLP and RF, allowing us to obtain a classification model for this database. The method proposed showed overall classification accuracy of 99.00%, on the heartbeats classification in the AAMI standard classes (N, S, V, F, and Q). The model's performance proved very satisfactory, demonstrating the advantages of the proposed method.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. Guyton, A., Hall, J.: Textbook of Medical Physiology. Elsevier Saunders, Amsterdam (2015)
2. Luz, E., Schwartz, W., Cámara-Chavez., G., Menotti, D.: ECG-based heartbeat classification for arrhythmia detection: a survey. Comput. Methods Programs Biomed. **127**, 144–164 (2016)
3. Luz, E., Nunes, T., Papa, J., Menotti, D.: ECG arrhythmia classification based on optimum-path forest. Expert Syst. Appl. **40**, 3561–3573 (2013)
4. Ye, C., Kumar, B., Coimbra, M.: Combining general multi-class and specific two-class classifiers for improved customized ECG heartbeat classification. In: 21st International Conference on Pattern Recognition ICPR 2012, pp. 2428–2431, Tsukuba, Japan (2012)
5. Dalvi, R., Zago, G., Aandreão, R.: Heartbeat classification system based on neural networks and dimensionality reduction. Res. Biomed. Eng. **32**(4), 318–326 (2016)
6. Koeppen, B., Stanton, B.: Berne & Levy Fisiologia. Elsevier, Rio de Janeiro (2009)
7. Pinheiro, P.: Eletrocardiograma (ECG): Entenda os resultados (2021). https://www.mdsaude.com/cardiologia/exame-eletrocardiograma-ecg/
8. ANSI/AAMI EC57:2012, testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. American National Standards Institute, Inc. (2012)
9. Moody, G., Mark, R.: The impact of the MIT-BIH arrhythmia database. IEEE Eng. Med. Biol. Mag. **20**, 45–50 (2001)
10. Prati, R.C., Batista, G.E.A.P.A., Monard, M.C.: A study with class imbalance and random sampling for a decision tree learning system. In: Bramer, M. (ed.) IFIP AI 2008. ITIFIP, vol. 276, pp. 131–140. Springer, Boston, MA (2008). https://doi.org/10.1007/978-0-387-09695-7_13
11. Rasesh, K., Dhuli, R.: Classification of imbalanced ECG beats using re-sampling techniques and adaboost ensemble classifier. Biomed. Sig. Process. Control **41**(4), 242–254 (2018)
12. Sannino, G., Pietro, G.D.: A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Futur. Gener. Comput. Syst. **86**, 446–455 (2018)

# Assessment of AlexNet for Oral Epithelial Dysplasia Classification

Viviane Mariano Silva[1]([✉]) [iD], A. L. D. Araújo[2] [iD], F. P. Fonseca[3] [iD], P. A. Vargas[2] [iD], M. A. Lopes[2] [iD], A. R. Santos-Silva[2] [iD], and M. C. Moraes[1] [iD]

[1] Institute of Science and Technology, Federal University of São Paulo (ICT-Unifesp), São José dos Campos, São Paulo, Brazil
viviane.mariano@unifesp.br
[2] Oral Diagnosis Department, Piracicaba Dental School, University of Campinas (UNICAMP), Piracicaba, São Paulo, Brazil
[3] Department of Oral Surgery and Pathology, Federal University of Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil

**Abstract.** The progression of potentially malignant lesions into oral squamous cell carcinoma (OSCC) is assessed based on the dysplasia grading, in which several cytological and architectural changes drive pathologists to assign a malignization risk for proper patient management. The present study implemented AlexNet, a Deep Learning (DL) model based on the binary system for grading oral epithelial dysplasia (OED), the only prognostic hallmark of oral potentially malignant disorders (OPMD). A TOTAL of 63 digital slides from two institutions were used. Annotated regions of interest (ROI) were segmented and fragmented into 56,406 smaller patches of $220 \times 220$ pixels. After data separation into training, validation, and testing sets (80%, 10%, and 10%), data augmentation was conducted in training/validation images. After carrying out the test, the trained AlexNet reached 91.57% accuracy, with 93.55% sensitivity, and 87.57% specificity. F1-score was 0.93857 and AUC was 0.9771. Consequently, showing potential for the proposed application.

**Keywords:** Artificial Intelligence · Oral Potentially Malignant Disorders · Binary System · Oral Cavity

## 1 Introduction

Oral squamous cell carcinoma (OSCC) is a disease of great magnitude worldwide with significant regional variations both in incidence and mortality. The disease has a better prognosis when diagnosed and treated early, which can be challenging if the patient is not under surveillance for oral potentially malignant disorders (OPMD).

OPMD can progress within a period of 5 years after the initial diagnosis [1]. The diagnosis is based on the exclusion of known lesions and further proofed by oral epithelial dysplasia (OED), one of the most important hallmark of malignization risk in OPMD [2]. Progression to OSCC can occur in 2.6% to 29.2% of patients depending on the dysplasia grade, and usually occurs in the dysplastic site [3].

The World Health Organization [4] acknowledged the binary system for OED grading initially proposed by Kujan [5] and recommends it despite the high interobserver disagreement. Attempts to analyze sources of disagreements [6] and reduce interobserver variability [7] when using the system are frequently under investigation.

The use of digital systems and digital image analysis associated with the development of algorithms for image analysis has motivated the use of Artificial Intelligence (AI) in the context of histopathological diagnosis. In this particular scenario, Deep Learning (DL) is employed to aid in the reproducibility of a grading system with a known high subjectivity.

In previous work, Muthu Rama Krishnan [8] segmented and classified the connective tissue cell population associated with oral submucosal fibrosis (OSF) using multi-level thresholding and support vector machine (SVM) as a classifier and reached an accuracy of 88.69%. In sequence, the authors used Higher Order Spectra, Local Binary Pattern, and Laws Texture Energy as feature descriptors and tested five classifiers (Decision Tree, Sugeno Fuzzy, Gaussian Mixture Model, K-Nearest Neighbor, Radial Basis Probabilistic Neural Network) to verify which combination of techniques would culminate into the best performance to classify OSF between dysplastic and non-dysplastic. The combination of Higher Order Spectra + Fuzzy achieved an accuracy of 95.7%.

Specifically, the use of previous and classical machine learning methods requires coherent and tuned feature extraction stages which can be affected by user, technical, or clinical variations. In this way, the use of convolutional neural networks (CNN) is shown as an automation alternative since the different convolutional kernels are responsible for the extraction of attributes. An important recent study used CNNs to grade dysplasia between mild, moderate and severe were published [9]. However, it was only applied to samples of small animals; additionally, in our knowledge, only segmentation [10–12] and classification [13–15] approaches through classical models have been applied in OPMD cases. Therefore, due to the gap involving binary grading of OPMD using deep learning algorithms such as CNN, the purpose of this work is to implement, train and evaluate a DL model for OPMD grading in a realistic pathological context. During the design of this work, the TRIPOD guidelines were used aiming at good practices in diagnostic modeling research, and can thus be classified as type 2b [16, 17].

## 2 Methods

### 2.1 Clinical Dataset

For CNN training, datasets from two institutions were used (Table 1), comprising a total of 63 H&E-stained glass slides of OPMD with personal and institutional variations (Fig. 1A). This study is in accordance with the Declaration of Helsinki and approved by the Piracicaba Dental Ethical Committee, Registration number CAAE: 42235421.9.0000.5418. Glass slides were scanned using the Aperio Digital Pathology System (Leica Biosystems, Wetzlar, Germany) with a spatial sampling of $0.47\mu$m per pixel, with automated focusing and magnification at $\times20$.

**Table 1.** Contribution of institution for build dataset.

| Institution | Patients (n) | WSI (n) | Areas LR (n) | Areas HR (n) | Patches LR (n) | Patches HR (n) |
|---|---|---|---|---|---|---|
| Piracicaba Dental School (Piracicaba, Brazil) | 39 | 48 | 100 | 174 | 12 344 | 28 636 |
| Federal University of Minas Gerais (Belo Horizonte, Brazil) | 8 | 15 | 36 | 57 | 5 814 | 9 612 |

The slides were classified based on the binary system for OED grading [4, 5] due to the system's prognostic significance and specific cytological and architectural changes. Four architectural and five cytological changes were considered as cut-offs for malignization risk and a consensus diagnosis was made when disagreements occurred. The number of architectural and cytological changes may widely vary along the epithelium causing some biopsies to have annotations of both high risk (HR) and low risk (LR) of malignant progression.

## 2.2  Pre-processing and Data Augmentation

First step of pre-processing is the proper delimitation of representative regions of interest (ROI), which were manually annotated for representative areas of HR and LR by experienced pathologists (Fig. 1A). The annotated ROI were segmented (Fig. 1B) and fragmented into small patches (220 × 220 pixels) (Fig. 1C). After this initial step, an eyeball screening was conducted by the biomedical engineer (VMDS) to remove patches with non-representative features (e.g., black squares and smaller sized patches). Patches with proper size and any small amount of tissue were maintained, resulting in a dataset with a total of 56,406 patches (Table 1) (Fig. 1D), which were resized (227 × 227 pixels) to fit the input layer of the DL architecture (Fig. 1E). Data sampling was conducted to randomly divide number of patches into training, validation and testing sets with 80%, 10% and 10%, respectively (Fig. 1F). To increase the robustness of the model and minimize overfitting, the images of the training set were submitted to data augmentation (rotation, mirroring, horizontal and vertical flipping, zoomming) (Fig. 1F).
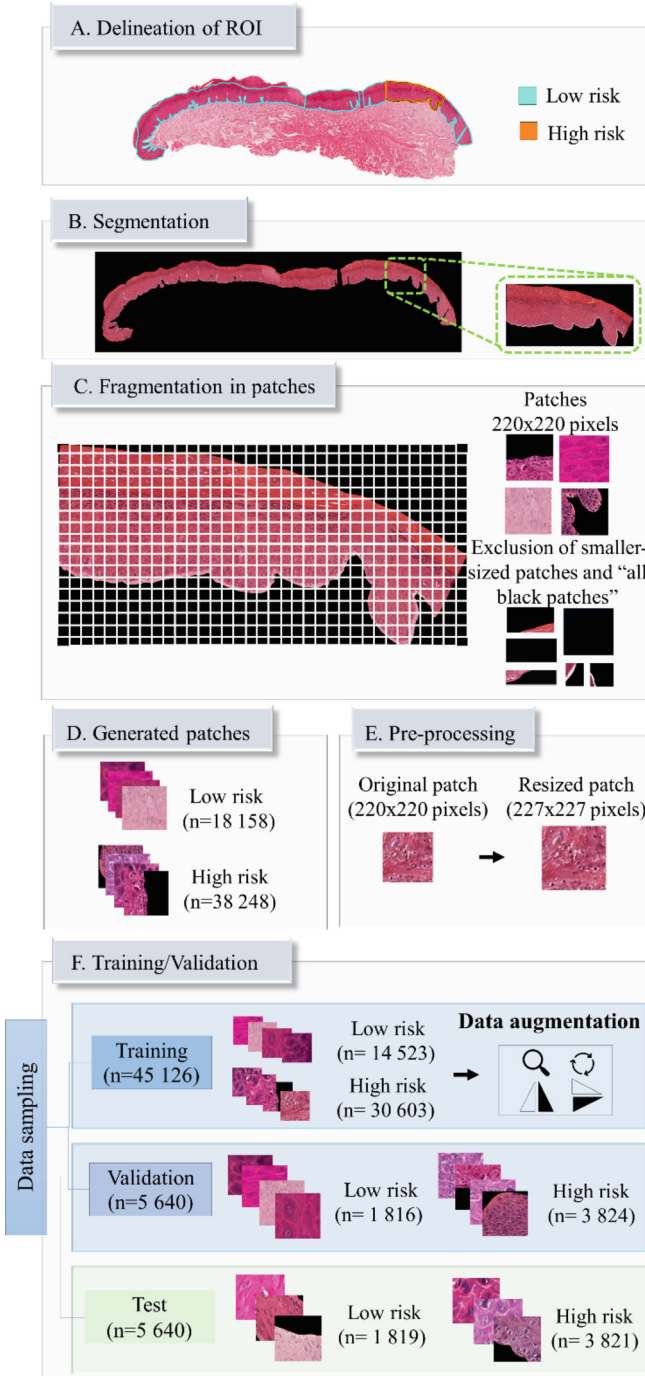
**Fig. 1.** Methodology

## 2.3 Architecture and Implementation

AlexNet [18] was implemented and trained for the classification task. The CNN architecture has a sequential arrangement of five convolutional layers and three fully connected layers (Fig. 2). Dropout and batch normalization layers were added to the network' structure after the activation layers. The activation function chosen was ReLu, with the exception of the last layer FC where softmax was used. The Adam optimizer was chosen to carry out the training, with an initial learning rate lr = 0.001. The algorithm was implemented using Python 3.6 and several open-source libraries specific to machine learning and image processing (TensorFlow, Keras, Scikit-Learn, and OpenCV). The training and validation were carried out for 75 epochs.

CNN training was conducted using an Intel CORE i7 3.50 GHz computer processor with 16 GB RAM and 1TB Hard Disk, available at the Signal and Image Processing Laboratory at the Institute of Science and Technology, Federal University of São Paulo (ICT-Unifesp). After training/validation conclusion, the test was carried out over the CNN with corresponding data (high risk = 3,821 and low risk = 1,819 patches).
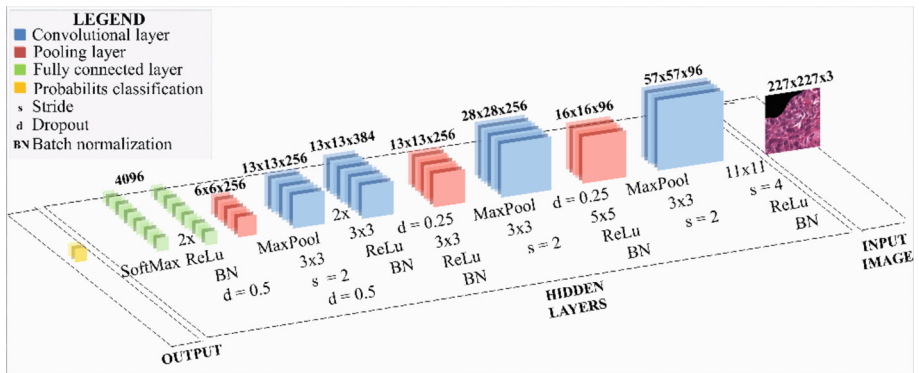


**Fig. 2.** AlexNet convolutional neural network architecture.

## 3 Results

AlexNet training performed in 75 epochs resulted in a model that reached a validation accuracy of 90.32%. The accuracy and loss curves obtained during CNN training are shown in Fig. 3, showing a close behavior between the training and validation accuracy curves. As it can be seen, there is a slight instability in the validation loss. Although there is not a great divergence between the training and validation loss, the irregularities indicate a need for better hyperparameter adjustments and/or a dataset with a greater variety of features.

By performing the test, the corresponding confusion matrix (Fig. 4). Considering HR as True Positive (TP), and LR as True Negative (TN), we obtained the corresponding ratio of TP(%) = 94.16%, TN(%) = 86.37%, FP(%) = 13.63%, FN(%) = 5.83%.
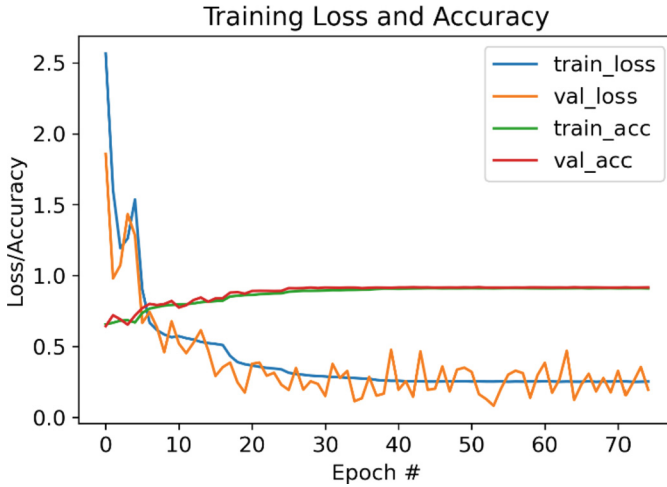
**Fig. 3.** Accuracy and loss curves from training and validation.

Additionally, different metrics were computed through the confusion matrix; resulting in a overall accuracy of 91.57%. The correspondent sensitivity and specificity parameters of the classification achieved 93.55% and 87.57%, respectively. The F1-score metric achieved a magnitude as high as 0.93857. The ROC curve, shown in Fig. 5, shows a great differentiation behavior in the classification task evidenced by an AUC of 0.9771. The model presented satisfactory outcomes; additionally, the higher TP demonstrates the potential of the model to precisely catch the relevant class.
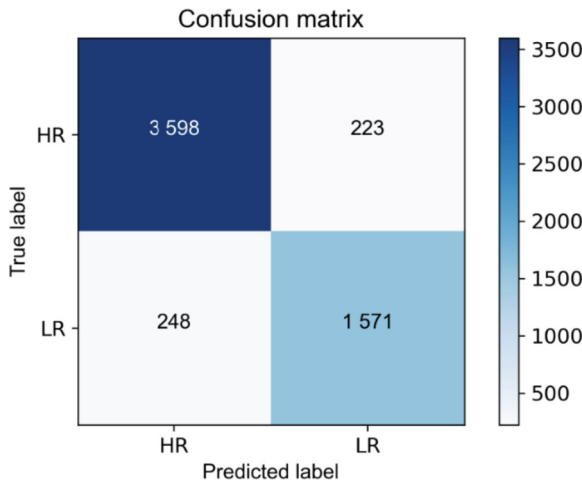


**Fig. 4.** Confusion matrix obtained by the test dataset.

Authors reassessed all 471 misclassified patches to identify possible reasons for misclassifications; as a result, some patterns were noted: a few reminiscent small-sized images (not excluded in the screening prior to training), images with keratinization, and patches with a great amount of black background. Images with artifacts, black gaps in the middle of the cropped image were also seen.
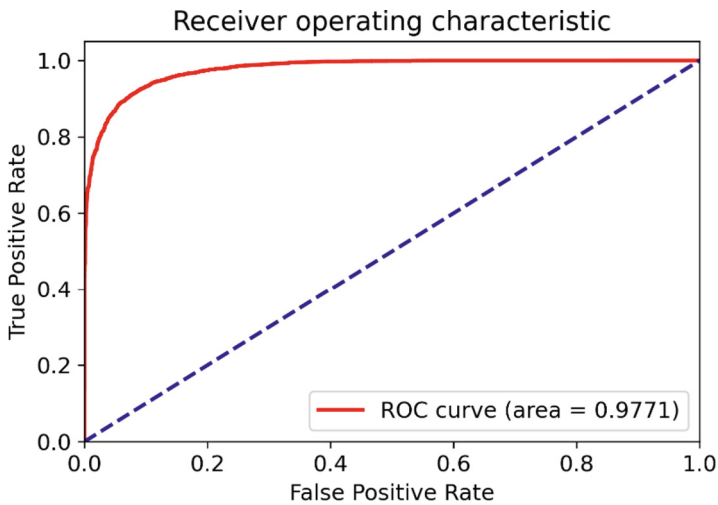


**Fig. 5.** Receiver operating characteristic curve.

## 4   Discussion

The accurate assessment of OED grading allows pathologists to provide targeted information on the progression of OPMD, which aids the clinicians to manage and choose proper interventions according to a low or high malignization risk. Unfortunately, cytological and architectural criteria on OED grading are subjective leading to high rates of interobserver disagreement when addressing the binary system. Speight8 reported that in a group of four observers the level of interobserver agreement can vary from 62% to 79.6%, while in a more recent study, authors indicated an interobserver agreement between 70% and 90%19. Therefore, attempts to provide a support system for pathologists' decisions when grading OED should overcome these values. The accuracy achieved during test assessment by the proposed DL-based architecture is 91.57%, which represents a superior ability to classify OED when compared to the divergence between pathologists.

The present dataset, even with a sufficiently large number of samples, has an imbalance between HR and LR images. This imbalance is not an outlier in clinical practice since the patient remains in close follow-up and biopsy is only performed when the lesion displays significant clinical changes, therefore, also presenting more dysplastic changes corresponding to lesions with higher malignization risk.

Patches with keratinization layers were very frequent within errors, which can be explained by the presence of this characteristic in both classes since the majority of leuko-plakias presents hyperkeratosis. The existence of an intersection between the attributes that characterize each class may explain the almost equally distribution of FP and FN.

Additionally, patches with a very large area of black background were also common in both classes due to the characteristic drop-shaped rete ridges. The exclusion of images with this feature prior to training is not ideal (even inviable) due to the importance in analyzing the basal layer cells. In the present study, attempts were made to avoid major artifacts to be included in ROI annotation. The annotations of the LR and HR region's limits are not precise and may include transitional areas.

Considering the lack of studies in the literature that seek to apply a similar method for OED grading using the binary classification system with AI, this study represents a major innovation in the area, and the accuracy of 91.57% using AlexNet architecture enlight great possibilities for future projects.

## 5   Conclusions

We proposed the use of AlexNet as DL-based system for dysplasia grading in OPMD. The use of DL models aims to eliminate inter-pathologist variability in the analysis of OED, a known pitfall. The accuracy of 91.57% achieved by the architecture represents a superior performance compared with the interobserver agreement. We also aim to approach the ability of CNN to differentiate high and low malignization risk lesions in clinical images from OPMD with annotations based on the outcome and on the histopathological correspondent classification.

**Conflict of Interest.** We declare that the authors have no financial relationship with any commercial associations, current and within the past five years, that might pose a potential, perceived or real conflict of interest. These include grants, patent licensing arrangements, consultancies, stock or other equity ownership, advisory board memberships, or payments for conducting or publicizing our study.

## References

1. Warnakulasuriya, S.: Clinical features and presentation of oral potentially malignant disorders. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. **125**(6), 582–590 (2018). https://doi.org/10.1016/j.oooo.2018.03.011

2. Van der Waal, I.: Potentially malignant disorders of the oral and oropharyngeal mucosa; present concepts of management. Oral Oncol. **46**(6), 423–425 (2010). https://doi.org/10.1016/j.oraloncology.2010.02.016

3. Hankinson, P.M., Mohammed-Ali, R.I., Smith, A.T., Khurram, S.A.: Malignant transformation in a cohort of patients with oral epithelial dysplasia. Br. J. Oral Maxillofac. Surg. **3**, S0266–4356(21)00085–1 (2021). https://doi.org/10.1016/j.bjoms.2021.02.019

4. El-Naggar, A.K., Chan, J., Takata, T., et al.: The 4th edition of the head and neck WHO blue book: editor's perspectives. Hum Pathol. (2017). https://doi.org/10.1016/j.humpath.2017.05.014 (2017)

5. Kujan, O., Oliver, R.J., Khattab, A., Roberts, S.A., Thakker, N., Sloan, P.: Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. Oral Oncol. **42**(10), 987–993 (2006). https://doi.org/10.1016/j.oraloncology.2005.12.014

6. Kujan, O., Khattab, A., Oliver, R.J., Roberts, S.A., Thakker, N., Sloan, P.: Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: an attempt to understand the sources of variation. Oral Oncol. **43**(3), 224–231 (2007). https://doi.org/10.1016/j.oraloncology.2006.03.009

7. Speight, P.M., et al.: Interobserver agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. **120**(4), 474–82.e2 (2015). https://doi.org/10.1016/j.oooo.2015.05.023

8. Muthu Rama Krishnan, M., et al.: Automated classification of cells in sub-epithelial connective tissue of oral sub-mucous fibrosis-an SVM based approach. Comput. Biol. Med. **39**(12), 1096–104 (2009). https://doi.org/10.1016/j.compbiomed.2009.09.004

9. Carvalho, R.H.O.: Nascimento MZ Classification of Oral Cavity Lesions Based on Deep Learning in Remote Environment. (Graduate) Federal University of Uberlândia (2021). https://repositorio.ufu.br/handle/123456789/34391

10. Krishnan, M.M.R., Choudhary, A., Chakraborty, C., Ray, A.K., Paul, RR.: Texture based segmentation of epithelial layer from oral histological images. Micron **42**(6), 632–641 (2011). https://doi.org/10.1016/j.micron.2011.03.003

11. Fraz, M.M., Shaban, M., Graham, S., Khurram, S.A., Rajpoot, N.M.: Uncertainty driven pooling network for microvessel segmentation in routine histology images. In: Stoyanov, D., et al. (eds.) OMIA/COMPAY-2018, pp. 156–164. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00949-6_19

12. Musulin, J., Štifanić, D., Zulijani, A., Ćabov, T., Dekanić, A., Car, Z.: An enhanced histopathology analysis: an AI-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. Cancers (Basel) **13**(8), 1784 (2021). https://doi.org/10.3390/cancers13081784

13. Krishnan, M.M.R., Acharya, U.R., Chakraborty, C., Ray, A.K.: Automated diagnosis of oral cancer using higher order spectra features and local binary pattern: a comparative study. Technol. Cancer Res. Treat. **10**(5), 443–455 (2011). https://doi.org/10.7785/tcrt.2012.500221

14. Krishnan, M.M.R., Venkatraghavan, V., Acharya, U.R., Pal, M., Paul, R.R., Min, L.C., et al.: Automated oral cancer identification using histopathological images: a hybrid feature extraction paradigm. Micron **43**(2–3):352–364 (2012). https://doi.org/10.1016/j.micron.2011.09.016

15. Krishnan, M.M.R., Shah, P., Chakraborty, C., Ray, A.K.: Statistical analysis of textural features for improved classification of oral histopathological images. J. Med. Syst. **36**(2), 865–881 (2012). https://doi.org/10.1007/s10916-010-9550-8

16. Moons, K.G.M., et al.: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann. Intern. Med. **162**(1), W1–73 (2015). https://doi.org/10.7326/M14-0698

17. Collins, G.S., Dhiman, P., Navarro, C.L.A., et al.: Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. **11**(7), e048008 (2021). https://doi.org/10.1136/bmjopen-2020-048008
18. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105 (2012)

# EEG Synchronization and Desynchronization Associated with Non-painful Thermal Stimuli

D. C. Santos-Cuevas[1(✉)] , D. D. Collina[2] , and C. J. Tierra-Criollo[1]

[1] Federal University of Rio de Janeiro, Rio de Janeiro 21941-914, Brazil
dianasantosc@peb.ufrj.br
[2] Federal Center for Technological Education of Minas Gerais,
Belo Horizonte, Minas Gerais 30510-000, Brazil

**Abstract.** Brain processes associated with non-painful thermal stimulation in the human brain have been little studied. The electroencephalography (EEG) signal can be used to evaluate objectively the functionality of small fibers (A$\delta$ and C) related to thermal stimulation. In the literature there is no agreement on the EEG signal modulation related to non-painful thermal stimulus. The aim of this work was to investigate the behavior of different brain rhythms of EEG signal during non-painful thermal warming and cooling stimuli. EEG signals were acquired from 14 subjects (men). The stimulus was applied to the right forearm by means of a thermode (Peltier), during a period of warming ($32\,°C \rightarrow 40\,°C$) and cooling ($32\,°C \rightarrow 24\,°C$). Event-related synchronization and desynchronization (ERS/ERD) showed a significant difference ($p < 0.05$), between warming and cooling, in the delta, theta and alpha bands, mainly in the ipsilateral temporal region. This finding suggests that non-painful warm and non-painful cold stimuli modulate the EEG signals differently.

**Keywords:** Brain oscillations · non-painful thermic stimulation · EEG

## 1 Introduction

Small-diameter peripheral neuropathies can occur by injury on myelinated (A$\delta$) and unmyelinated (C) thin fibers [1–3]. These types of fibers are related to pain and temperature sensations (warm and cold), and may evoke non-painful or painful sensations in humans, depending on the variation in the intensity of the stimulus applied. The stimuli are classified as: i) non-painful thermal stimulus within the warm range of $[30\,°C – 45\,°C]$ and for cold in the range $[17\,°C – 35\,°C]$; (ii) painful thermal stimuli have temperatures above 45 °C and below 5 °C for warming and cooling, respectively [4, 5].

Loss of thermal sensitivity can happen in peripheral neuropathies caused, for example, by diabetes [6] and Hansen's disease [7].

Diagnosing techniques for thermal sensitivity can be: (i) subjective, thus dependent on the individual's response to thermal stimuli (commonly used in clinical assessment), called psychophysical assessment, where a relationship is established between the intensity of the stimulus and the individual's perception, e.g. tests that use test tubes filled with

warm or cold water and thermal sensitivity quantification tests (QST-T) [8]; (ii) objective techniques that do not depend on the individual's response, e.g. functional magnetic resonance imaging fRMI [9–13], positron emission tomography PET [14, 15], based on electroencephalography (laser transient evoked potential [16], CHEPs and thermode steady-state response [17, 18]). Objective techniques can be used with individuals with cognitive limitations.

Steady-state studies based on EEG signals have focused on analyzing the behavior of brain rhythms. Changes in the power of these brain rhythms related to a thermal stimulus (event) have been identified by a decrease in oscillatory activity (event-related desynchronization - ERD) or by an increase in oscillatory activity (event-related synchronization - ERS). The most common thermal stimulation procedures are: stimulation, warm and cold water [19, 20], contact thermode [17, 18] and laser [16].

Studies with EEG signals related to painful stimuli, warm and cold, have been much more frequent than with non-painful stimuli. The first study with non-painful thermal warm and cold stimulation was recently reported [19]. The authors stimulated the subjects' left hand in (warm/cold) water and, found different EEG power spectra for both thermal stimuli as compared to baseline. They also found a decrease in power in the delta and alpha bands for both stimuli. In addition, the authors reported a decrease in theta activity in the contralateral area only for cold.

In another study [20] with EEG signals, two types of stimuli were used (water and metal thermostat) with different non-painful warm temperatures. The results showed that the neural responses in the different EEG bands were sensitive to the type of stimulus. In both stimuli, an increase in the alpha band in all regions of the cortex. An inverse behavior of power was observed in the theta band, which increased with the thermostat and a decreased with water. In addition, the power of the beta band increased only for water-based stimulation, in the bilateral frontal and right temporal regions.

In the study [17], the authors performed non-painful cold thermal stimuli, in the left hand with a thermode in a temperature range of [35 °C to 30 °C]. An increase in power was found in the delta, theta and gamma bands and a decrease in power in the alpha and beta bands. In the study [18] the authors performed thermal stimulation with laser and thermode, stimulating the subjects' right forehead with a temperature range of 33 °C to 41 °C. The results showed a decrease in the power in the alpha and beta bands to the thermode stimulus, while with the laser the changes were inverse.

In the work [21] they reported that they used personalized thermal stimulators. Stimulations were performed with five different temperatures on the right hand: painful cold (15 °C); painful warm (45 °C); non-painful cold (25 °C); non-painful warm (40 °C) and; baseline (33 °C). Painful thermal stimuli were found to provoke a greater decrease in alpha power compared to non-painful stimuli.

Information on brain processing of non-painful thermal stimuli is still under discussion, with little consensus on the behavior of the corresponding brain rhythms in EEG signals, requiring further investigation.

The aim of the present study was to investigate the behavior of brain rhythms on the EEG during non-painful thermal warm and cold stimulation applied by means of thermode (Peltier).

## 2   Materials and Methods

### 2.1   Subjects

EEG data were recorded from 14 healthy subjects, aged 21–50 years old (32.3 ± 11.4 years), all male, right-handed, and none of them had a history of neurological disorder, or any drug use that affects nervous system function. This experiment was approved by the Ethics Committee of the Federal University of Minas Gerais (CAAE-0450.0.203.000–11). In accordance with the Declaration of Helsinki, consent was obtained from all subjects before they entered this study.

### 2.2   Thermal Stimulation

All stimuli considered in this study were generated by a thermal stimulator (Biostimulator – UFMG, [22]), whose main function is temperature control by means of a thermal stimulating probe (thermode) based on a Peltier module with a flat stimulation surface of 30 mm x 30 mm, the temperature range is from 0 ℃ to 50 ℃, with a resolution of 0.1℃ and the temperature variation rate up to 5 ℃/s (in the present study, 1 ℃/s was used).

### 2.3   Design Experiment

The study design consisted of the following consecutive steps: (i) preparation - All subjects were positioned in an armchair, with the right upper limb resting on the arm support of the armchair. The thermode was positioned in the proximal third of the inner region of the right forearm (see Fig. 1a). The subject was instructed to remain relaxed, with eyes open and with the gaze fixed on a point marked on the wall in front of him and asked to move as little as possible during the data collection. The experiments took place in a controlled environment, with a temperature stabilized at 25 °C; (ii) spontaneous EEG signal acquisition for 3 min without thermal stimulation; (iii) determination of thresholds of perception to warm (LPC) and to cold (LPF) - the sensitivity to warm and cold was evaluated, each individual being instructed to press a button as soon as he perceived the warming and the cooling starting at a baseline of 32 °C and not exceeding the limits of 40 °C for warming and 24 °C for cooling; and (iv) Recording of the EEG signal during thermal stimulation: during five 15 min blocks of stimulation (see Fig. 1b) - with approximately 28 epochs (see Fig. 1c) each block; and 3 min at baseline between blocks.

Each epoch consisted of 6 stimulation stretches (Tr) with a duration of approximately 5 s each Tr1 is moderate warming, in the range of (32 °C → 40 °C); Tr2 cooling (40 °C → 32 °C); Tr3 baseline (32 °C); Tr4 is moderate cooling, in the range of (32 °C → 24 °C); Tr5 warming (24 °C → 32 °C); Tr6 baseline. A total of 124 epochs (approximately 1 h and 40 min of examination).
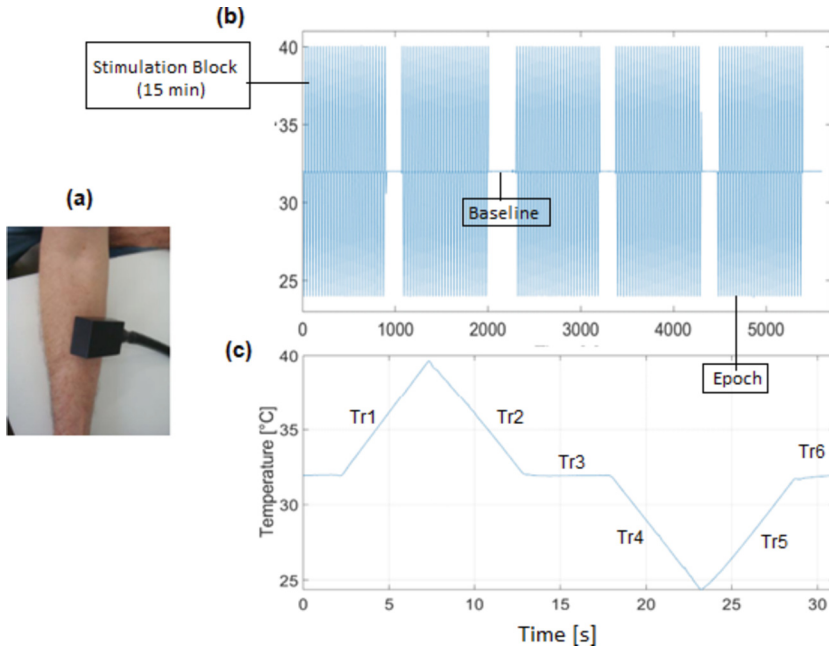
**Fig. 1.** Experiment Design a) Stimulation setup. The thermode was place at the in the proximal third of the inner region on the subject's right forearm b) stimulation blocks with approximately 15 s and 28 epochs each; c) representation epoch with 6 stimulation stretches with a duration of approximately 5 s each

### 2.4 Acquisition of EEG Signals

The EEG signals were recorded by the BrainNet BNT–36 equipment (EMSA – Rio de Janeiro), with a high pass filter of 0.1Hz and a low pass of 100 Hz, a sampling frequency of 600 Hz and a notch filter of 60Hz. The EEG electrodes were located according to the international 10–20 system using silver/silver chloride electrodes positioned in a cap. Contact with the scalp and electrode was performed with an electrolytic gel. The reference was binaural and the ground electrode was placed on the forehead. Eye movements and eye blinks were recorded using two additional surface electrodes placed one above and one below the right eye. The impedance was kept at less than 10 k$\Omega$. The thermal stimulation instants were synchronized with the EEG recording throughout the collection.

### 2.5 EEG Signal Processing

All EEG processing steps were performed using the MATLAB (MathWorks, Natick, MA, USA). EEG signals were subsampled at 300 Hz. The signals were then passband filtered from 1 to 45 Hz using a 4$^{th}$ order bidirectional (zero phase) Butterworth filter.

The EEG signal epochs underwent artifact rejection, using the technique of rejection of 3 times the standard deviation with the following algorithm (a) Compute the standard

deviation σ of a section of the signal that, visually inspected, is considered non-artifact; (b) Exclude epochs in which more than 5% of the continuous samples, or a total of more than 10% of the samples exceeded the threshold of ±3 σ [23]. After artifact rejection, 80 epochs were considered artifact-free for all electrodes. Subsequently, the Butterworth filter (explained above) was applied in the following frequency ranges: delta (1–4 Hz); theta (4–8 Hz); alpha (8–13 Hz); beta (13–30 Hz) and gamma (30–45 Hz).

## 2.6   Synchronization and Desynchronization Analysis (ERS/ERD)

Event-related synchronization and desynchronization (ERS/ERD) analysis provides a measure of relative power in a frequency band (relative to a reference). A decrease in signal power is defined as desynchronization (ERD) [24]. The opposite is called event-related synchronization (ERS) [24]. First, bandpass filtering must be performed depending on the band of interest.

The ERS/ERD is defined as:

$$P_{j,f} = \frac{1}{L-1} \sum_{i=1}^{L} \left( x_{ij,f} - \overline{x_{j,f}} \right)^2 \tag{1}$$

$$R_f = \frac{1}{k} \sum_{j=r_0}^{r_0+k} P_{j,f} \tag{2}$$

$$ERS/ERD_{j,f} = \frac{P_{j,f} - R_f}{R_f} \times 100\% \tag{3}$$

where $L$ is the total number of epochs; $x_{ij,f}$ is the sample j of the epoch i of the signal already filtered in the band of interest $f$; $\overline{x_{j,f}}$ is the average value of sample $j$ from all epochs in the band of interest $f$; $R_f$ is the average power in the reference window (base value), which goes from sample $r_0$ to sample $r_o + k$; $ERS/ERD_{j,f}$ is expressed as a percentage value.

## 2.7   Interval of EEG Signal of Interest

The ERS/ERD analysis was performed on: (i) warm interval that corresponds to higher temperatures than LPC (see Fig. 2a) with a reference window of 1 s (black box) before starting the stimulus; (ii) Cold interval that corresponds to temperatures below the LPF (see Fig. 2b) with a reference window of 1 s (black box) before starting the cold stimulus.

Once the ERS/ERD values were obtained for each subject, the general average (Grand Average) of the subjects was performed. Topographic maps were constructed with the average ERS/ERD for each interval.
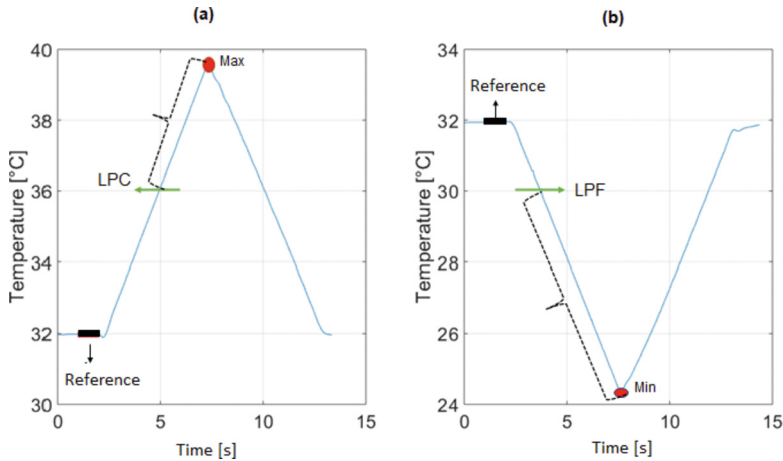
**Fig. 2.** Interval of EEG signal of interest a) Warm interval corresponds to higher temperatures than LPC; b) Cold interval corresponds to temperatures below the LPF a) Warm Segment; b) Cold Segment

## 2.8 Statistical Analysis

The ERS/ERD values for the warm and cold intervals, for each brain rhythm, were statistically compared (n = 14 subjects) using the paired Wilcoxon test ($\alpha = 5\%$), whit the null hypothesis of no difference in the ERS/ERD between the non-painful warm and cold intervals.

## 3 Results

The average thermal perception threshold (10 repetitions) for each subject was approximately 36 °C for warm and 30 °C for cold.

Figure 3. Shows the topographic maps of the ERS/ERD grand average of the 14 subjects (see Fig. 3).

In the delta band desynchronization (ERD) was observed in almost all leads (except Cz, Pz and O2 which showed ERS) for the warm interval in relation to the baseline (32°C). In the cold interval there was synchronization (ERS) in most leads (except in C3, Cz, C4 and P4 which showed ERD).

Significant difference between the warm and cold intervals ($p < 0.05$) was found only in leads F3 ($p = 0.0295$) frontal contralateral to the stimulus and T6 ($p = 0.0494$) temporal ipsilateral to the stimulus.

In the theta band, a global ERD was observed in all leads for the warm interval, while in the cold interval ERS was observed in almost all leads (except in C3, Cz, C4, P3 and Pz and O2, which presented ERS). No significant difference was found.

In the alpha band, an ERD was identified in all leads for the warm interval, while for the cold interval there was an ERS in almost all leads (except F3, Fz, F4, F8 and C3 which showed an ERD). A significant difference between the warm and cold intervals was found in leads P4($p = 0.0295$) and T6($p = 0.0085$) ipsilateral to the stimulus.
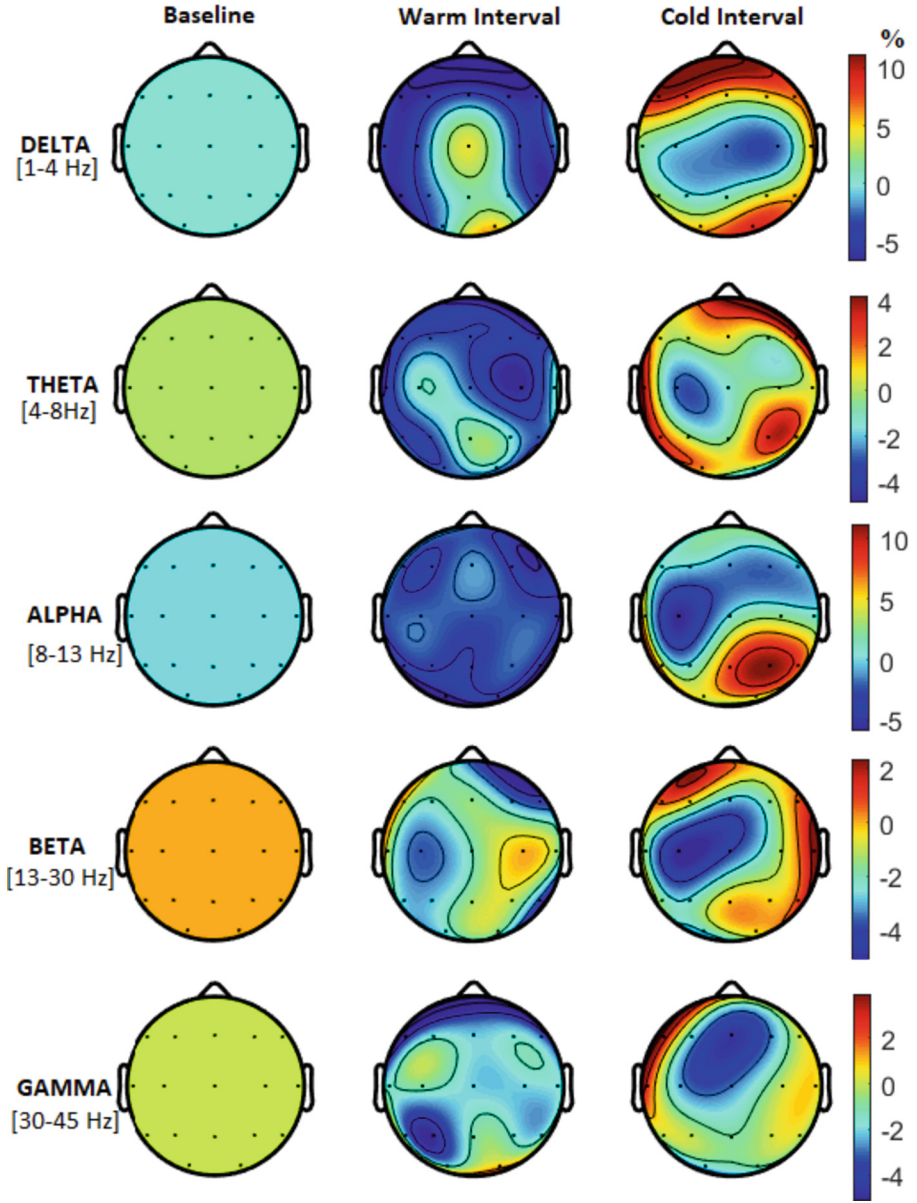
**Fig. 3.** Topographic Maps (group-average) ERS/ERD during baseline, warm and cold in delta, beta, alpha, beta and gamma frequency bands

The beta band showed an ERD for all leads in the warm interval. In the cold interval an ERS was observed in almost all electrodes (except F7, F8, T4, Pz and T6 which exhibited an ERS). A significant difference between the warm and cold intervals was found in lead T6 (p = 0.0107);

Finally, in the gamma band, an ERD was observed in all leads, for the warm t interval. While for the cold interval it showed an ERS in leads F7, F8, T3, T4, P3, P4 and T6 and an ERD in leads F3, Fz, F4, C3, Cz, C4, T5, P6, O1 and O2. No statistical difference was found.

## 4   Discussion

This study analyzed the behavior of brain rhythms in the EEG during non-painful thermal stimulation of warm and cold, applied with a Peltier thermode.

In the non-painful warm interval, all bands (delta, theta, beta, alpha and gamma) showed ERD as compared to the 32 °C baseline. Similar results were found in other study [19] with non-painful thermal warm stimulation with immersion of the left hand in a container with water at temperatures between 40 °C–43ºC, but they did not analyze the gamma band. Baseline was obtained with EEG signal recorded at an average temperature between the warm and cold stimulus (12 °C–15ºC). An opposite result was found in [20] showing an increase in power (ERS) in the bands: alpha (in all leads); and in beta (contralateral). The authors used two reference baselines of 25 °C and 32 °C. In addition, stimulation was performed with water and thermostat.

The non-painful cold stimuli in the present study elicited an ERS in the delta, theta, beta and gamma bands, while an ERD was observed in the alpha band (in the frontal region and in the contralateral central region). Similar results in delta, alpha and gamma bands were reported in [17]; in theta and beta bands the findings were the opposite, i.e., an ERD.

When statistically comparing the ERS/ERD, between the intervals of warm and cold, no significant difference was found in any of the bands in another work [19]. In the present study, a significant difference ($p < 0.05$) was found between the warm and cold intervals, in the delta (F3, T6), alpha (P4 and T6), and beta (T6) bands. Note that the ipsilateral temporal region (T6) is in all these bands. It is important to mention that the temporal lobe is a region related to memory management, sensory recognition, attention [25], and homeostatic emotions (temperature, itching and pain) [26].

Each of the temporal lobes border the insula and the limbic system (composed of the anterior cingulate cortex -ACC, tonsils, hypothalamus, among others). Studies with fMRI [9–13] and PET [14, 15] have shown that non-painful thermal stimulation mainly activates the anterior cingulate cortex (aACC), insula and thalamus. One study [13] showed that most cold-related activations were in the posterior insula, and warm-related insular activations were found in the anterior insula.

In the present study, the significant difference found in the ipsilateral parietal lobe between non-painful warm and cold suggests that this region may reflect a different modulation of the EEG signal for both stimuli. However, further studies should be performed to gather more evidence on this result.

## 5   Conclusion

In this study, a different modulation in the EEG signal was observed due to non-painful thermal stimuli of cold and warm, using ERS/ERD, mainly in the ipsilateral temporal region. In the literature, there is little consensus on the behavior of different brain

rhythms to non-painful thermal stimuli. The lack of consistent responses may in part be a consequence of the different methods used in the application of thermal stimuli, which do not allow a sufficiently uniform experimental control. Further investigations would be needed to understand the connections between different brain regions related to non-painful thermal stimuli of warm and cold. To this end, other EEG signal processing techniques, in addition to power, can be explored, such as, for example, techniques to assess the flow of cerebral information and entropy.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. Jäning, W.: Peripheral thermoreceptors in innocuous temperature detection. Thermoregulation: From Basic Neuroscience to Clinical Neurology. Elsevier. Handb Clin Neurol Part I. Elsevier (2018)
2. Teliban, A., Bartsch, F., Struck, M., Baron, R., Jänig, W.: Axonal thermosensitivity and mechanosensitivity of cutaneous afferent neurons. Eur. J. Neurosci. **33**(1), 110–118 (2011)
3. Schepers, R.J., Ringkamp, M.: Thermoreceptors and thermosensitive afferents. Neurosci. Biobehav. Rev. **34**(2), 177–184 (2010)
4. Greenstein, B., Greenstein, A.: Color Atlas of Neuroscience: Neuroanatomy and Neurophysiology, 2nd edn. Thieme, Stuttgart -New York (2000)
5. Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., Hudspeth, A.J., Mack, S.: Principles of Neural Science. 5nd edn. McGraw Hill (2014)
6. Zhou, L.: Small fiber neuropathy. Semin. Neurol. **39**(05), 570–577 (2019)
7. Vijayan, B.V., Dominic, M.R., Nair, V.C.P.: Leprous neuropathy: observational study highlighting the role of electrophysiology in early diagnosis. J. Neurosci. Rural Pract. **12**(3), 530–534 (2021)
8. Fruhstorfer, H., Lindblom, U., Schmidt, W.C.: Method for quantitative estimation of thermal thresholds in patients. J. Neurol. Neurosurg. Psychiatry **39**(11), 1071–1075 (1976)
9. Becerra, L.R., et al.: Human brain activation under controlled thermal stimulation and habituation to noxious warm: an fMRI study. Magn. Reson. Med. **41**(5), 1044–1057 (1999)
10. Yagishita, T., Sadato, N., Okada, T, et al.: Brain activation by thermal stimulation in humans studied with fMRI. In: Yutaka, T, Tadakatsu, O. (eds). Elsevier Ergonomics Book Series, vol 3, pp. 17–20 (2005)
11. Tseng, M.T., Tseng, W.Y., Chao, C.C., Lin, H.E., Hsieh, S.T.: Distinct and shared cerebral activations in processing innocuous versus noxious contact warm revealed by functional magnetic resonance imaging. Hum. Brain Mapp. **31**(5), 743–757 (2010)
12. Kwana, C.L., Crawley, A.P., Mikulis, D.J., Davis, K.D.: An fMRI study of the anterior cingulate cortex and surrounding medial wall activations evoked by noxious cutaneous warm and cold stimuli. Pain **85**(3), 359–374 (2000)
13. Davis, K.D., Kwan, C.L., Crawley, A.P., Mikulis, D.J.: Functional MRI study of thalamic and cortical activations evoked by cutaneous warm, cold, and tactile stimuli. J. Neurophysiol. **80**(3), 1533–1546 (1998)

14. Minoshima, S., Casey, K.L.: Cerebral responses to warmth and warm and cold pain measured by positron emission tomography. Curr. Rev. Pain **3**(4), 316–320 (1999)
15. Craig, A.D., Chen, K., Bandy, D., Reiman, E.M.: Thermosensory activation of insular cortex. Nat. Neurosci. **3**(2), 184–190 (2000)
16. Mulders, D., et al.: Dynamics of the perception and EEG signals triggered by tonic warm and cool stimulation. PLoS One **15**(4), e0231698 (2020)
17. Fardo, F., Vinding, M.C., Allen, M., Jensen, T.S., Finnerup, N.B.: Delta and gamma oscillations in operculo-insular cortex underlie innocuous cold thermosensation. J. Neurophysiol. **117**(5), 1959–1968 (2017)
18. Wang, X., Wanniarachchi, H., Wu, A., Gonzalez-Lima, F., Liu, H.: Transcranial photobiomodulation and thermal stimulation induce distinct topographies of EEG alpha and beta power changes in healthy humans. Sci. Rep. **11**, 18917 (2021)
19. Chang, P.F., Arendt-Nielsen, L., Chen, A.C.: Comparative cerebral responses to non-painful warm vs. cold stimuli in man: EEG power spectra and coherence. Int. J. Psychophysiol. **55**(1), 73–83 (2005)
20. Lv, B., Su, C., Yang, L., Wu, T.: Effects of stimulus mode and ambient temperature on cerebral responses to local thermal stimulation: an EEG study. Int. J. Psychophysiol. **113**, 17–22 (2017)
21. Tayeb, Z., Dragomir, A., Lee, J.H.: Distinct spatio-temporal and spectral brain patterns for different thermal stimuli perception. Sci. Rep. **12**, 919 (2022)
22. Collina, D.D.: Quantificação de limiares térmicos em fibras finas. Tese (Doutorado em Engenharia Elétrica). Universidade Federal de Minas Gerais. Brazil (2012)
23. Tierra-Criollo, C.J.: Monitorização objetiva da resposta à estimulação somato-sensitiva utilizando parâmetros espectrais. Tese (Doutorado em Engenharia Biomédica), COPPE/UFRJ, Rio de Janeiro. COPPE/UFRJ. Brasil (2001)
24. Pfurtscheller, G., Lopes da Silva, F.H.: Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. **110**(11), 1842–1857 (1999)
25. Chen, A.C., Rappelsberger, P.: Brain and human pain: topographic EEG amplitude and coherence mapping. Brain Topogr. **7**(2), 129–140 (1994)
26. Craig, A.D.: Interoception: the sense of the physiological condition of the body. Curr. Opin. Neurobiol. **13**(4), 500–505 (2003)

# Biomedical Optics and Systems and Technologies for Therapy and Diagnosis

# Qualitative and Quantitative Analysis of Bulk-Fill Composite Resin Restorations Using Optical Coherence Tomography

Fernanda Kely C. Santos and Patricia A. Ana(✉)

Federal University of ABC, Arcturus 03, Sao Bernardo do Campo 09606-070, Brazil
patricia.ana@ufabc.edu.br

**Abstract.** The optical coherence tomography (OCT) allows the imaging of soft and hard tissues at real time with high resolution. Although studies that use OCT to evaluate marginal defects of dental restorations are reported, there are no studies that evaluate the microstructural defects of bulk-fill composite resins. Here, it is proposed a methodology for qualitative and quantitative analysis of bulk fill composite resins restorations using OCT. Thirty human molar teeth with class II cavities were randomly assigned to 6 experimental groups to be restored with six different bulk-fill composite resins. All restorations were scanned with an OCT system (930 nm) and images were analyzed using the ImageJ software. It was measured the marginal gaps, and the presence of bubbles, filler accumulations and cracks. The qualitative analysis also evaluated the presence of steps, material excess and defects resulting from the finishing and polishing of the restorations. Data were statistically analyzed (ANOVA, $p < 0.05$). Gaps were found in most of the restorations performed, and some significant differences between the different resins were found, considering the size of the gaps detected. The Surefil SDR resin showed the highest number of internal bubbles and defects, while the X-tra Base resin showed no bubbles or structural defects. Detection of polishing errors, excess of material and cracks were also possible with the use of OCT. It was demonstrated the feasibility of this technique in quality assessment and monitoring of clinical performance of class II restorations performed with bulk-fill composite resins, which can predict the longevity of this treatment.

**Keywords:** Optical coherence tomography · Enamel · Composite resin

## 1 Introduction

Optical coherence tomography (OCT) is a diagnostic technique that allows obtaining high-resolution cross-sectional images in in vitro, ex vivo and in vivo analyses, in real time. Currently, OCT is widely used in ophthalmology, and can be considered a gold standard exam routine for diagnosing diseases that affect the retina, such as glaucoma or macular degeneration [1]. Typically, an OCT image reaches a depth of about 2 mm in scattering tissues such as skin, and up to 2 cm in tissues that are transparent to the near-infrared wavelengths, such as the human eye [2].

Recently, OCT has been used for several purposes, expanding its application in areas such as dermatology, dentistry, and cardiology. New techniques aim to reduce the time to perform the scans or even extract other information from the backscattered profile, such as the functionality of the tissue detected by the polarization of the light reflected by the tissue. Several light sources and combinations with mirrors are also being tested to increase the resolution and image quality of the technique [3].

The advances in the technology of OCTs are remarkable. In mid-2009 the common spatial resolution of the technique was around 20 μm. The literature has shown that the spatial resolution along the axial axis (depth) can reach 7.7 μm or even 4.5 μm depending on the tissue composition [4]. Today, with the implementation of the OCT technique in the frequency domain, it is possible to obtain resolution of 3 μm [5].

In dentistry, OCT is a method for the early diagnosis of caries, periodontal disease, and oral cancer [6]. Due to its excellent spatial resolution, OCT is ahead of interproximal radiography for detecting pathologies in the early stages [7]. Furthermore, the possibility of 3D image reconstruction overcomes another limitation of dental radiography. Studies have also shown that the use of OCT for detecting microfractures and rupture lines is reliable and even more efficient than techniques such as transillumination, proving to be efficient in the detection of pathologies such as cracked tooth syndrome [8]. OCT can also be an efficient tool in the detection of marginal failures of dental restorations, especially compo-site resin restorations located on the proximal (mesial and distal) surfaces of the teeth, as well as in monitoring the wear of these restorations. It is also an advantage considering the limited resolution of the interproximal radiographs or the absence of direct visualization through conventional clinical examination (performed with an exploratory probe and clinical mirror). However, there are still no studies that assess the adaptation and marginal sealing of bulk-fill composite resins (that have low shrinkage stress) through OCT, which motivated the present study.

In this way, this study aimed to evaluate the marginal integrity and quantify the defects in bulk-fill composite resin restorations performed in human molar teeth by OCT technique.

## 2   Materials and Methods

### 2.1   Experimental Design

A randomized *in vitro* study was carried out with 30 human molar teeth with OM (occlusal-mesial) or OD (occlusal-distal) cavities, which were randomly distributed into 6 different experimental groups of 5 samples each, to be restored with 6 types of compos-ite resin. Afterwards, both restorations of each tooth were evaluated by OCT, in which 8 scans were performed on each restoration, four in the bucco-lingual direction and four in the cervico-occlusal direction. The images were evaluated using ImageJ software, and the dimensions of the marginal gaps (tooth-restoration interface) on the buccal, lingual and cervical walls of the restorations, as well as the presence of irregularities and bubbles, were quantified. Statistical analysis was performed using Analysis of Variance and Tukey's test, considering the significance level of 5%. Statistical analysis was performed considering the restorations as an experimental unit, the type of composite resin as a variation factor and the distance, in micrometers, as the response variable.

## 2.2   Samples Preparation

Thirty intact human third molar teeth were provided by the Biobank of Human Teeth of the School of Dentistry of University of Sao Paulo. This work was carried out in ac-cordance with the approval of the Research Ethics Council (CEP) of the Federal University of ABC, under the number CAAE 49461415.1.0000.5594.

After cleaning and removing organic material, two 3x3x1 mm cavities (OM and OD) were made in each of the teeth using high-speed diamond drills. The samples were then randomly distributed into six experimental groups with 5 teeth each, where each group was restored with a different composite resin, as shown in Table 1.

**Table 1.**  Characteristics and composition of the composite resins used in this study.

| Composite resin | Type | Matrix | Dispersed phase |
|---|---|---|---|
| SonicFill (Kerr) | nanohybrid | Bis-GMA, TEGDMA, EBPADMA | $SiO_2$, glasses and oxides. 83.5% by weight, 67% by volume |
| Surefil SDR (Dentsply) | Low viscosity for basis | Modified UDMA, EBPADMA, TEGDMA | Ba-Al-F-B-Si glass, St-Al-F-Si glass. 68% by weight, 44% by volume |
| Tetric EvoCeram (Ivoclar Vivadent) | nanohybrid | Bis-GMA, UDMA, EBPADMA | Ba-Al-Si glass, prepolymerized particles, spherical oxides. 79–81% by weight, 60–61% by volume |
| IPS Empress Direct (Ivoclar Vivadent) | nanohybrid | vitreous matrix | Ba-Al-Si glass, ytterbium trifluoride, silica dioxide, mixed oxides. 75–79% by weight |
| X-tra Base (Voco) | Low viscosity for basis | MMA, Bis-EMA | Inorganic grains, pyrogenic silicic acid. 75% by weight and 58% by volume |
| Filtek Bulk Fill (3M) | Low viscosity for basis | Bis-GMA, TEGDMA, Bis-EMA, EBPADMA | Zirconia, silica, ytterbium trifluoride. 64.5% by weight, 42.5% by volume |

To perform the restorations, all walls were etched with 37% phosphoric acid (FGM Produtos Odontológicos, Brazil) for 15 s; then, the cavities were washed with water for 10 s and dried with a light jet of air. Afterwards, the adhesive system (Adper Single

Bond 2, 3M ESPE, USA) was applied with a micro brush, followed by light curing for 10 s. The composite resins were inserted into the cavities in a single layer using an anti-adherent spatula and polymerized using a halogen lamp light curing light (Optilight Plus™/Gnatus, Brazil) for 20 s, following the resin manufacturer's instructions.

The samples were then stored in distilled water at room temperature for 24 h. Then, the excess of resin was removed using fine diamond burs, subsequently finishing was carried out using a low-speed handpiece with rubber tips and polishing paste with aluminum oxide grains.

## 2.3 OCT Evaluation

A commercially available optical coherence tomography system was used (Callisto 930 nm/Thorlabs, USA). This is a spectral domain system, which has a super luminescent diode as a light source and emits light with a central wavelength of 930 nm. The spatial resolution of the system is 7 µm, achieving a depth of up to 1.7 mm.

For imaging, the probe was positioned at a right angle to the surface of the restoration at 30 mm. In each restoration, four vertical scans (cervico-occlusal direction) and four horizontal scans (bucco-lingual direction) were performed as shown in Fig. 1. In the cervico-occlusal scans, it was evaluated the cervical wall of the restorations (Fig. 2), while that, in the bucco-lingual scans, it was evaluated the buccal and lingual walls of the restorations. A total of 16 images were obtained for each sample.
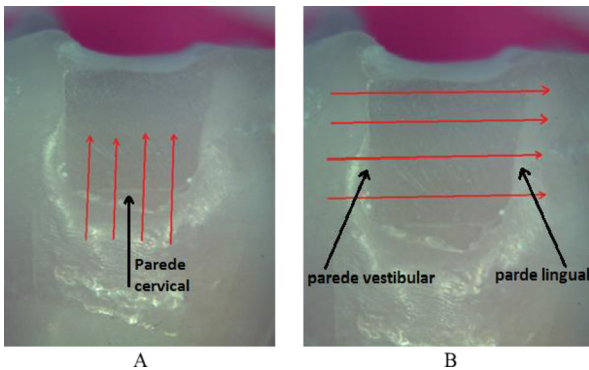


**Fig. 1.** Representation of the scans (red arrows) performed in each restoration. A: cervico-occlusal scans; B: bucco-lingual scans.
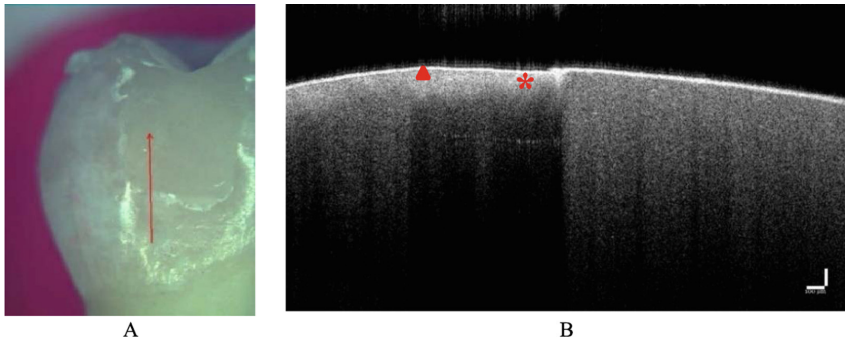
**Fig. 2.** A: Representation of a cervico-occlusal scan (red arrow); B: Cross-sectional OCT image evidencing the enamel surface ($\Delta$) and the resin surface (*). The scale bar refers to the refractive index n = 1.0.

## 2.4  Image Analysis

The qualitative and quantitative analysis of OCT images were achieved using the ImageJ 1.51s (National Institutes of Health, USA) software by a blind, trained, and calibrated examiner. The marginal gap analysis (interfacial adhesive defects) was performed by the measurement of the distance between the enamel wall and the composite resin. The measurements were done at a depth of 500 $\mu$m from the enamel surface, according to the Fig. 3. Each gap thickness measurement was performed 3 times, in order to take into account, the operator's measurement error.

The structural analysis was performed through the qualification and quantification of defects such as bubbles, cracks, agglomerates of resin fillers and defects of marginal adaptation. Finally, still related to the qualitative analyses, the morphological characteristics of the restorations were evaluated, such as the presence of marginal steps, excess material on the edges of the restorations and defects resulting from the finishing and polishing of the restorations.
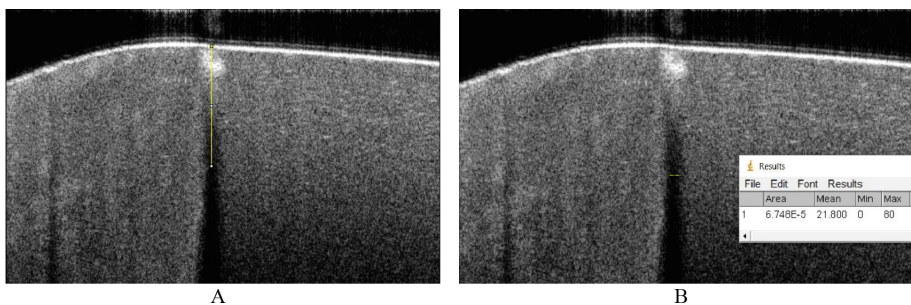


**Fig. 3.** Marginal gap measurements from OCT images (yellow lines). A: Standardization of the 500 $\mu$m depth; B: gap measurement.

## 2.5  Statistical Analysis

The measurements of marginal gaps, on the three faces of each restoration (buccal, lingual, and cervical) were statistically evaluated, separately for each face, in relation to the homogeneity and normality of variances, using Shapiro-Wilk and Levene tests. Thus, the data were analyzed by analysis of variance (ANOVA) and Tukey's post hoc test, at a statistical significance level of 5%, using the Biostat 5.0 software.

# 3  Results and Discussion

## 3.1  Marginal Gap Analysis

Figure 4 shows the difference in backscattering signal of enamel and resin. In Fig. 4(A) it is noticed only scarce bright areas at the interface tooth-restoration, indicating a good seal. In Fig. 4(B) it is evidenced, by the red circle, some bright pixels suggestive of deposition of polishing material on the edge of the restoration. Also, in the area highlighted by the red rectangle, we can see, from the left side towards the right side, a region of higher signal intensity, which corresponds to the tooth (dental enamel), fol-lowed by a region of signal drop, which corresponds to a "void", or absence of material and, on the right side, a light-er region, with higher signal intensity, which corresponds to the interface of the restoration.
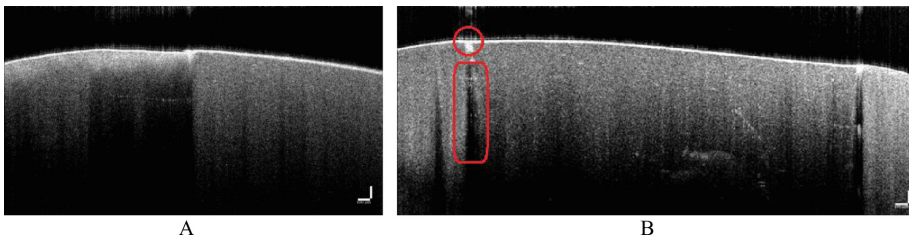


**Fig. 4.**  Representative cross-sectional images obtained by OCT of composite resin restorations. A: absence of gap between the enamel and restoration; B: the red rectangle evidences the presence of a gap.

This aspect corroborates the literature [9] and suggests that there is a gap between the tooth and the restoration, which represents a failure in the restorative procedure and can compromise the longevity of the restoration. The formation of a gap at the tooth-restoration interface is usually related to polymerization shrinkage, in which forces related with the contraction tension can break the connection to the cavity walls and can lead to gaps. This occurrence represents one of the main problems related to the failure of restorations. Gaps can correspond to 1.67 to 5.68% of the total volume of the restoration [10]; also, they can allow filling with oral fluids, which contain bacteria that can cause secondary caries, as well as post-operative hypersensitivity [11].

This aspect was found some restorations, and therefore it was possible to measure the tooth-restoration distance. The results of marginal gap measurements in buccal, lingual,

and cervical walls of restorations are shown in Figs. 5, 6 and 7, respectively. It is noted that the spaces found in restorations made of different resins diverged considerably. Sonic Fill resin presented the lowest tooth-restoration mean distance ($50.3 \pm 9.5$ µm) on both the buccal and lingual walls, while the X-tra base resin presented the highest mean value on the buccal ($74.5 \pm 10.1$ µm) and cervical ($64.8 \pm 18.8$ µm) walls. The Tetric EvoCeram resin presented the higher distance on the lingual wall ($79.6 \pm 9.0$ µm).

In the cervical wall, however, a different behavior was observed, in which the IPS Empress resin presented the lowest average value of the gaps between the tooth and the restoration ($33.9 \pm 15.5$ µm), while the other resins presented similar values among them.
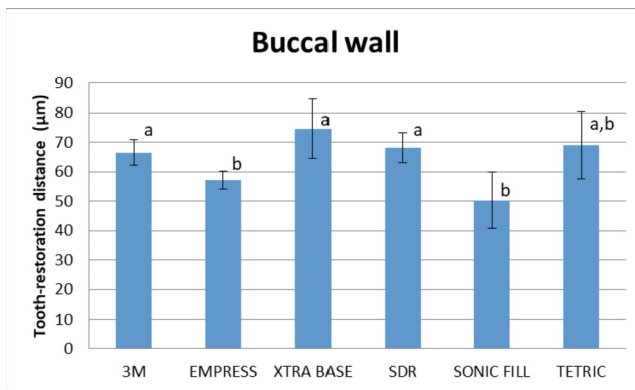


**Fig. 5.** Mean of the distances between the tooth and the buccal wall of the composite resin restorations obtained for the different experimental groups. The bars indicate standard deviation.
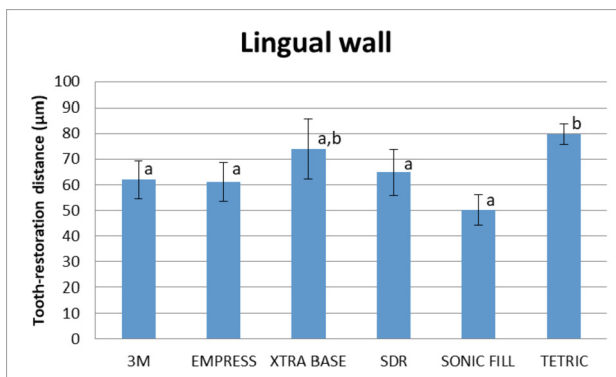


**Fig. 6.** Mean of the distances between the tooth and the lingual wall of the composite resin restorations obtained for the different experimental groups. The bars indicate standard deviation.

OCT has been used as a valuable technique for measuring gaps at the tooth-restoration interface. One of the first works that performed such an evaluation using a time domain
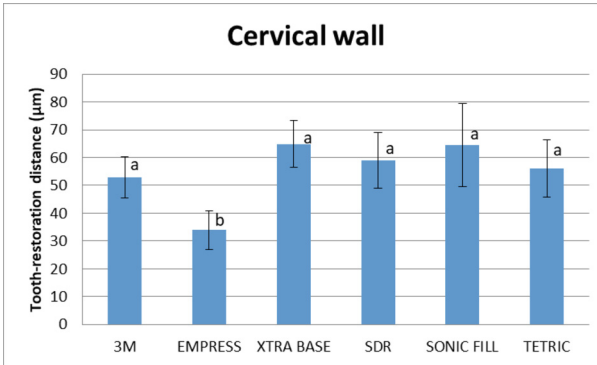
**Fig. 7.** Mean of the distances between the tooth and the cervical wall of the composite resin restorations obtained for the different experimental groups. The bars indicate standard deviation.

optical coherence tomograph (TD-OCT) reported the visualization of gaps as large as 50 µm in size [9]. In this study, we used an SD-OCT (spectral domain) system, which has higher resolution and faster image acquisition speed. For this reason, it was possible to measure gaps of up to 30 µm.

It is worth noting that the gaps observed may be related to the presence of air or the large thickness of the transparent adhesive layer [12]. In any of these hypotheses, the quality and longevity of the restoration may be compromised, and this fact reinforces the need for constant clinical evaluation of these restorations.

In the present study, we found gaps of dimensions between 30 and 80 µm, larger than those reported in the literature [13]. It is known that the formation of these gaps is related to the polymerization contraction which, in turn, can be influenced by different factors, such as the type of resin, the restorative technique, the shape and size of the cavity, among others. In this study, bulk-fill resins, which were developed to simplify the restorative procedure, were evaluated. For that, they have a higher depth of polymerization, which could eliminate the need for layering. It is for this reason that, in the present study, the restorations were performed in a single increment. These are resins that use modified filling technologies that minimize volumetric shrinkage and/or modify viscoelastic behavior, which, according to manufacturers, can result in improved marginal integrity [14]. However, the literature shows that the contraction of these resins is not inferior to conventional composite resins, which justifies the findings of the present study [15].

## 3.2 Structural Analysis

In addition to the gaps identified in the tooth-restoration interface, some imperfections can also be observed inside the resin volume. Dark spots on the OCT image were detected (Fig. 8A), which evidenced that no portion of the incident light was backscattered in that region. For this reason, it is likely that they are the representation of air bubbles formed inside the resin due to the process of accommodation.

Figure 8B shows some irregularities inside the resin, however they appear as clear spots in the OCT image. The bright regions indicate a higher backscattering signal and may represent structures of greater density, such as an accumulation of resin fillers.

Figure 8C presents flaws inside the resin, resulting in air bubbles (indicated by the blue arrow). White spots are also observed (red arrow) which may indicate accumulation of resin filler, since a large portion of light was backscattered in this region. In addition, the interface region has large gaps (yellow arrows), which indicates poor adaptation of the restoration. Figure 8D shows an image of one of the IPS Empress restorations. It is easily noticeable a large black region in the bottom of the resin, where practically no portion of the light is backscattered, but all transmitted. It is possible that the cavity made in the tooth has not been filled by the composite resin. It is worth noting that the thickness recommended by the literature for the layer of the adhesive system used in the present work varies between 4.2 and 6.4 µm (for dentin and enamel, respectively). Thus, it is observed that the adhesive layer shown in Fig. 8D is much thicker than recommended (around 150 µm), which represents a risk for the longevity of the restoration.
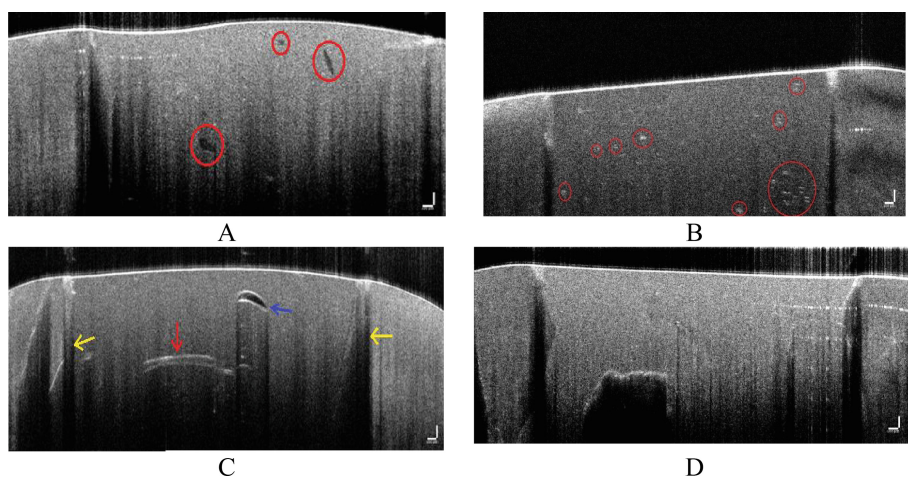


**Fig. 8.** Representative cross-sectional images obtained by OCT of composite resin restorations. A: presence of bubbles; B: presence of filler accumulation; C: presence of higher bubbles and marginal gaps; D: presence of defects of marginal adaptation.

Table 2 shows the result of the structural analysis of the restorations, with the quantification of bubbles and filler accumulations in each experimental group. It is noted that some resins showed significantly more filler accumulation than others. All restorations made with Filtek Bulk Fill and Tetric EvoCeram showed filler buildup. In contrast, restorations made with Sonic Fill, X-tra Base and Surefill SDR resins showed little evidence of filler buildup, in approximately only half of the restorations. Furthermore, even in restorations where these accumulations were present, the amounts presented are considerably smaller when compared to the two resins mentioned above, thus showing a more homogeneous structure of these materials.

Restorations made with Sonic Fill resin also showed better performance considering the air bubble formation. The Sonic Fill resin system consists of a handpiece that allows the sonic adaptation of this resin, and this fact allows better marginal adaptation and more homogeneity of the material. In the present study, none of the 10 restorations

**Table 2.** Quantification of filler and bubbles present in the restorations.

|  | Filtek | IPS Empress | X-tra base | Surefil SDR | Sonic Fill | Tetric |
|---|---|---|---|---|---|---|
| Number of filler accumulations | 85 | 63 | 13 | 7 | 16 | 144 |
| Restorations with filler accumulations | 10 | 10 | 5 | 4 | 6 | 10 |
| Number of bubbles | 2 | 39 | 7 | 2 | 0 | 2 |
| Restorations with bubbles | 1 | 10 | 3 | 2 | 0 | 1 |

made with this resin presented bubbles, showing the ability to spread and fill the entire cavity volume. The other composite resins evaluated are applied in the classic way: their adaptation through the cavity volume is performed manually using a spatula.

An interesting piece of evidence is that the resin that presented the highest number of air bubbles, IPS Empress, is the only nanohybrid composite resin used in the present study. This resin requires its application in increments of a maximum of 2 mm thick, always followed by light curing. Although small portions of the resin are used in each increment, which theoretically would allow the resin to be better distributed during application, this fact was not confirmed in the present study. This result indicates the superiority of bulk-fill resins, both in terms of their better adaptation and the greater agility of the tooth restoration process, since they are performed by a unique increment.

### 3.3 Morphological Analysis

The OCT analysis also enabled the evaluation of the surface morphology of the restorations, as well as the assessment of defects resulting from the finishing and polishing procedure. In Fig. 9, representative images of the main defects found in the evaluated restorations can be seen. The main defects were the excess of material (Fig. 9A), which causes the formation of steps in the margins of the restorations; the formation of surface cracks in the material (Fig. 9B), which can cause aesthetic and structural failures in the restoration; the lack of material at the edges of the cavity (Fig. 9C), and the poor adaptation of the material in relation to the shape of the cavity (Fig. 9D).
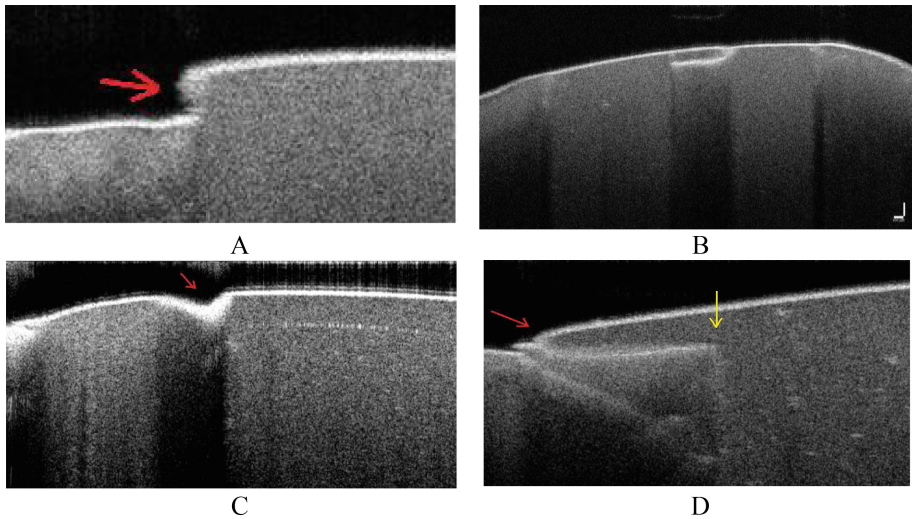
**Fig. 9.** Representative cross-sectional images obtained by OCT of composite resin restorations. A: excess of resin (red arrow); B: presence of a crack on the resin surface; C: presence of a marginal defect due to the polishing procedure; D: presence of excess of resin.

## 4  Conclusion

The optical coherence tomography technique is useful to evaluate bulk-fill composite resin restorations, allowing the quantification of defects, bubbles, steps, large load particles and cracks at the tooth-restoration interface, as well as margin excesses or polishing failures of the restorations. Also, although bulk-fill composite resins have different optical behaviors, they behave similarly to a nanohybrid composite resin in relation to their marginal adaptation.

## References

1. Marschall, S., Sander, B., Mogensen, M., et al.: Optical coherence tomography—current technology and applications in clinical and biomedical research. Anal. Bioanal. Chem. **400**, 2699–2720 (2011)
2. Pan, Y., Lankenau, E., Welzel, J., et al.: Optical coherence-gated imaging of biological tissues. IEEE J Select Top Quant Elec **2**, 1029–1034 (1996)
3. Ali, S., Gilani, S.B.S., Shabbir, J., et al.: Optical coherence tomography's current clinical medical and dental applications: a review. F1000Res **22**, 310 (2021)
4. Hariri, S., Moayed, A.A., Dracopoulos, A., et al.: Limiting factors to the OCT axial resolution for in-vivo imaging of human and rodent retina in the 1060nm wavelength range. Optics Exp **17**, 24304–24316 (2009)

5. Karth, P.A.: Optical Coherence Tomography. American Academy of Ophthalmology. http://eyewiki.aao.org/Optical_Coherence_Tomography

6. Katkar, R.A., Tadinada, S.A., Amaechi, B.T., et al.: Optical coherence tomography. Dent. Clin. North Am. **62**, 421–434 (2018)

7. Hsieh, Y.-S., Ho, Y.-C., Lee, S.-Y., et al.: Dental optical coherence tomography. Sensors (Basel) **13**, 8928–8949 (2013)

8. Lee, S.-H., Lee, J.-J., Chung, H.-J., et al.: Dental optical coherence tomography: new potential diagnostic system for cracked-tooth syndrome. Surg. Radiol. Anat. **38**, 49–54 (2016)

9. Melo, L.S.A., Araujo, R.E., Freitas, A.Z., et al.: Evaluation of enamel dental restoration interface by optical coherence tomography. J. Biomed. Opt. **10**, 064027 (2005)

10. Davidson, C.L., Gee, A.J.: Feilzer a the competition between the composite-dentin bond strength and the polymerization contraction stress. J. Dent. Re. **63**, 1396–1399 (1984)

11. Deliperi, S.: Bardwell DNAn alternative method to reduce polymerization shrinkage in direct posterior composite restorations. J. Am. Dent. Assoc. **133**, 1387–1398 (2002)

12. Ishibashi, K., Ozawa, N., Tagami, J., et al.: Swept-source optical coherence tomography as a new tool to evaluate defects of resin-based composite restorations. J. Dent. **39**, 543–548 (2011)

13. Oskoee, S.S., Bahari, M., Navimipour, E.J., et al.: Factors affecting marginal integrity of class II bulk-fill composite resin restorations. J. Dent. Res. Dent. Clin. Dent. Prospects **11**, 101–109 (2017)

14. Misilli, T., Gonulol, N.: Water sorption and solubility of bulk-fill composites polymerized with a third generation LED LCU. Braz. Oral Res. **31**, e80 (2017)

15. Behery, H., El-Mowafy, O., El-Badrawy, W., et al.: Gingival microleakage of class II bulk-fill composite resin restorations. Dent. Med. Probl. **55**, 383–388 (2018)

# Wearable Influence on Breathing Pattern Measured by Unrestricted Barometric Plethysmography in Wistar Rats

Rodrigo Moreira Felgueira[1,2](✉) , Isabela Carvalho Velloso de Oliveira[3] ,
Carolina Lourenço Marques[3] , Fernando José de Freitas Junior[1] ,
Mariana Correia de Oliveira Alves[3] , Erasmo Assumpção Neto[2] ,
Soraia Katia Pereira Costa[3] , and Henrique Takachi Moriya[1]

[1] Biomedical Engineering Laboratory, University of São Paulo, São Paulo, Brazil
rodrigofelgueira@usp.br
[2] Pediatrics Department, Universidade Nove de Julho, São Bernardo, Brazil
[3] Institute of Biomedical Sciences, University de São Paulo, São Paulo, Brazil

**Abstract.** This study aimed to evaluate and to describe the influence of a commercial chest vest that may have influence on Wistar rodents' respiratory pattern. Such wearables can be used under current conditions in sets of remote monitoring devices in animals with different models of diseases, both in specific terms and in the context of integrated telehealth and the Internet of Things (IOT). We assessed the respiratory frequency of the animals with and without the wearable using unrestricted whole-body plethysmography (WBP) on three consecutive days of analysis. WBP is a non-invasive technique for measuring lung function that uses well-defined concepts of lung mechanics and allows serial and prolonged recording parameters, in an environment where the animal can perform daily activities, such as eating and sleeping. Our results showed that the commercial wearable altered the respiratory pattern of the animals. It seems that is necessary to develop a customized wearable that would reduce such influence, in order to improve future studies using biosensors.

**Keywords:** wearable technology · barometric plethysmography · breathing pattern · Wistar rats

## 1 Introduction

Wearable Technology (WT) is inserted in the context of remote monitoring of biological signals, based on non-invasive and wearable sensors. Wearable devices can monitor and record real-time information about the physiological condition and body movement activities. Health monitoring systems based on wearable sensors can comprise different types of flexible sensors that can be integrated into textile fibers, clothing and elastic bands or directly attached to the human body. The sensors are capable of measuring physiological signals such as electrocardiogram (ECG), electromyogram (EMG), heart rate (HR), body temperature (BT), electrodermal activity (EDA), arterial oxygen saturation ($SAO_2$), blood pressure (BP), and respiratory rate (RR) [1, 2].

The first step before the development of non-invasive sensors is the selection of the wearable that will be integrated into their use. The ideal wearable needs to find certain characteristics that allow its proper use: 1) They must be comfortable and not alter the user's behavior; 2) They cannot impair the capture of the desired biological signals or the data transmission capacity of the sensors. 3) They cannot significantly modify the expression of the biological signal to be measured [1, 2].

The use of WT for monitoring respiratory health is described in the literature. Four main areas of interest for respiratory health are described: pulse oximetry, lung ventilation, activity tracking, and air quality assessment [2].

Lung ventilation is the product of respiratory rate (or number of breaths per minute) and tidal volume (the volume of air that is inhaled and exhaled with each breath). Monitoring lung ventilation means monitoring these two fundamental parameters and all parameters that can be derived since inspiration and expiration are distinct, i.e., inspiratory time, expiratory time, lung cycle (i.e., the ratio between the inspiratory time and the total time of the respiratory cycle), flow inspiratory (that is, the relationship between inspired volume and inspiratory time) and expiratory flow (i.e., the relationship between exhaled volume and expiratory time) [3–5].

These parameters are fundamental to assess an animal's ventilatory function. Respiratory rate, for example, represents a vital sign used to monitor the progression of various diseases. An abnormal respiratory rate is a key prognostic factor and an important marker of severe disease [5–7].

However, such studies need to investigate technical aspects related to the use of wearables in the behavior and breathing pattern of animals. One of the ways to assess this influence would be through the use of non-invasive devices for measuring lung mechanics in experimental animals [3, 6–8].

In this context, unrestricted Whole-Body Plethysmography (WBP) uses a non-invasive pulmonary function measurement technique that assumes the principle of gas dynamics established by the Boyle-Mariotte Law that allows us to use the WBP as a tool for measuring respiratory pattern variables and pulmonary mechanics, without the need to install invasive or contact devices, allowing comfort and longer monitoring [3, 6, 9].

WBP uses a closed system chamber where the animal is positioned in dorsal decubitus, continuously ventilated at constant bias flow. The rodent is placed to be evaluated without restrictions on body movement. According to Fig. 1, there is a main chamber (A) where the animal is kept, and a reference chamber balanced with atmospheric pressure (B). Between both chambers A and B there is a differential pressure transducer (C). The airflow in the main chamber is measured at the top by another sensor called a pneumotachograph (D). At the side end, an aerosol inlet with a closed valve can be seen, which allows the administration of gas. The air outlet is positioned on one of the sides of the chamber (3). In the lower part, there is an inlet orifice for the bias flow (2), which maintains a continuous flow of gas, allowing oxygenation and preventing hypercapnia [3–5].

Since it is possible to measure the variation of pressure and flow in the chamber through the placement of these sensors, it is feasible to correlate such variables with the pattern and lung mechanics of the animals. Applying the concept of gas dynamics, it is possible to acquire pressure variables and correlate them with volume variables [3, 7–9].
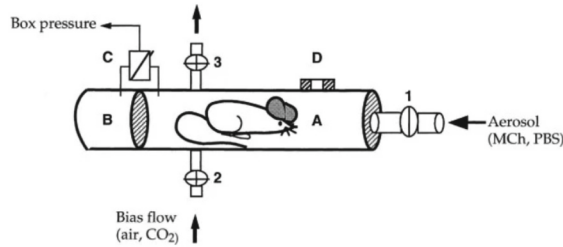
**Fig. 1.** Scheme of Whole-Body Plethysmography [4], description in the text above

The hypothesis suggested for the study is that the use of wearables in experimental rats can induce changes in behavior and respiratory dynamics. In addition to behavior analysis, the proposal is to use WBP to assess the respiratory rate pattern of these animals on three consecutive days of analysis.

## 2   Material Methods

This is a descriptive experimental study with association of hypothesis.

A.  Description of groups

Five male rats of the *Rattus norvegicus* species, Wistar breed, were analyzed. They were young adult animals aged between 6 and 12 weeks (registration number 9956280119 of the Ethics Committee of the Pharmacology Department of the Institute of Biomedical Sciences of the University of São Paulo). The rodents were kept in the rat vivarium under controlled conditions of temperature (22ºC) and lighting (light-dark cycle of 12/12h), with water and food ad libitum.

B.  Equipment description

The equipment used was an unrestricted whole-body plethysmograph (PLY4211, BUXCO, USA). The equipment was calibrated before each experiment. (DSI Buxco Bias Flow Manual). The data recorded for the experiment was the respiratory rate. The normal respiratory rate for animals of this species when awake is 110 breaths per minute (range 100 to 140 per minute) [10].

C.  Experiment description

The animals were acclimatized in an acrylic chamber for at least 10 min before recordings. 5-min time series records were generated for 3 consecutive days (2019–12-04, 2019–12-05 and 2019–12-06) in the mornings. Each animal was submitted to 2 recording batteries per day: the first without wearable and the second with the wearable.

D.  Wearable description

The garment selected was a chest vest whose main material was nylon and a Velcro closure system (Chest vest, size S, Ferplast Jogging, Brazil) (Fig. 2).

E.  Statistical analysis

Given the small sample size of animals, the dataset of the functional experiments was expressed through descriptive statistics. Location analyzes were performed by calculating the mean and quartiles, and dispersion measurements by calculating the range and standard deviation using a commercial software (Graphpad, Prism, USA). Data visualization was performed using histogram and boxplot graphics.



**Fig. 2.** Tested commercial chest vest.

## 3   Results

The first part of the experiment involved placing the rodents without the vest inside the acrylic chamber. The animals were kept in an unrestricted manner, as shown in Fig. 3.



**Fig. 3.** Chamber used in experimentation.

The second part of the experiments consisted of placing the commercial wearable on the rodents, as illustrated in Fig. 4, for later start of the experiment battery inside the acrylic chamber.

In Table 1 there is a description of the weights of the animals in grams (g) on the 1$^{st}$ day of the experiment, in addition to their tolerance to the use of the wearable. Of the animals tested, three had good wearable tolerance.

**Fig. 4.** Wistar rat with tested vest

**Table 1.** Description of animals and experiments

| Animal | Weight (g) | Tolerated the wearable |
| --- | --- | --- |
| 1 | 553 | Yes |
| 2 | 536 | No |
| 3 | 340 | No |
| 4 | 328 | Yes |
| 5 | 330 | Yes |

Figure 5 shows the summary table of the experiments performed on the three animals that tolerated the use of the wearable during the three days (rats 1, 2 and 5). The variable shown in the graphs is the respiratory rate, through the number of respiratory intercussions per minute (ripm).

## 4   Discussion

The proposal for the study was that wearables in experimental rats could induce changes in behavior and respiratory dynamics. For this purpose, a commercial garment was used in animal testing at WBP [7].

After placing the wearable, two of the five animals (Rat 3 and Rat 4) became extremely agitated and were unable to remain with it. As a result, the team responsible for the experiments chose to exclude them from the analysis. This is an important factor to consider when choosing a wearable for use in breath testing. The other 3 animals, after a process of environmentalization with the wearable, showed good tolerance to its use, being possible to proceed with experiments.

As for the results that evaluated the respiratory rate pattern measured by the WBP, two outcomes were evaluated that can translate the influence of the wearable on the behavior and the ventilatory capacity of the animal. First, dispersion measures such as range, standard deviation and quartiles were observed. The increase in the variability of the values reflects a difficulty in adapting the animal to the clothing. On the other hand,

measures of central tendency, such as median, reflecting the evolution of ventilatory capacity, since respiratory rate is one of the components of minute volume (MV).

Rat 1 presented median, quartiles and range with a tendency to stability when they were without a vest. However, when exposed to the wearable, there is a tendency to hypoventilation and an increase in variability, which denotes a possible negative influence of the wearable on comfort and respiratory capacity.

Rat 2 showed a tendency towards stability in the first two days of the experiment, however, on the 3$^{rd}$ day, median and dispersion values were very different, which can be attributed to some failure in the WBP record for the group without vest. When exposed to the wearable, this animal showed a large increase in dispersion and median values, which can be attributed to the discomfort generated by the use of clothing.
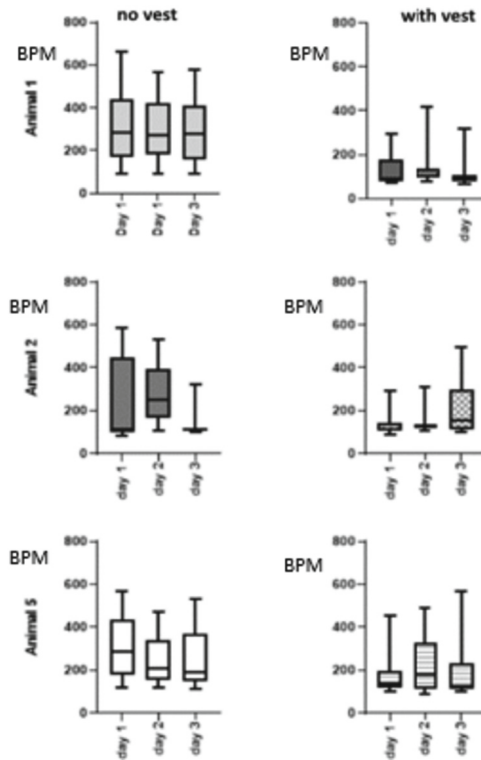


Fig. 5. Summary table of the experiment in rats that tolerated the use of the wearable. BPM = breath per minute

Rat 5, following the same trend as Rat 1, evolved with median stability and dispersion in the three days when he was without vest. On the other hand, in the phase in which he was wearing a vest, his data showed a tendency for an increase in the variability of the dispersion measurements, as well as a large oscillation in the median in the three days, which can also be attributed to the negative influence of the wearable, both in the sense of change your behavior as your breathing capacity.

# 5   Conclusions

The accomplished experiment showed that the option of the selected commercial wearable can influence the behavior and lung mechanics of the animal. In this context, it would be important to go for the development of a customized wearable that is more comfortable and has less influence on the behavior and pulmonary ventilation of rats.

**Conflict of Interest.**   The authors declare that they have no conflict of interest.

# References

1. Majunder, S., Mondal, T., e Deen, M.J.: Wearable sensors for remote health Monitoring. Sensors (Basel) **17**(1), 130 (2017). https://doi.org/10.3390/s17010130
2. Aliverti, A.: Wearable technology: role in respiratory health and disease. Breathe **13**, e27–e36 (2017). https://doi.org/10.1183/20734735.008417
3. Lundblad, L.K.A., Irvin, C., Adler, A., Bates, J.H.T.: A reevaluation of the validity of unrestrained plethysmography in mice. Appl. Physiol. **93**, 1198–1207 (2002). https://doi.org/10.1152/japplphysiol.00080.2002
4. Hamelmann, E., Schwarze, J., Takeda, K., et al.: Noninvasive measurement of airway responsiveness in allergic mice using barometric plethysmography. Am. J. Respir. Crit. Care Med. **156**, 766–775 (1997)
5. Kirschvink, N., Vincke, G., Onclinx, C.: Comparison between pulmonary resistance and Penh in anaesthetised rats with tracheal diameter reduction and after carbachol inhalation. J. Pharmacol. Toxicol. Methods **51**, 123–128 (2005). https://doi.org/10.1016/j.vascn.2004.10.002
6. Hülsmann, S., Khan, A., Hagos, L., Hindermann, M., Nägel, T., Dullin, C.: Evaluation of a mechanical lung model to test small animal whole body plethysmography. Sci. Rep. **11**(1), 1–14 (2021). https://doi.org/10.1038/s41598-021-96355-y
7. Glaab, T., Braun, A.: noninvasive measurement of pulmonary function in experimental mouse models of airway disease. In: Lung (Vol. 199, Issue 3, pp. 255–261). Springer (2021). https://doi.org/10.1007/s00408-021-00443-
8. Pazhoohan, S., Feijani, F. A., Mehrabbeigi, F., Palizvan, M. R., & Hajihashemi, S. Effect of nasal airflow on respiratory pattern variability in rats. Physiol. Pharmacol. (Iran) **26**(1), 70–78 (2022). https://doi.org/10.52547/phypha.26.1.7
9. Deng, J., Chen, X.K., Guo, F.Z., et al.: Respiratory function tolerance of rats with varing degrees of thoracic volume reduction. Orthop. Surg. (2023). https://doi.org/10.1111/OS.13630
10. Schulz, H., Muhle, H.: Chapter 16: Respiration in Part 6: Physiology in the laboratory Rat. Academic press, Cambridge (2000)

# Optimization of Design
# and Manufacturing Parameters
# of One-Plane Bevel Tipped 3D Printed
# Microneedles

Isabella Villota[(✉)] , Paulo-Cesar Calvo , Oscar Campo ,
and Faruk Fonthal

Universidad Autónoma de Occidente, Cali V. 760030, Colombia
isabella.villota@uao.edu.co
http://www.uao.edu.co

**Abstract.** This paper presents the micro needle's fabrication process
with a single-plane beveled needle tip (MNTB1). Sixteen micronee-
dles were measured at four different angles ($-45°,0°,45°,90°$) to deter-
mine that and compared with the theoretical design values. This study
shows the pre and post-printing process, where $0°$ was found to be
the best printing angle, achieving dimensions in the geometry of the
microneedle of $45,346°$ in experimental grade $(\theta)$, $0,448\,\mathrm{mm}$ in length
and $0,248\,\mathrm{mm}$ external diameter$(\alpha)$. On the other hand, excess resin is
properly removed in the post-printing process when the microneedles are
exposed to $10\,\mathrm{min}$ in 99% isopropyl alcohol. Then the microneedles are
subjected to a curing process at $60°\mathrm{C}$ for $30\,\mathrm{min}$. Finally, the micronee-
dles are subjected to a drilling process to enhance the $0,100\,\mathrm{mm}$ inner
diameter. This fabrication process of a microneedle by 3D printing shows
good results for implementation in a transdermal drug delivery device.

**Keywords:** Microneedles · Transdermal drug delivery · 3D printing ·
Computer Aided Design

## 1 Introduction

Transdermal drug delivery (TDD) is an increasingly attractive method because
of its advantages compared to conventional drug delivery methods. This method
allows transporting molecules directly to the dermis in a painless way [1,2],
where microneedles (MN) take a great relevance because they allow transporting
the drug without reaching the nerve endings of the skin [3] due to their small
dimensions.

MNs have lengths from $0,150\,\mathrm{mm}$ to $1,500\,\mathrm{mm}$; these lengths are sufficient
to release the drug into the epidermis. The width of the MNs can be between
$0,050\,\mathrm{mm}$ and $0,250\,\mathrm{mm}$ [4].

Among the different TDD devices, the transdermal patch composed of MN
has been studied by other researchers, showing positive results in the efficient
transdermal application of specific drugs such as insulin [5–8].

Micro-molding techniques are used to fabricate the microneedles found in TDD devices [3]; these techniques require high-cost tooling. For this reason, other methods of MN fabrication have been explored; one of them is with the help of 3D printing technologies, which in recent years have revolutionized different fields, such as pharmaceutical, industrial, biomedical, and other sciences, because they have been able to obtain cost-effective and functional prototypes and products [9].

One of the most widely used techniques in 3D printing is a stereolithography laser (SLA) using UV-sensitive liquid resin; this technique has recently been used for the rapid and cost-effective fabrication of MNs. These use biocompatible materials for insulin delivery anticancer drugs, among others [10–14]. However, 3D-printed MN fabrication presents significant challenges. One is that the print quality can be easily affected due to sensitivity in some manufacturing parameters, such as print angle, pre-cure, and post-cure of the part.

For this reason, it is necessary to continue investigating more about this method to fabricate MN and overcome the challenges. In this work, we show the fabrication process of an MN with a single-plane beveled needle tip (MNTB1) inspired by a single-bevel hypodermic needle that was mechanically evaluated in a previous study [15] seeking to answer *How does the printing angle and curing process influence the final quality of the microneedles in a 3D printing process?*. During the last decades, a range of external and internal diameters found with which an MN can be fabricated to support the insertion process in the skin mechanically. Also, a safety factor of 4 is used in medical instruments from these ranges, a value chosen to fabricate the microneedle in this work. The MN was manufactured with a class 1 biocompatible resin with the SLA method; the effect of the printing angle and the washing and curing conditions after printing were examined, and the first two parameters were optimized, ensuring the printing quality of the MN.

## 2    Materials and Methods

### 2.1    Materials

The material used to fabricate the microneedles was Surgical Guide class 1 biocompatible resin (Formlabs), and 99% isopropyl alcohol was used to wash the printed parts. The material used to manufacture the microneedles was Surgical Guide class 1 biocompatible resin (Formlabs). The resin complies with ISO 10993-1:2018 and is designed for relatively high elongation and ultimate tensile strength. Its flexural strength is $\geq 70$ MPa, and its flexural modulus is $\geq 2000$ MPa. The resin requires post-curing to achieve biocompatibility and optimal mechanical properties.

Also, 99% isopropyl alcohol was used to wash the printed parts, as indicated by the resin manufacturer's specifications.

## 2.2   Computer-Aided Design (CAD)

The MN was designed in Solidworks software. A one-bevel hypodermic needle inspired its geometry, with a length of 0,450 mm and an angle ($\theta$) of 45°. Previously, a mechanical study was carried out to ensure a safety factor of 4 to support the insertion load to the skin. With this premise, the external ($\alpha$) and internal ($\beta$) diameter sizes were 0,232 mm and 0,100 mm, respectively, see Fig. 1.

Also, the microneedle array was designed with a rectan-gular-shaped base of 8,4 mm×4 mm with eight microneedles; this was done only to evaluate the dimensions of the different MN impressions and to have some measurements needed to ensure the values were obtained without wasting material.

## 2.3   3D Printing

The MN designs were exported from Solidworks to Pre-Form, where supports were added to the designs with a contact point of 0,350 mm and a layer thickness of 0,050 mm set.

On the SLA laser technology Form2 printers (Formlabs, GoPrint3D, Ripon, North Yorkshire, UK); the MNs manufactured. Four printing angles (-45°, 0°, 45°, and 90°) were set in the MN manufacturing process to evaluate the effect of this angle on the print quality of the MN.
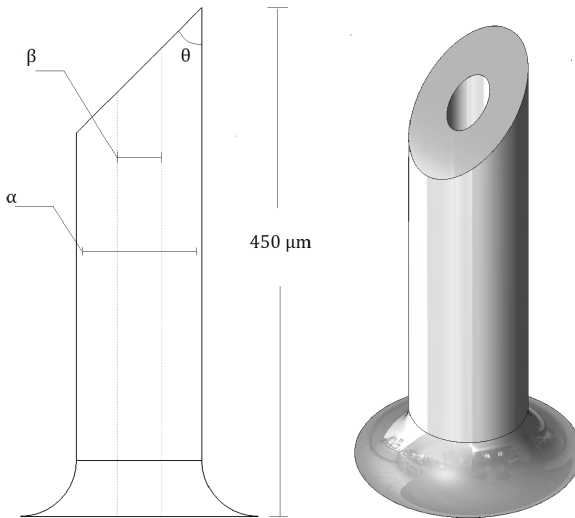


**Fig. 1.** Schematic illustration of the MNTB1 design.

## 2.4   Post-impression Process

Once the MN were printed, they were immersed in 99% isopropyl alcohol and placed in a Cole-Parmer 8890 ultrasonic cleaner operating at a frequency of 47khz at different washing times (2 min, 4 min, 6 min, 8 min, 10 min) to observe which was the appropriate time to clean these MN.

   Once the MNs were washed and dried, they were subjected to the curing process in FormCure. In this process, the parts were exposed to a temperature of 60°C for 30 min, given by the manufacturer. These parameters are necessary to achieve the maximum possible strength and stability, and the curing process guarantees the incompatibility of the part. Following the curing of the MN, an inner diameter drilling procedure was performed on a router CNC.

## 2.5   Measuring Macroscope

The measurement of the dimensions of the 3D-printed MNs in each test was done with a VR-3000 Measuring Macroscope. To perform the measurements, it was necessary to make a magnification of 40X and 80X.

## 2.6   Statistical Methodology

The experiment was based on varying the printing angle to measure the response variables (bevel angle, height, and external diameter).

   In this case, the completely randomized experimental design was selected because the application of the effect of the angle position was performed on homogeneous experimental units (the microneedle with the same resin and geometry). Minitab software was used for this procedure.

# 3   Results

## 3.1   Optimization of Printing Angle

The microneedles were printed at four angles ($-45°$, $0°$, $45°$, $90°$). At the angle of -45°, see Fig. 2b; it was observed that the geometry of the MN took an inclination, see Fig. 2f and its structure was different from the designed model of the microneedle, see Fig. 2a. When the angle was 45°, see Fig. 2c, the form of the MN was similar to a triangle, i.e., it did not achieve a close resemblance to the designed model of the needle, see Fig. 2g. When the angle was 90°, see Fig. 2d; the structure appeared far from the theoretical model, as the MN took on a flattened shape, see Fig. 2h. In comparison, with the printing angle of 0°, see Fig. 2e, the appearance of the MN was very similar to that of the theoretical model, see Fig. 2i.

   In addition to a review of the appearance of the printed structures, the height, angle, and external diameter were measured at 16 MN for each printing test angle; this number of microneedles corresponds to the sample size taken in the experiment. This sample size allowed us to obtain measurements with a reliability

of 95%, a sampling error of 3, and a test power of 83%, showing that the number of samples was adequate.

Furthermore, an analysis of variance found that adequate control was exercised in the experiment and that the effect of the angle position could explain the variation in the measurements.
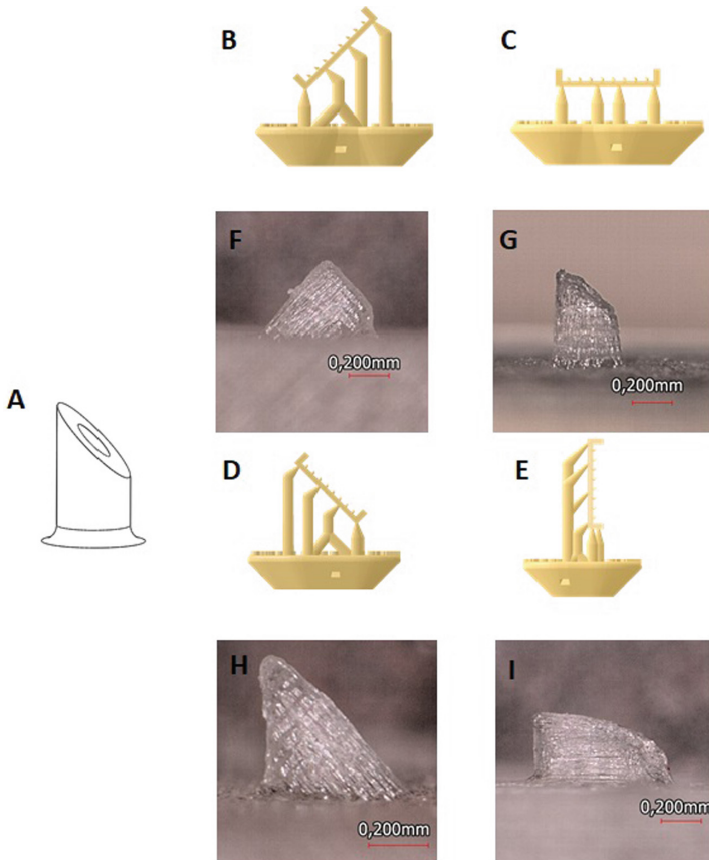


**Fig. 2.** MN's fabricated at different printing angles, a) CAD model, b) Printing base at $-45°$, c) Printing base at $45°$, d) Printing base at $90°$, e) Printing base at $0°$, f) Printing at $-45°$, g). Printing at $45°$, h). Printing at $90°$, i). Printing at $0°$.

The mean of the experimental measurements was the value used to compare the dimensions obtained experimentally with those of the designed model. Where it was observed that the angle of $0°$ allowed the best printing resolutions and also the closest measurements to the CAD model, obtaining in this printing angle the following experimental measurements: an angle of $45,346°$, a length of $0,448$ mm and an external diameter of $0,248$ mm, as shown in Table 1, where it can also

be observed that the angles of $-45°$ and $90°$ are the least appropriate for the microneedle design proposed in this work since these angles generate values of geometric dimensions farther away from the CAD model.

**Table 1.** Experimental Values of Outer Diameter, Bevel Angle, and Length of the microneedle vs printing angle.

| Printing angle (°) | Outer diameter (mm) | Bevel angle (°) | length (mm) |
|:---:|:---:|:---:|:---:|
| −45 | 0,449 ± 0,020 | 82,141 ± 2,306 | 0,420 ± 0,016 |
| 0 | 0,248 ± 0,007 | 45,346 ± 0,583 | 0,448 ± 0,006 |
| 45 | 0,413 ± 0,013 | 41,974 ± 1,629 | 0,430 ± 0,013 |
| 90 | 0,726 ± 0,021 | 74,958 ± 2,745 | 0,373 ± 0,008 |

## 3.2 Post-printing Process

Before subjecting the MNs to the curing process, the manufactured MNs should be washed to remove the remaining resin pieces, see Fig. 3a. For this purpose, the MNs were soaked in 99% alcohol at different time intervals (2 min, 4 min, 6 min, 8 min, 10 min) to observe which time was suitable for this process to optimize it. For this, the process was repeated three times at different times, where it was obtained that a 2 min wash exposure was not enough since the MN still had excess resin, see Fig. 3b.

When the MN was exposed to 4 min of washing, it was observed that the MN was much cleaner than in the previous case, but excess resin could still be observed, see Fig. 3c; the same happened when the MN was subjected to alcohol for 6 min, see Fig. 3d, since ex-cess resin was still observed, especially in all the inclination. When the MN was exposed to 8 min in alcohol, the piece could be seen as almost clean, see Fig. 3e; however, small resin residues were appreciated that did not appear when the MN was exposed to 10 min in alcohol, see Fig. 3f.

The three times the cleaning process was carried out, the same occurred in each washing time, which allowed us to know that 10 min is the best exposure time for the washing process.

## 3.3 Inner Diameter

To obtain the internal angle ($\beta$) of 0,100 mm, it was necessary to make a drilling process after curing the MNs because the printer does not give a good resolution at such small internal dimensions, i.e., it leaves traces of the inner diameter but does not reach to perform correctly from side to side. Therefore, a 0,100 mm microdrill CNC is used to remark and enhance the inner diameter, as seen in Fig. 4.
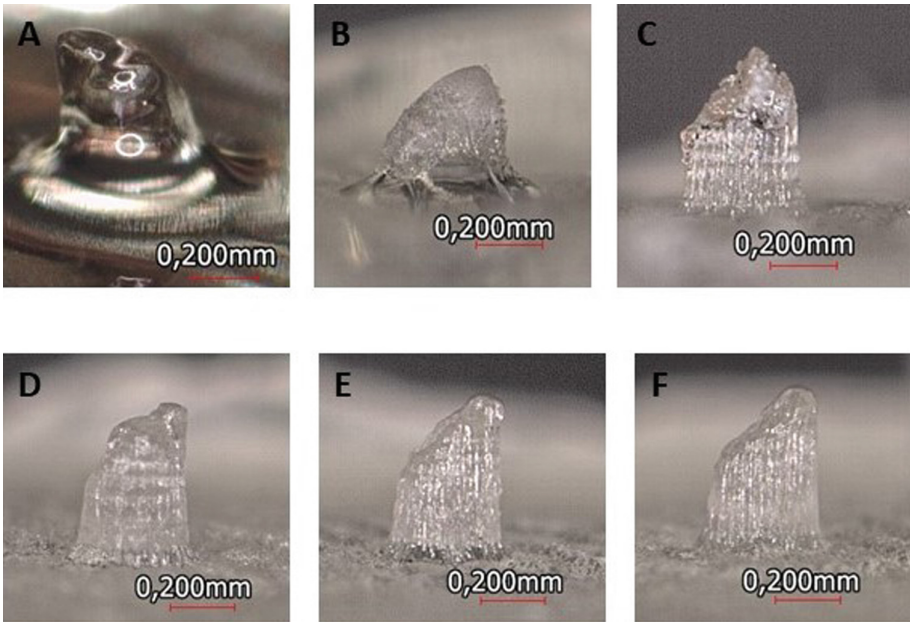
**Fig. 3.** Post-printing process; remove the remaining resin pieces (a), washed in alcohol 2 min (b), 4 min (c), 6 min (d), 8 min (e) and 10 min (f).
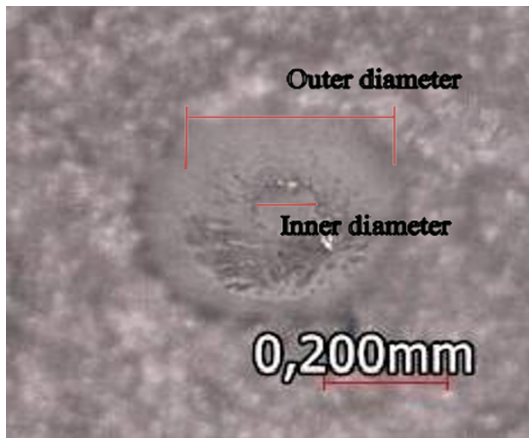


**Fig. 4.** Microneedle internal diameter drilling - Top view

# 4  Discussion

SLA 3D printing is a method that achieves high-resolution prints without the need for expensive equipment and high manufacture times, such as micro-molding, electroplating, and in-mold etching. This has made this method attractive for research in the field of transdermal drug delivery, and especially research focused on microneedles manufacture that is part of TDD systems. The manufacture of 3D microneedles using SLA-type printers has been performed in several studies [11, 12, 16–18]. However, one of the challenges in the previous studies was to achieve good accuracy in microneedle tip printing and microchannel formation as in the present study. The high resolution of the microneedle tip is of great importance because it is essential in the skin penetration process.

In the MN printing process, the printing details can be as small as the diameter of the laser focal spot, which makes it possible to produce small-sized parts. However, when working with micrometer dimensions even smaller than the laser focal spot, it is very likely to generate differences in the dimensions and shapes of the printed model.

In the present study, we printed a microneedle inspired by a single-bevel hypodermic needle of 0,450 mm in height. The best printing quality of this MN was obtained at an angle of 0° concerning the printing platform, unlike some studies such as Economidou et al. [3]. In that study, they manufactured three microneedle shapes with a length of 1,000 mm using class 1 biocompatible resin. Also, they changed the printing angles (0°, 90°, and 45°) for each of the shapes. The authors established that 45° was the angle at which the MN had the best sharpness because at this angle enough amount of support structure was printed before starting to manufacture the MNs. Also, in the Yeung et al. study, they manufacture an MN with a length of 1,000 mm in 3D printing with LT Clear Class IIa biocompatible dental resin [19]. In the above study, they also changed the printing angles (−45°,0°,45°,90°). The authors concluded that the 45° printing angle allowed the best printing quality due to resin flow. However, these two presented studies do not discuss how this printing angle would influence MN with lengths much shorter than 1 mm and with another type of geometry.

On the other hand, Xenikakis et al. manufactured MNs of 1 mm in length and different geometries with Class IIa bio-compatible resin (NextDent Ortho Rigid) [20]. In that study, they examined different printing angles (0°, 45°, -45°, 90°). The authors found that printing angles of 0°, 45°, and 90° cause resin flow into the microchannel section. For this reason, they opted for a printing angle of −52, 63° so that the inclined surface of the hollow microneedles would be parallel to the printing platform. Considering the results obtained in this study, it was evident that inner hole printing is a challenge even in 3D printing and even more so when they are dimensions such as those proposed in this study.

Finally, another study by Choo et al. manufactured 3D printed microneedles with Clear resin, where they reported that at a printing angle of 0°, the MN prints similar to the input design. However, at an angle of 45°, they obtain a sharper microneedle tip [17]. In addition, they varied the microneedle height from 0,600 mm to 2,000 mm, finding that as the input height increased in the input

design, the accuracy of the MN print increased. That study provides evidence that the smaller the printing dimensions, the more challenging the print quality will be. On the other hand, they show that although the 0° angle was not the one that printed the sharpest tip, the print obtained at this angle was very close to the model designed in CAD.

Although previous studies were able to manufacture 3D-printed microneedles with a height $\geq 1,000\,\text{mm}$, these do not show how sensitive the printing is when the lengths are much less than 1,000 mm. Therefore, in this study, we manufacture a 0,450 mm MN and show that for this length, the printing angle that achieves the best print resolution is 0°.

## 5    Conclusions

In the present research work, important information is presented for the fabrication process of an MN by 3D printing, analyzing different effects in the pre and post-printing pro-cesses, finding the right printing angle to allow the closest fabrication of the MN to that of the CAD model, the right time to remove the excess resin to the MN. A drilling process was also performed to polish the inner diameter of the MN.

It was observed that an impression angle of 0° provides the best resolution in the fabrication of the MN, obtaining a very similar appearance and dimensions of the geometry very close to those of the CAD model, being then the angle chosen as the most suitable for the application presented in this work.

Also, it was found that when the MN is subjected to isopropyl alcohol in the cleaning process for 10 min, it is possible to adequately remove the resin residues that remain in the MN just after printing, thus avoiding the MN being left with inadequate dimensions, that is, that are close to those of the CAD model which can affect at the time of being used in a TDD device.

On the other hand, it was found that having dimensions in orders as small as micrometers, the printer cannot make the inner diameters correctly. Therefore, it was necessary to make the inner diameter of the MN by drilling on a CNC with a 0,100 mm drill bit. This drawback obtained with the resolution of printing at such small dimensions is one of the challenges faced by 3D printing technology, which is worth further exploring and optimizing.

The good result obtained in manufacturing an MN that is mechanically suitable for the skin penetration process shows that this study could be relevant to the MN manufacturing processes to be used in a transdermal drug delivery method.

Also, the study shows this fabrication system as a promising option when using new technologies such as 3D printing that allows better and more efficient management of resources than conventional methods, contributing more and more to the applications and challenges that 3D printing brings.

Finally, in future work, the option of exploring the behavior of different printing parameters with different materials and different MN sizes (mechanically validated) to obtain a more general knowledge of the behavior of microneedle

fabrication in SLA 3D printing is proposed. The option of validating the functionality of these microneedles experimentally in a skin insertion process is also considered.

**Conflict of Interest** The authors declare that they have no conflict of interest

# References

1. Han, T., Bhusan Das, D.: Potential of combined ultrasound and microneedles for enhanced transdermal drug permeation: a review. **89**, 312–328. ISSN 09396411, https://doi.org/10.1016/j.ejpb.2014.12.020, https://linkinghub.elsevier.com/retrieve/pii/S0939641114003750

2. Rabiei, M., Kashanian, S., Samavati, S.S., Jamasb, S., McInnes, S.J.P.: Nanomaterial and advanced technologies in transdermal drug delivery. **28**(4), 356–367. ISSN 1061-186X, https://doi.org/10.1080/1061186X.2019.1693579, https://www.tandfonline.com/doi/full/10.1080/1061186X.2019.1693579

3. Economidou, S., Pere, C.P., Okereke, M., Douroumis, D.: Optimisation of design and manufacturing parameters of 3D printed solid microneedles for improved strength, sharpness, and drug delivery. **12**(2), 117. ISSN 2072-666X, https://doi.org/10.3390/mi12020117, https://www.mdpi.com/2072-666X/12/2/117

4. Waghule, T., et al.: Microneedles: a smart approach and increasing potential for transdermal drug delivery system. **109**, 1249–1258. ISSN 07533322, https://doi.org/10.1016/j.biopha.2018.10.078, https://linkinghub.elsevier.com/retrieve/pii/S0753332218348091

5. Ita, K.: Dissolving microneedles for transdermal drug delivery: advances and challenges. **93**, 1116–1127. ISSN 07533322, https://doi.org/10.1016/j.biopha.2017.07.019, https://linkinghub.elsevier.com/retrieve/pii/S0753332217324149

6. Moussi, K., Bukhamsin, A., Hidalgo, T., Kosel, J.: Biocompatible 3D printed microneedles for transdermal, intradermal, and percutaneous applications. **22**(2), 1901358. ISSN 1438-1656, 1527-2648, https://doi.org/10.1002/adem.201901358, https://onlinelibrary.wiley.com/doi/10.1002/adem.201901358

7. Xenikakis, I., et al.: Fabrication and finite element analysis of stereolithographic 3D printed microneedles for transdermal delivery of model dyes across human skin in vitro. **137**, 104976. ISSN 09280987, https://doi.org/10.1016/j.ejps.2019.104976, https://linkinghub.elsevier.com/retrieve/pii/S0928098719302398

8. Johnson, A.R., et al.: Single-step fabrication of computationally designed microneedles by continuous liquid interface production. **11**(9), e0162518. ISSN 1932-6203, https://doi.org/10.1371/journal.pone.0162518, https://dx.plos.org/10.1371/journal.pone.0162518

9. Pedde, R.D., et al.: Emerging biofabrication strategies for engineering complex tissue constructs. **29**(19), 1606061. ISSN 09359648, https://doi.org/10.1002/adma.201606061. https://onlinelibrary.wiley.com/doi/10.1002/adma.201606061

10. Pere, C.P.P., et al.: 3D printed microneedles for insulin skin delivery. **544**(2), 425–432. ISSN 03785173, https://doi.org/10.1016/j.ijpharm.2018.03.031, https://linkinghub.elsevier.com/retrieve/pii/S0378517318301789

11. Economidou, S.N., et al.: 3D printed microneedle patches using stereolithography (SLA) for intradermal insulin delivery. **102**, 743–755. ISSN 09284931, https://doi.org/10.1016/j.msec.2019.04.063, https://linkinghub.elsevier.com/retrieve/pii/S0928493118330704

12. Uddin, M.J., et al.: 3D printed microneedles for anticancer therapy of skin tumours. **107**, 110248. ISSN 09284931, https://doi.org/10.1016/j.msec.2019.110248, https://linkinghub.elsevier.com/retrieve/pii/S0928493119305661

13. Dabbagh, S.R., Sarabi, M.R., Rahbarghazi, R., Sokullu, E., Yetisen, A.K., Tasoglu, S.: 3D-printed microneedles in biomedical applications. **24**(1), 102012. ISSN 25890042, https://doi.org/10.1016/j.isci.2020.102012. https://linkinghub.elsevier.com/retrieve/pii/S2589004220312098

14. Yao, W., et al.: 3D printed multi-functional hydrogel microneedles based on high-precision digital light processing. **11**(1), 17. ISSN 2072-666X, https://doi.org/10.3390/mi11010017. https://www.mdpi.com/2072-666X/11/1/17

15. Villota, I., Calvo, P.C., Campo, O.I., Fonthal, F.: Microneedles: one-plane bevel-tipped fabrication by 3D-printing processes. **27**(19), 6634. ISSN 1420-3049, https://doi.org/10.3390/molecules27196634. https://www.mdpi.com/1420-3049/27/19/6634

16. Krieger, K.J., Bertollo, N., Dangol, M., Sheridan, J.T., Lowery, M.M., O'Cearbhaill, E.D.: Simple and customizable method for fabrication of high-aspect ratio microneedle molds using low-cost 3D printing. **5**(1), 42. ISSN 2055-7434, https://doi.org/10.1038/s41378-019-0088-8. https://www.nature.com/articles/s41378-019-0088-8

17. Choo, S., Jin, S., Jung, J.: Fabricating high resolution and high dimensional microneedle mold through the resolution improvement of stereolithography 3D printing. **14**(4), 766. ISSN 1999-4923, https://doi.org/10.3390/pharmaceutics14040766, https://www.mdpi.com/1999-4923/14/4/766

18. Mathew, E., Pitzanti, G., dos Santos, A.L.G., Lamprou, D.A.: Optimization of printing parameters for digital light processing 3D printing of hollow microneedle arrays. **13**(11), 1837. ISSN 1999-4923, https://doi.org/10.3390/pharmaceutics13111837. https://www.mdpi.com/1999-4923/13/11/1837

19. Yeung, C., et al.: A 3D-printed microfluidic-enabled hollow microneedle architecture for transdermal drug delivery. **13**(6), 064125. ISSN 1932-1058. https://doi.org/10.1063/1.5127778. http://aip.scitation.org/doi/10.1063/1.5127778

20. Xenikakis, I., et al.: Fabrication of hollow microneedles using liquid crystal display (LCD) vat polymerization 3D printing technology for transdermal macromolecular delivery. **597**, 120303. ISSN 03785173, https://doi.org/10.1016/j.ijpharm.2021.120303. https://linkinghub.elsevier.com/retrieve/pii/S0378517321001071

# Ultrasound Coupled Radial Vibration Mode: Influence on Cardiac Sonothrombolysis

Wilton Ruas Silva[✉] [iD] and Sergio S. Furuie [iD]

Universidade de São Paulo, Escola Politécnica, São Paulo, SP 05508-010, Brazil
wilton.ruas@usp.br

**Abstract.** Sonothrombolysis is a technique that uses ultrasound waves with microbubbles in the bloodstream to reduce thrombi in the coronary arteries of patients diagnosed with infarction. The energy generated by the cavitation of microbubbles in the thrombus region is the main cause of their rupture. To generate cavitation, low frequency ($<1$ MHz) and high acoustic pressure ($>100$ kPa) waves, and, thereafter, specific transducers, are necessary; however, their geometry causes strong radial vibration modes, which can reduce their useful transmitted energy. This study aimed to measure the influence of the radial mode on the effectiveness of a rod ceramic transducer designed for a sonothrombolysis device. Five units of a rod Navy II ceramic were used to produce 1 element transducer and a four-element cell, to measure their sensitivities, bandwidth, and acoustic pressure levels using the pulse-echo method, a hydrophone receiver and computing treatment to raise their frequency spectra and transmitted acoustic fields. Once the diameter-to-thickness ratio of the ceramic used was only 1.59, the radial vibrational mode showed to be dominant for a pulse excitation burst, providing, in its resonant frequency, an amplitude response 6 dB higher $(1{:}0,480)$ than its thickness mode. Due to the geometry of ceramic and to the Poisson effect, the radial mode also produced a coupled thickness mode in its radial resonant frequency. The results showed that the typical radial vibration modes, existing in some low-frequency piezoelectric rod ceramics can generate a useful coupled thickness mode, producing the necessary acoustic pressure to be used in a sonothrombolysis equipment.

**Keywords:** Sonothrombolysis · cavitation · rod piezoelectric transducer · diameter-to-thickness ratio · coupled vibration mode

## 1 Introduction

Based on data released in 2018 by the World Health Organization, heart disease is still the leading cause of death in the world [1]. One of the main reasons for this high mortality rate is the lack of a fast intervention, especially in cases of myocardial infarction diagnosis, in which prolonged lack of blood flow, caused by clots (thrombi), can cause irremediable heart muscle loss. Considering the high loss of lives, new techniques are being tested aiming to reduce the fatality rate of these cases. Sonothrombolysis is a medical technique that uses ultrasound waves, together with the application of microbubbles in the bloodstream and the use of fibrinolytic drugs, to reduce thrombi in the coronary

arteries, in patients diagnosed with infarction, promoting an increase in the patency of these arteries. The energy generated by the stable and inertial cavitation of the microbubbles in the region of the thrombi, caused by the mechanical waves of the ultrasound, is the main cause of their rupture.

The use of specific transducers, with frequencies lower than 1 MHz, which provide a lower attenuation and a greater penetration of the mechanic waves in the heart (keeping the Mechanical Index (MI), within the safety ranges indicated for the use of ultrasound in diagnostic mode (MI < 1.9), but with a sufficient strength to cause inertial cavitation [2]), is believed to lead to better results of sonothrombolysis, compared to the results from using commercial diagnostic ultrasound equipment so far.

For sonothrombolysis application, using a ceramic disk for building the transducers to achieve the desired safe energy for cavitation, once a low frequency (<1 MHz) is also desired [3], and considering that the ceramic is a half wavelength resonator, there might be a compromise in the transducers disk diameter-to-thickness (*d/t*) ratio design choice to guarantee a reasonable small-sized transducer and the required low frequency and its related large thickness.

Generally, it is assumed that the ceramic will present only a piston-type surface vibration, therefore with a dominant single thickness extensional mode. Although, there is realistically a large spectrum of other vibration modes and such assumption is considered valid for *d/t* > 20 [4].

Considering the tradeoffs of our application, a rod 10.15 mm-diameter and 6.4 mm-thickness Navy II ceramic (*d/t* ratio equals to 1.59), with an expected 235 kHz thickness mode resonance frequency, was chosen. Measurements showed [5] that when the *d/t* ratio is higher than 0.7 and lower than 2.5, as in our case, it is not valid to assume that there will be only one vibration mode.

Sections 2 and 3 show the evaluation and measurement of the existent dominant vibration modes, using a *d/t* ratio equal to 1.59, and their influence, particularly of the radial mode, on using a ceramic with such *d/t* ratio to generate the necessary acoustic pressure [3], for a transducer array specifically aimed at a sonothrombolysis portable equipment.

## 2   Material and Methods

### 2.1   Material

Five units of a rod Navy II ceramic model 805, from APC, Ltd (www.americanpiezo. com), with a 10.15 mm diameter and a 6.4 mm thickness, were used to produce 1 element transducer and 1 cell, as shown in Fig. 1, designed as a linear array of 4 elements (to evaluate possible radial cross-talking effect) to measure their relative pulse-echo sensitivity, bandwidth, and acoustic pressure levels. A 2.9 mm matching layer, built with resin and alumina, in a proportion based on the equations of Sayers&Tait [6], was built with an acoustic impedance of 4.3Mrayl. Once our transducer application was not related to image, a 3 mm air backing was also added aimed to maximize the transmitted acoustic energy.

**Fig. 1.** Four-transducer linear array cell.

## 2.2 Methods

For most applications, the diameter-to-thickness ratio (*d/t*) is generally a consequence of the piezoelectric ceramic disc operation frequency and the desired acoustic beam profile. Many experimental and conceptual studies have shown the existence of different vibration modes in piezoelectric disks and its relationship with the *d/t* ratio. These studies were usually conducted by evaluating the transmitted signal frequency spectrum [7].

For *d/t* ratio smaller than 6.6, coupled edge and radial modes, near to the thickness extensional one, are expected [8]. As the diameter decreases, the radial mode resonance frequency increases, getting closer to the thickness center frequency of the transducer [2].

In this study, both pulse-echo and hydrophone receiver methods were used for evaluating the general acoustic characteristics of our transducer, based on the E1065/20 standard [9]. For comparison purposes and cross-talking checking, some of the experiments were executed using a single-element transducer and a four-element-transducer cell.

## 3 Results

### 3.1 Experimental Setup

A water tank (550 mm X 300 mm X 150 mm) with a 3D servomechanism control system, a NH4000 Precision Acoustics hydrophone, an Olympus Panametrics ultrasonic square wave pulser/receiver unit, model 5077PR and a Keysight InfiniiVision MS0X2004A oscilloscope were used to execute the experiments.

Figure 2 shows the water tank in the hydrophone measurement. During the experiments, the water was deionized and in ambient temperature ($T_{water}$ = 20 °C).

### 3.2 Experimental Results

Firstly, a pulse-echo measurement was made on a single element transducer to evaluate its thickness response for a negative 200 V shock excitation pulse. The transducer was
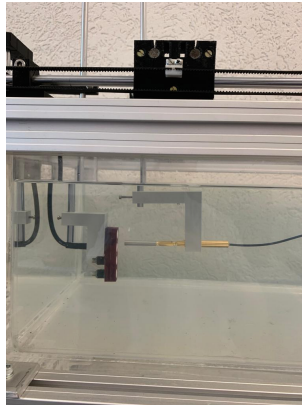
**Fig. 2.** Hydrophone receiver measurement

placed, along its thickness axis, 10 cm away from a 5 cm-thick reflector. The reflected pulse is presented in Fig. 3.
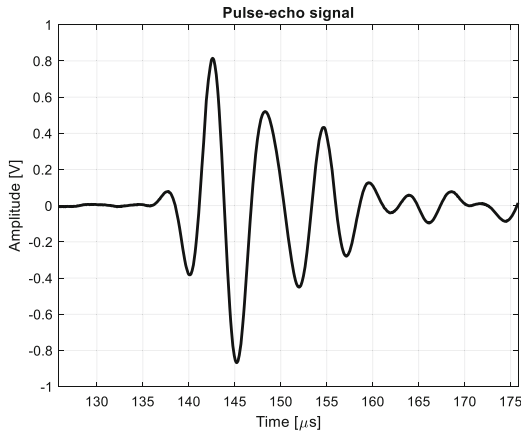


**Fig. 3.** Pulse-echo response:time signal

In Fig. 4, the signal frequency spectrum is presented, showing that a dominant vibration mode appeared in a peak at 170 kHz frequency and, what would be the native thickness vibration mode, occurred around the expected frequency of 240 kHz. This supposed native thickness mode has an amplitude peak response about 6 dB lower than the dominant mode at 170 kHz. Figure 5 presents some details of the dominant vibration modes.

Figure 6 shows the detected uncoupled native thickness vibration mode details, normalized by its peak value, in dB. The figure shows a peak frequency ($f_p = 236.2$ kHz) and a center frequency ($f_c = 240.6$ kHz), for a 6 dB bandwidth, near to the expected length (thickness) mode resonance frequency given by the ceramic manufacturer ($f_r =$
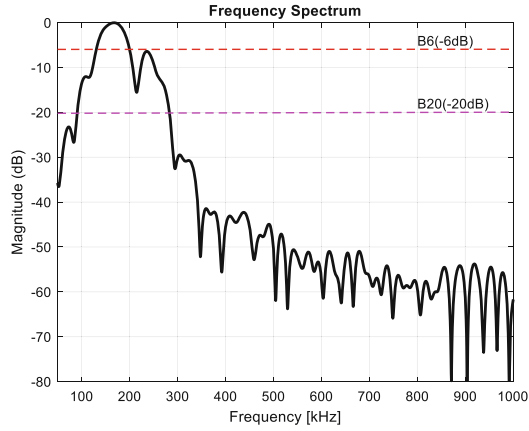
**Fig. 4.** Pulse-echo response: frequency spectrum
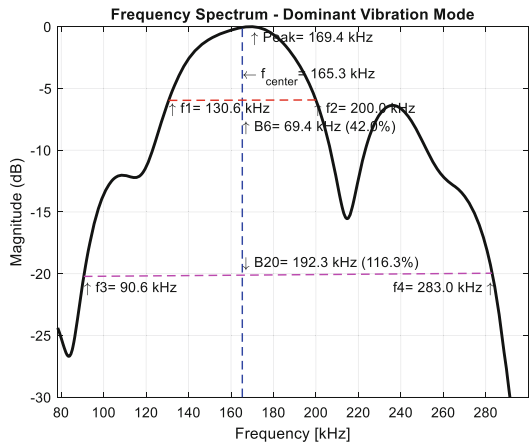


**Fig. 5.** Dominant vibration modes

234.38 kHz), considering, due to its low diameter-to-thickness ratio, this ceramic much more a cylinder than properly a disc.

The ceramic electric impedance and phase curves (Fig. 7) were also used to seek other vibration mode frequencies to evaluate the origin of the unexpected additional extensional mode in a frequency around 170 kHz.

The curves in the figure above show the main vibrations modes for a rod ceramic with their first harmonics, and it is also noticeable a radial resonance frequency near to 170 kHz and a thickness resonance frequency near to 240 kHz.

To evaluate the theoretical general form of the pressure amplitude field generated by our transducer in a given position, we used the expected pressure amplitude in a point $A$ in a distance $r$ from the center of the transducer, with an angle $\theta$ in relation to the transducer longitudinal axis $z$.
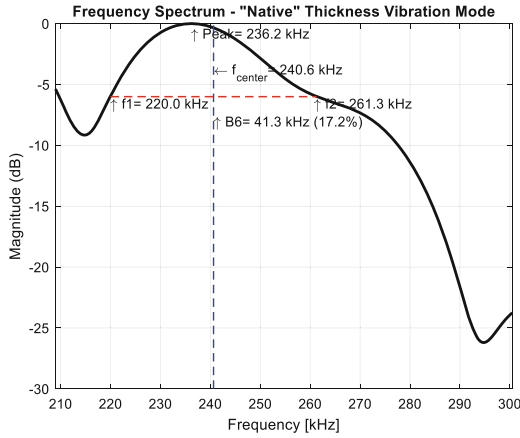
**Fig. 6.** Native thickness vibration mode
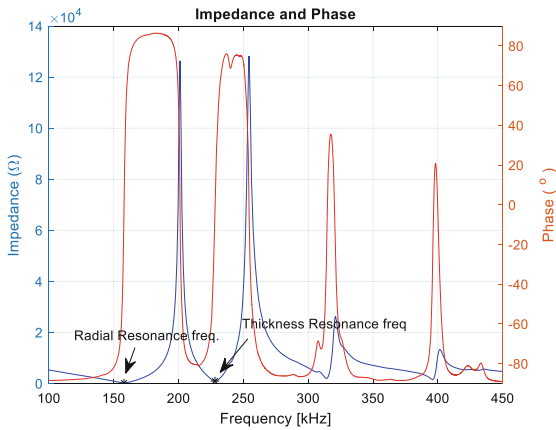


**Fig. 7.** Transducer ceramic electric impedance

For a given point, the normalized pressure amplitude $P(\theta)$ will be given by the equation below [10]:

$$P(\theta) = |\frac{2.J_1(k.a.sen\theta)}{k.a.sen\theta}| \tag{1}$$

where $J_1$ is the first order Bessel function, $k$ is the wavenumber (spatial frequency of the wave, given by $k = 2\pi/\lambda$, where $\lambda$ is the wavelength) and $a$ is the radius of the ceramic disc. Figure 8 shows the expected normalized radial pressure acoustic field profile, in dB, for vibration modes using central frequencies of $f_{rad} = 170$ kHz and $f_{thick} = 240$ kHz, to be compared with the experimental acoustic field, obtained by pulser excitation and hydrophone measurements.

Considering the theorical acoustic pressure profile, and due to the thickness mode vibration, a main longitudinal axial lobe and two secondary sided ones would be expected. As seen in Fig. 8, the dominant radial vibration would generate just one lobe.
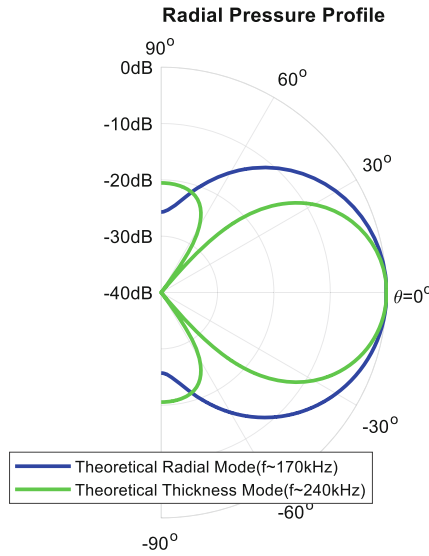


**Fig. 8.** Theoretical acoustic pressure profile
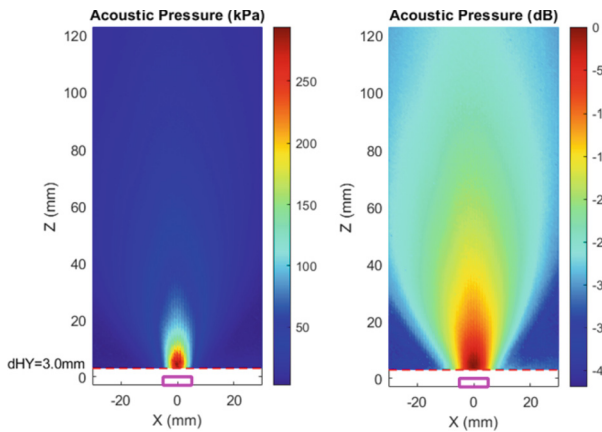


**Fig. 9.** XZ Plan Measured Pressure Acoustic Field

Using a hydrophone, the acoustic pressure field in the longitudinal plan was measured. Figure 9 shows the measured acoustic field along the XZ Plan for our single element transducer excited by a 200 V negative pulse. The Z axis was considered the longitudinal one.

Considering that the ceramic used had a diameter $d = 10.5$ mm, and, for the thickness mode, a central frequency around 240 kHz, we would have a far field transition at $z = 4.12$ mm. As shown in Fig. 9, with this pulse excitation for a single transducer, it is possible to reach an acoustic pressure higher than 100 kPa (value expected to be necessary to produce microbubbles cavitation for sonothrombolysis) [11], up to a 40 mm-distance from the face of the transducer.

Another point observed is the absence of the expected two-sided lobes in the longitudinal plan.

To verify the possible effects of the dominant radial vibration mode, a 4-cell linear array was built and, using the hydrophone, the radial plan acoustic field was raised for the plan 6 mm away from the face of the transducer.

Figures 10 and 11 show the results which allow verifying that, exciting one specific transducer with the same pulse of 200 V, the radial dominant mode does not affect its nearby transducers, once the pressure level caused by the excited transducer is less than 20 dB in its closer neighborhood.
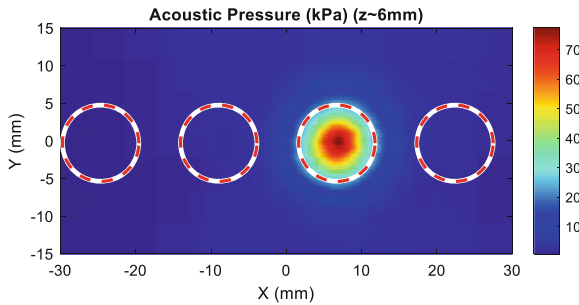


**Fig. 10.** XZ Plan Measured Pressure Acoustic Field
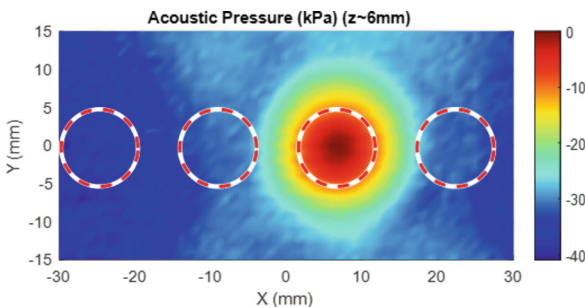


**Fig. 11.** Measured Pressure Acoustic Field in dB

## 4   Discussion

Sonothrombolysis, as a very promising ultrasound technique, is driving many different studies related to the application of piezoelectric materials, once the mechanism of enhancing thrombolysis is directly influenced by the acoustic parameters of the piezoelectric ultrasound transducers, like its central frequency and peak-to-peak acoustic pressure, both associated with the MI and cavitation generation [12].

In spite the fact the specific mechanisms of sonothrombolysis have not been completely elucidated, studies made *in vitro* models suggest that the process of the clot rupture is directly linked to the microbubbles cavitation and radiation force caused by the ultrasound waves [13].

In this context, a study using a canine graft occluded by thrombus, has showed that peak negative pressures generated by a 1 MHz diagnostic ultrasound, with a Mechanical Index higher than 0.5, but lower than 1.9, used with a continuous microbubble infusion, were effective at recanalizing acute intravascular thrombi. This study also showed the importance of cavitation, generated by high MI impulses, for producing thrombolysis using ultrasound *in vivo* [14].

Other study, this time using a 120-kHz continuous wave ultrasound with a peak-to-peak- acoustic pressure amplitude of 440 kPa, has demonstrated an enhancement of thrombolysis of blood clots in an *ex vivo* porcine carotid artery model, by using such ultrasound waves to promote and maximize the microbubbles stable cavitation activity [15].

From May 2014 to September 2015, at the Heart Institute of the University of Sao Paulo (InCor), the first human trials demonstrated the positive effect of ultrasound-guided cavitation of microbubbles during acute ST-segment elevation myocardial infarction [16]. This study showed that 1.3 MHz and 1.8 MHz diagnostic ultrasound short pulses ($<5$ $\mu$s), with a MI about 1.3, were effective at restoring both epicardial and microvascular flow [17].

All studies and trials above used diagnostic or therapeutic ultrasound equipment not specifically conceived for sonothrombolysis. As already mentioned, sub-megahertz frequency range seems to be more suitable for such application, considering its higher acoustic transmission efficiency [3] aiming at promoting microbubbles cavitation. Therefore, considering its specific use for cardiac sonothrombolysis, it becomes important to evaluate the dominant transducer vibrational modes, particularly the radial vibration mode, and its influence on providing the necessary acoustic pressure to generate cavitation when compared to the mentioned successful studies using diagnostic ultrasound transducers.

In our case, due to the piezo ceramic geometric features, a dominant radial mode, in a frequency around 170 kHz, appears when the transducer is excited by a pulse. Considering this frequency, that mode presents a transition far field in x = 1.16 mm; therefore a very small amount of its energy could be detected either by a reflected signal, in the pulse-echo measurements, or by the hydrophone placed in the longitudinal plan; however, a strong signal in this frequency was detected not only from a 100-mm distant reflector, on the face of the rod, but also by the hydrophone receiver.

This strong signal, at the frequency of the radial mode, is generated by the displacement of the piezoelectric material along the z-axis (within its thickness) due to the

Poisson effect in the ceramic ($\nu = -\varepsilon_{rad}/\varepsilon_{long}$, where $\nu$ is the Poisson's ratio, $\varepsilon$ is the Strain, defined in elementary form as the change in length divided by the original length [18]). Therefore, the expansion/contraction in the radial direction is associated with a contraction/expansion in the ceramic thickness direction, creating a coupled thickness mode vibration, near to the 170 kHz frequency, stronger than the original native thickness mode at around 240 kHz.

Such additional coupled vibration generates an extra acoustic pressure, in a frequency near to the expected thickness mode, which improves the transmitted energy in the longitudinal axis of our transducer, helping to generate the necessary acoustic pressure to cause the desired microbubbles cavitation on the cardiac sonothrombolysis, without creating undesired cross-talking in a 2D-transducer-matrix array designed to generate the ultrasound waves in the cardiac region.

## 5   Conclusions

The overall results showed that, in some low frequency piezoelectric ceramics (resonance frequency lower than 1 MHz), their typical radial vibration modes, rather than being a problem and due to the Poisson effect, can produce a useful coupled thickness vibration mode that can improve the effectiveness of the transducer, providing the necessary acoustic pressure to be used in a cardiac sonothrombolysis equipment.

Additionally, to the exposed measurements, a complete computing simulation, creating a virtual cardiac region, through k-wave [19] and Matlab [20], using the evaluated rod transducer in an 8X8 matrix array configuration was executed and showed that the solution can produce the necessary acoustic pressure in the entire cardiac region.

For future work, an 8X8 matrix array prototype, using the ceramics evaluated in this study, shall be produced and tests *in vitro* shall be performed to evaluate its effectiveness to generate the expected cavitation in microbubbles. Other specific study, with transducers using ceramics with both different diameter-to-thickness ratio and nominal resonant frequencies, shall be carried out to compare their efficiency for cardiac sonothrombolysis use.

**Conflict of Interest.**   The authors declare that they have no conflict of interest.

## References

1. World Health Organization: The top 10 causes of death (2018). https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed 05 July 2019
2. Hoskins, A., Martin, P., Thrush, K.: Diagnostic Ultrasound: Physics and Equipment, 3rd Edn., vol. 46, no. 4. CRC Press, Boca Raton (2019)

3. Bader, K.B., Bouchoux, G., Holland, C.K.: Sonothrombolysis. Adv. Exp. Med. Biol. **880**, 339–362 (2016). https://doi.org/10.1007/978-3-319-22536-4_19

4. Kunkel, H.A., Locke, S., Pikeroen, B.: Finite-element analysis of vibrational modes in piezo-electric ceramic disks. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**(4), 316–328 (1990). https://doi.org/10.1109/58.56492

5. de Jong, N., Bom, N., Souquet, J., Faber, G.: Vibration modes, matching layers and grating lobes. Ultrasonics **23**(4), 176–182 (1985). https://doi.org/10.1016/0041-624X(85)90027-7

6. Sayers, C.M., Tait, C.E.: Ultrasonic properties of transducer backings. Ultrasonics **22**(2), 57–60 (1984). https://doi.org/10.1016/0041-624X(84)90022-2

7. Guo, N.: The vibration characteristics of piezoelectric discs, Imperial College of Science, Technology and Medicine (1990)

8. Shaw, E.A.G.: On the resonant vibrations of thick barium titanate disks. J. Acoust. Soc. Am. **28**(1), 38–50 (1956). https://doi.org/10.1121/1.1908218

9. ASTM: Standard Practice for Evaluating Characteristics of Ultrasonic Search Units, ASTM E1065/E1065M-20, vol. 03.03, no. 06, p. 24 (2020)

10. Kinsler, L.E., Frey, A.R., Coppens, A.B., Sanders, J.V.: Fundamentals of Acoustics, 4th edn. (2000)

11. Bader, K.B., Holland, C.K.: Gauging the likelihood of stable cavitation from ultrasound contrast agents. Phys. Med. Biol. **58**(1), 127–144 (2013). https://doi.org/10.1088/0031-9155/58/1/127

12. Goel, L., Jiang, X.: Advances in sonothrombolysis techniques using piezoelectric transducers. Sensors **20**(5), 1288 (2020). https://doi.org/10.3390/s20051288

13. Bader, K.B., Gruber, M.J., Holland, C.K.: Shaken and stirred: mechanisms of ultra-sound-enhanced thrombolysis. Ultrasound Med. Biol. **41**(1), 187–196 (2015). https://doi.org/10.1016/j.ultrasmedbio.2014.08.018

14. Xie, F., et al.: Treatment of acute intravascular thrombi with diagnostic ultrasound and intra-venous microbubbles. JACC Cardiovasc. Imaging **2**(4), 511–518 (2009). https://doi.org/10.1016/j.jcmg.2009.02.002

15. Hitchcock, K.E., et al.: Ultrasound-enhanced rt-PA thrombolysis in an ex vivo porcine carotid artery model. Ultrasound Med. Biol. **37**(8), 1240–1251 (2011). https://doi.org/10.1016/j.ultrasmedbio.2011.05.011

16. Mathias, W., et al.: Diagnostic ultrasound impulses improve microvascular flow in patients with STEMI receiving intravenous microbubbles. J. Am. Coll. Cardiol. **67**(21), 2506–2515 (2016). https://doi.org/10.1016/j.jacc.2016.03.542

17. Porter, T.R., Mathias, W.: Cardiovascular sonothrombolysis. Curr. Cardiol. Rep. **21**(8), 86 (2019). https://doi.org/10.1007/s11886-019-1167-z

18. Lakes, R.: Meaning of Poisson's ratio, Dept. of Engineering Physics, University of Wiscosin (2022). http://silver.neep.wisc.edu/~lakes/PoissonIntro.html

19. Treeby, B.E., Cox, B.T.: K-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave-fields. J. Biomed. Opt. **5**(2), 021314 (2010)

20. MathWorks: MATLAB_R2016b, The MathWorks Inc., Natick (2016)

# Identification of UV Filters in SPF 30 Sunscreens by Raman Spectroscopy

Vera Lúcia Taba[1](✉) , Agnes C. Trindade[1] , Pedro A. Marrafa[1] ,
Cintia R. de Oliveira[1] , and Landulfo Silveira Jr.[1,2]

[1] Universidade Anhembi Morumbi (UAM), PPG Stricto Sensu em Engenharia Biomédica, São José dos Campos, SP, Brazil
veralu.taba@gmail.com
[2] Centro de Inovação, Tecnologia e Educação (CITÉ), São José dos Campos, SP, Brazil

**Abstract.** Exposure to ultraviolet A (UVA) and ultraviolet B (UVB) radiation can cause premature aging, sagging skin, hyperchromias and several mutations in the DNA of cells. When UV radiation reaches the skin, it also produces free radicals and the cumulative effect of exposure to sunlight can even lead to skin cancer. Due to these harmful factors, the use of sunscreen is recommended as a preventive measure. Nowadays, Sun Protection Factor (SPF), determined *in vivo*, is a universal indicator of the efficacy of sunscreen products against sunburn. Faced with the numerous options of brands, it is essential to know if the compositions described on the packaging labels are really present in the product. Thus, this study aims to identify the UV filters present in seven SPF 30 sunscreens by Raman spectroscopy. Some UV filters informed in the composition of some samples, could not be identified, since the characteristic Raman peaks were not found in the literature, such as: bemotrizinol, diethylamino hydroxybenzoyl hexyl benzoate, octyltriazone and homosalate. Raman spectroscopy can be an effective technique to identify UV filters in sunscreens, in addition to being used in quality control. The lack of identification of some UV filters occurred due to the absence of Raman spectra in the literature. This fact suggests the need to continue the study to obtain the Raman spectra of these and other UV filters, in order to contribute to the Raman spectra database.

**Keywords:** Sunscreen · UV Filter · SPF · Raman Spectroscopy

## 1 Introduction

When reaching the skin without protection, UV radiation causes several chemical and morphological reactions, altering some molecules present in skin cells. Cumulative exposure to UV radiation can even lead to mutations in the cells' DNA. UV radiation also produces free radicals, which cause premature aging and cell death, and can cause more harmful effects such as skin cancer [1–3]. Due to these factors harmful to the skin, the use of sunscreen is recommended. Several studies demonstrated that the use of sunscreen minimizes the effects of radiation on the skin [4, 5].

The harmful effect of skin exposure to UV radiation is mediated by short-wavelength (290-320 nm) UVB that reaches and affects the epidermis and long-wavelength (320-400 nm) UVA that penetrates into the dermis. UVA radiation has subdivisions, classified as UVA II (320–340 nm) and UVA I (340–400 nm). Excessive acute exposure to UVB results in sunburn, which causes erythema. Due to the cumulative effect, even the multiple acute UVB exposure early in life, were associated with basal cell carcinoma and melanoma [6].

The cases of skin cancer resulting from excessive sun exposure increases every year. This neoplasm corresponds to about 30% of all malignant neoplasms in Brazil. Skin tumors are related to some risk factors, especially exposure to sunlight, such as: prolonged and repeated exposure, especially in childhood and adolescence; people who work in direct sunlight are more vulnerable to non-melanoma skin cancer; having a low phototype, with red or blond hair, or being albino; have a familiar history of skin cancer [7].

Sunscreen is any cosmetic preparation intended to come into contact with the skin and lips, with the sole or main purpose of protecting it against UVB and UVA radiation by absorbing, dispersing or reflecting incident radiation [8].

Sunscreens are essential in the practice of photoprotection and may contain physical filters to reflect and/or chemical filters to absorb UV radiation. In the USA, sunscreens are regulated by the Food and Drug Administration (FDA). European countries follow the guidelines of the European Cosmetics Association (COLIPA). In Brazil and Mercosul countries, the Agência Nacional de Vigilância Sanitária (ANVISA) regulates sunscreen production [6, 8].

In a photoprotective formulation there may be several solar filters, but it is important to note that some are incompatible and, if combined, will reduce the final SPF of the product. For instance, the avobenzone is stable and effective when combined with octocrylene but becomes unstable when combined with cinnamates such as cinoxate. On the contrary, the combination of another active component can increase the level of sun protection, improving photostability [9].

To assess the quality and effectiveness of sunscreens, SPF, UVA-PF, and water resistance must be determined using *in vivo* methods, following USA and European regulations [8, 10–14].

Sunscreens are classified according to a Solar Protection Factor (SPF) system, which measures the protection offered by a particular sunscreen against the appearance of skin erythema, which is basically caused by UVB radiation [6]. The Minimum Erythematous Dose (MED) is the minimum dose of ultraviolet radiation required to produce the first noticeable erythematous reaction with clearly defined edges, observed between 16 and 24 h after exposure to UV radiation, according to the adopted methodology. The SPF is primarily an indicator of effectiveness against UVB light. This standardized analysis requires the application of 2 mg/cm$^2$ of sunscreen in the test area [6].

The SPF is the value obtained by the ratio between the MED in a skin protected by a sunscreen (MEDp) and the MED in the same skin when unprotected (MEDu) [8, 13]:

$$SPF = \frac{MEDp}{MEDu} \tag{1}$$

The Minimum Pigmenting Dose (MPD) is the minimum dose of UVA radiation required to produce a persistent pigment darkening of the skin with clearly defined edges, observed between 2 and 4 h after exposure to UVA radiation. UVA Protection Factor (UVA-PF) is the value obtained by the ratio between the MPD in a skin protected by a sunscreen (MPDp) and the MPD in the same skin, when unprotected (MPDu). The UVA-PF must correspond to at least 1/3 of the SPF value declared on the labeling [8, 13]:

$$UVA - PF = \frac{MPDp}{MPDu} \tag{2}$$

The effectiveness of a sunscreen depends on the amount applied, as shown in Table 1. The current FDA standard for sunscreen application is 2 mg/cm$^2$ for the product to deliver the SPF described on the label. Studies suggest that sunscreen use by consumers is only 25% to 50% of the ideal amount, with an adult applying approximately 1/2 teaspoon to each arm, face, and neck. About a teaspoon should be applied to each leg, chest and back [9].

**Table 1.** SPF variation according to the amount of sunscreen applied [9].

| SPF | Amount of sunscreen applied to the skin | | | |
|---|---|---|---|---|
| | 2.0 mg/cm$^2$ | 1.5 mg/cm$^2$ | 1.0 mg/cm$^2$ | 0.5 mg/cm$^2$ |
| 2 | 2.0 | 1.7 | 1.4 | 1.2 |
| 4 | 4.0 | 2.8 | 2.0 | 1.4 |
| 8 | 8.0 | 4.8 | 2.8 | 1.7 |
| 15 | 15.0 | 7.6 | 3.9 | 2.0 |
| 30 | 30.0 | 12.8 | 5.5 | 2.3 |
| 50 | 50.0 | 18.8 | 7.1 | 2.7 |

The components of sunscreens differ in their absorption spectra, as shown in Table 2. Ideally, a sunscreen should provide protection against UV radiation across its full spectrum [9].

Several studies used Raman spectroscopy to identify photoprotective ingredients, solvents and other chemical compounds present in sunscreen formulations, especially the physical and chemical filters of such formulations [15–19]. These studies demonstrated that Raman spectroscopy can be an alternative for quality control and identification of UV filters in sunscreens.

Raman spectroscopy consists of obtaining the Raman scattering, discovered in 1928 by C. V. Raman, which depends on the processes of inelastic scattering of a monochromatic light beam by the molecule due to the polarization induced by the incident light beam. The Raman scattering measures the difference between the energy of the scattered photon and that of the incident photon. Such variation allows obtaining information about the chemical composition of the studied sample [20].

Table 2. UV absorption limit (nm) of sunscreens [9].

| UV Filter | Type | UV absorption limits (nm) |
|---|---|---|
| Aminobenzoic acid (PABA) | Organic or chemical | 260–313 |
| Homosalate | Organic or chemical | 300–310 |
| Cinoxate | Organic or chemical | 270–328 |
| Octyl methoxycinnamate | Organic or chemical | 270–328 |
| Octyl salicylate | Organic or chemical | 300–310 |
| Padimate O | Organic or chemical | 290–315 |
| Phenylbenzimidazole Sulfonic acid | Organic or chemical | 290–320 |
| Trolamine salicylate | Organic or chemical | 260–320 |
| Methyl anthralin | Organic or chemical | 290–320 |
| Oxybenzone | Organic or chemical | 270–350 |
| Sulisobenzone | Organic or chemical | 270–360 |
| Dioxybenzone | Organic or chemical | 260–380 |
| Avobenzone | Organic or chemical | 310–400 |
| Titanium dioxide ($TiO_2$) | Inorganic, physical or mineral | 250–400 |
| Zinc oxide (ZnO) | Inorganic, physical or mineral | 250–380 |
| Padimate O | Organic or chemical | 290–315 |

This study aims to identify the presence of physical and chemical solar filters described in the composition of six sunscreens and a lip balm, from seven different brands, all with SPF 30, through Raman spectroscopy analysis. The justification for the study is the constant need for methods that could be used in quality control of pharmaceutical products.

## 2  Materials and Methods

In this study, seven products with SPF 30 from different brands were used as samples, called P1 to P7 in the experiment, as follows:

P1: body sunscreen;
P2: colored facial sunscreen;
P3: facial and body sunscreen;
P4: lip sunscreen;
P5: body sunscreen;
P6: anti-aging facial sunscreen;
P7: facial sunscreen.

In addition to the seven samples studied, Tinosorb® M, Eusolex® 6300, avobenzone and $TiO_2$ solar filters were also analyzed for comparison with sunscreen samples and/or references in the literature.

The samples were submitted to Raman spectroscopy (model P1, Lambda Solutions Inc., MA, USA, excitation wavelength: 830 nm, laser power: 450 mW, 4 cm$^{-1}$ spectral resolution) under the same environmental conditions and laser power. Despite this, during the experiment it was necessary to decrease the laser power when analyzing sample P2, which has a beige color; such pigment was characterized as a chromophore and made analysis difficult.

All readings were obtained in triplicate. After the experiment, the spectra were submitted to pre-processing in the Microsoft Excel software. The spectra were subjected to baseline removal. Normalization was also performed by the area under the curve (1-norm) to eliminate the variable intensity. Thus, it was possible to observe that the characteristic peaks appear with approximate wavelength, but difference in intensity, most likely due to the concentration being different in each analyzed sample.

The mean spectra obtained in triplicate from P1 to P7 were plotted in a single graph to identify, through the positions and intensities of the peaks, which sunscreens had the same composition, and if the concentrations of the compounds were different or the same in the sunscreens studied.

## 3 Results and Discussion

The P1 sunscreen has in its composition the active ingredients: avobenzone [15, 16, LAB*], oxybenzone [16] and octocrylene [15]. In the obtained Raman spectrum shown in Fig. 1, it was possible to observe the presence of characteristic peaks of these compounds obtained through the published literature (Table 3, which will be seen later [15, 16, 21–23]). The good resolution of the spectrometer (4 cm$^{-1}$) allowed getting spectra with well resolved bands in the fingerprint range.
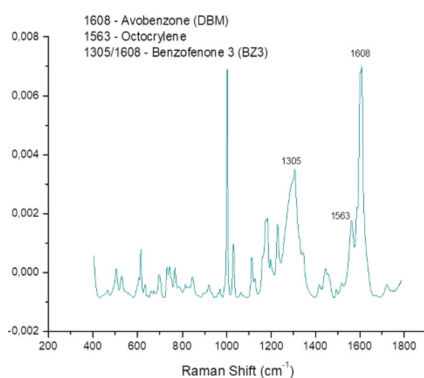


**Fig. 1.** Raman spectrum of the sample P1.

Among the UV filters described in P2 sample composition are octisalate [16], octinoxate [15, 16, 22] and benzophenone-3 [16]. These compounds identified in the Raman spectrum shown in Fig. 2. TiO$_2$, which should appear around 440 cm$^{-1}$ [23], was not identified. This may be due to the presence of the dye in the formulation, which is

characterized as a chromophore/fluorophore, which caused absorption and consequent emission of fluorescence, making it difficult to collect the Raman spectrum. There was destruction of the sample and, consequently, a lower signal-to-noise ratio in relation to the other products. The P2 analysis was repeated with lower laser power to obtain adequate scattering, also improving the signal-to-noise ratio due to fluorescence.
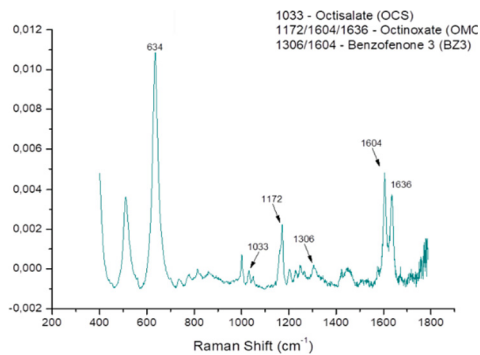


**Fig. 2.** Raman spectrum of the sample P2.

For the bemotrizinol described in the composition of sample P3, no reference literature was found, but the other compounds such as avobenzone [15, 16, LAB*], octisalate [16] and octocrylene [15] were identified from the characteristic peaks indicated in Fig. 3.
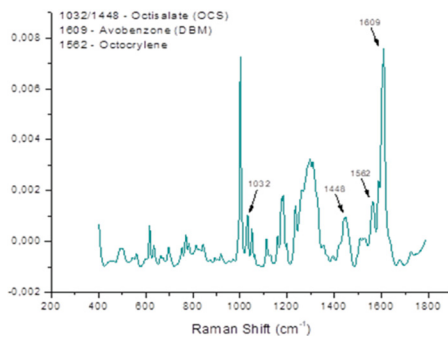


**Fig. 3.** Raman spectrum of the sample P3.

Sample P4 has in its composition two compounds that have no reference spectrum in the literature, namely Uvinul® A Plus and homosalate. Octinoxate [15, 16, 22] and $TiO_2$ [23] were possible to be identified in the Raman spectrum, shown in Fig. 4.
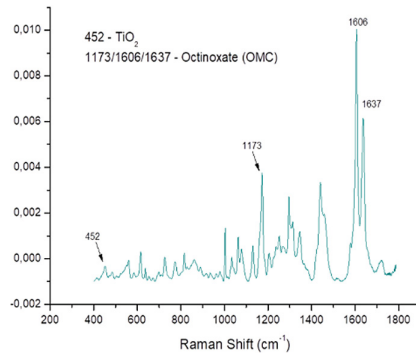
**Fig. 4.** Raman spectrum of the sample P4.

Sample P5 has in its packaging as main ingredients $TiO_2$ [23], Tinosorb® M [LAB*], octisalate [16], octinoxate [15, 16, 22], avobenzone [15, 16, LAB*], Eusolex® 6300 [LAB*] and octocrylene [15], all these compounds were possible to be identified in the Raman spectra on Fig. 5 according to the characteristic peaks of each ingredient described in the literature.
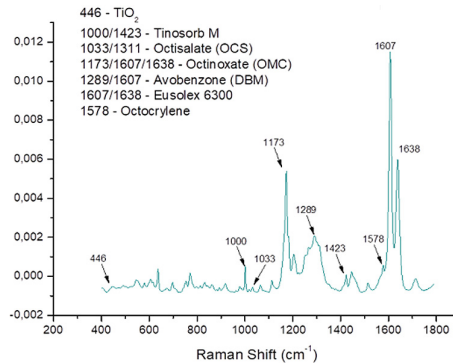


**Fig. 5.** Raman spectrum of the sample P5.

In sample P6, only homosalate and Uvinul® A Plus that are not described in the literature could not be identified in the spectrum of Fig. 6. Were identified $TiO_2$ [23], Tinosorb® M [LAB*], octinoxate [15, 16, 22], ODPABA [16], and octocrylene [15].
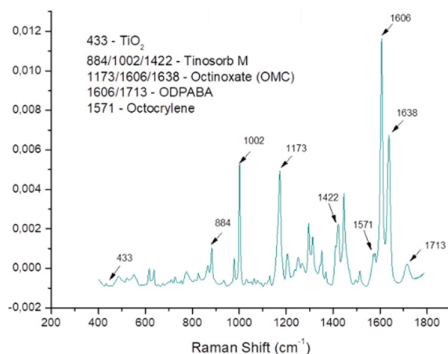
**Fig. 6.** Raman spectrum of the sample P6.

The compounds bemotrizinol and Uvinul® T150 described in the label of sample P7 have no reference in the literature, making it difficult to identify them in the spectrum shown in Fig. 7, in which the characteristic Raman peaks of the other photoprotective actives appear, whose wavelengths were confirmed in the literature: $TiO_2$ [23], octocrylene [15], avobenzone [16] and benzoic acid (ODPABA) [16].
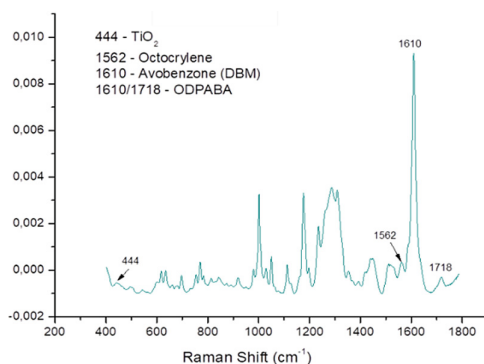


**Fig. 7.** Raman spectrum of the sample P7.

The Tinosorb® M and Eusolex® 6300 solar filters (Fig. 8 and Fig. 9, respectively), whose characteristic peaks could not be compared to the literature, were identified by the experiment (LAB)* and may need confirmation, through dilutions and comparisons with sunscreens that have the same actives, and/or a new search for references, which would imply in the continuity of the study. The fact that these samples were analyzed pure (without dilution) may have made it difficult to read the Raman peaks from ingredients in low concentrations.
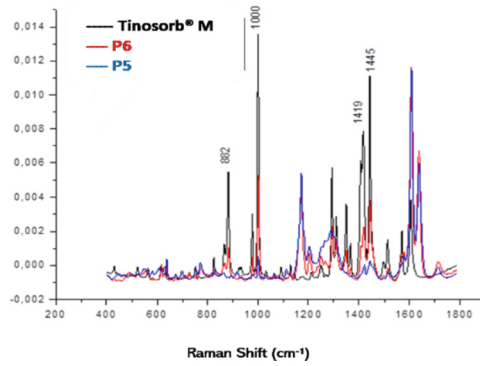
**Fig. 8.** Raman spectra of the samples P5, P6 and Tinosorb® M.
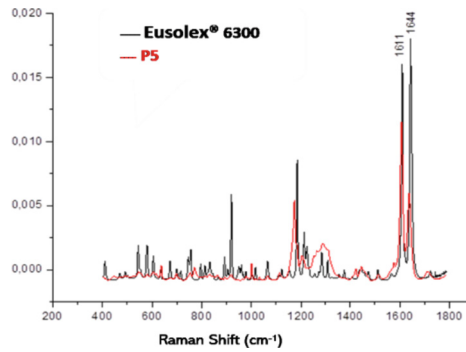


**Fig. 9.** Raman spectra of the sample P5 and of the Eusolex® 6300.

Table 3 shows all UV filters, marked with X, present in the respective sample compositions and the characteristic Raman peaks, according to the literature and/or experiment (LAB)*.

# 4 Conclusions

In this study, Raman spectroscopy analysis was performed to identify UV chemical filters in seven SPF 30 products from different brands. Most of the ingredients informed in the respective compositions were detected and identified through the characteristic Raman peaks described in the literature and/or obtained by Raman spectra. However, the

**Table 3.** UV filters present in samples P1 to P7 and characteristic Raman peaks.

| UV filters (INCI name) | USAN** name / trademark [21] | UV filters present in samples, according to manufacturers | | | | | | | Characteristic Raman peaks | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | P1 | P2 | P3 | P4 | P5 | P6 | P7 | Literature / LAB* ($cm^{-1}$) | Samples ($cm^{-1}$) |
| 4-Methylbenzylidene camphor | Enzacamene Eusolex® 6300 | | | | | X | | | 1611 LAB* 1644 LAB* | P5: 1607/1638 |
| Butylmethoxydibenzoylmethane | Avobenzone BMBM/DBM Parsol® 1789 | X | | X | | X | | X | 1280/1601 [16] 1611 [15] 1607 LAB* | P1: 1608; P3: 1609 P5: 1289/1607 P7: 1610 |
| Benzyl salicylate | Benzoic acid (ODPABA) | | | | | | X | X | 1183/1277/1607/ 1701 [16] | P6: 1606/1713 P7: 1610/1718 |
| Benzophenone-3 | Oxybenzone BP3/BZ3 | X | X | | | | | | 1228/1305/1602/ 1622 [16] | P1: 1305/1608 P2: 1306/1604 |
| Bis-ethylhexyloxyphenol methoxyphenyl triazine | Bemotrizinol Tinosorb® S | | X | X | | | X | X | Not found in the literature | ————— |
| Diethylamino hydroxybenzoyl hexyl benzoate | Uvinul® A Plus | | | | X | | X | | Not found in the literature | ————— |
| Ethylhexyl methoxycinnamate or Octyl methoxycinnamate | Octinoxate EHMC/OMC Uvinul® MC 80 | | X | | X | X | X | | 1170/1604/1634 [16] 1177/1605 [15] 1176/1608/1640 [22] | P2: 1172/1604/1636 P4: 1173/1606/1637 P5: 1173/1607/1638 P6: 1173/1606/1638 |
| Ethylhexyl salicylate or Octyl salicylate | Octysalate EHS/OCS Neo Heliopan® OS | | X | X | | X | | | 1033/1249/1325/ 1466/1674 [16] | P2: 1033 P3: 1032/1448 P5: 1033/1311 |
| Ethylhexyl triazone | Octyltriazone Uvinul® T150 | | | | | | | X | Not found in the literature | ————— |
| Homosalate | Homomenthyl salicylate Eusolex® HMS | | X | | X | | X | | Not found in the literature | ————— |
| Methylene bis-benzotriazolyl tetramethylbutylphenol | Bisoctrizole Tinosorb® M | | | | X | X | | | 882 LAB* 1000 LAB* 1419 LAB* 1445 LAB* | P5: 1000/1423 P6: 884/1002/1422 |
| Octocrylene | Octocrylene Uvinul® N539 T | X | | X | | X | X | X | 1560 /1564 [15] | P1: 1563 // P3: 1562 P5: 1578 // P6: 1571 P7: 1562 |
| Titanium dioxide | CI 77891 Eusolex® T2000 | | X | | X | X | X | X | 450 LAB* 600 LAB* 440 [23] | P2: TiO$_2$ wasn't found P4: 452 // P5: 446 P6: 433 // P7: 444 |

[*] LAB: Samples or standards, not found in the literature that were submitted to analysis in the spectrometer to verify the peaks with greater intensity, which were considered characteristic Raman peaks, after comparing the products.

** USAN: United States Adopted Names Council

characteristic Raman peaks of the actives: bemotrizinol, diethylamino hydroxybenzoyl hexyl benzoate, octyltriazone and homosalate, reported in the composition of some samples, were not found in the literature, and therefore could not be confirmed.

Among the UV filters described in the respective compositions and referenced in the literature, only $TiO_2$ was not identified in sample P2. The aforementioned sample is a colored sunscreen, in which the dye is characterized as a chromophore/fluorophore, causing absorption and emission of fluorescence, which made it difficult to collect the Raman spectrum.

Raman spectroscopy may be considered a tool for identifying active ingredients in sunscreens, but it would be necessary to continue this preliminary study to confirm the Raman peaks characteristic of ingredients in which there is still no reference in the literature.

**Conflict of Interest.** The authors have no conflict of interest to declare.

# References

1. Rangarajan, M., Zats, J.: Effect of formulation on the topical delivery of α- tocopherol. J. Cosmet. Sci. **54**, 161–174 (2003)
2. Palm, M.D., O'Donoghue, M.N.: Update on photoprotection. Dermatol. Ther. **20**, 360–376 (2007)
3. Sgarbi, F.C., Carmo, E.D., Rosa, L.E.B.: Radiação ultravioleta e carcinogênese. Rev. Cienc. Med. **16**, 245–250 (2007)
4. Lautenschlager, S., Wulf, H.C., Pittelkow, M.R.: Photoprotection. Lancet **370**, 528–537 (2007)
5. Baron, E.D., Kirkland, E.B., Domingo, D.S.: Advances in photoprotection. Dermatol. Nurs. **20**, 265–272 (2008)
6. Costa, A.: Tratado Internacional de Cosmecêuticos. Guanabara Koogan, Rio de Janeiro (2012)
7. Instituto Nacional de Câncer (INCA): Câncer de pele não melanoma. https://www.inca.gov. br/tipos-de-cancer/cancer-de-pele-nao-melanoma. Accessed 06 June 2022
8. Agência Nacional de Vigilância Sanitária (ANVISA): Resolução N.° 30. Regulamento Técnico Mercosul sobre Protetores Solares em Cosméticos. https://bvsms.saude.gov.br/bvs/sau delegis/anvisa/2012/rdc0030_01_06_2012.html. Accessed 06 June 2022
9. Draelos, Z.D.: Cosmecêuticos - Procedimentos em Dermatologia Cosmética. Elsevier, Rio de Janeiro (2005)
10. Food and Drug Administration (FDA): sunscreen drug products for over-the counter human use; tentative final monograph. Fed. Reg. 59:28194–28302. https://www.federalregister. gov/documents/2019/02/26/2019-03019/sunscreen-drug-products-for-over-the-counter-hum an-use. Accessed 06 June 2022
11. Verheugen, G.: Commission recommendation of 22 September 2006 on the efficacy of sunscreen products and the claims made relating thereto. Official J. Eur. Union **265**, 39–43 (2006)

12. Cosmetics Europe: guidelines for evaluating sun product water resistance. https://www.cosmeticseurope.eu/files/7914/6407/7400/Guidelines_for_Evaluating_Sun_Product_Water_Resistance_-_2005.pdf. Accessed 06 June 2022

13. Fuehring, S.: Standardisation mandate assigned to CEN concerning methods for testing efficacy of sunscreen products. Brussels, Belgium: European commission. https://ec.europa.eu/growth/tools-databases/mandates/index.cfm?fuseaction=search.detail&id=346. Accessed 06 June 2022

14. Cosmetics Europe: In vitro method for the determination of the UVA protection factor and "critical wavelength" values of sunscreen products. COLIPA: Auderghem, Belgium, pp. 1–29 (2011)

15. da Costa, M.M., Alves, L.P., Osório, R.A.L., Pacheco, M.T.T., Jr Silveira, L.: Detecting active ingredients of insect repellents and sunscreens topically in skin by Raman spectroscopy. J. Biomed. Opt. **23**(10), 107003 (2018)

16. Beyere, L., Yarasi, S., Loppnow, G.R.: Solvent effects on sunscreen active ingredients using Raman spectroscopy. J. Raman Spectrosc. **34**, 743–750 (2003)

17. Egawa, M., Iwaki, H.: In vivo evaluation of the protective capacity of sunscreen by monitoring Urocanic acid isomer in the stratum corneum using Raman spectroscopy. Skin Res. Technol. **14**, 410–417 (2008)

18. Oladepo, S.A., Loppnow, G.R.: Ultraviolet resonance Raman spectroscopy as a robust spectroscopic tool for in situ sunscreen analysis. Anal. Chim. Acta **628**, 57–66 (2008)

19. Tippavajhala, V.K., Mendes, T.O., Martin, A.A.: In vivo human skin penetration study of sunscreens by confocal Raman spectroscopy. AAPS Pharm. Sci. Tech. **19**, 753–760 (2018)

20. Cartaxo, S.B.: FT-Raman spectroscopy for the differentiation between cutaneous melanoma and pigmented nevus. Acta Cir. Bras. **25**, 351–356 (2010)

21. Osterwalder, U., Sohn, M., Herzog, B.: Global state of sunscreens. Photodermatol. Photoimmunol. Photomed. **30**, 62–80 (2014)

22. Cheng, J., Li, Y.S., Roberts, R.L., Walker, G.: Analysis of 2-ethylhexyl-p-methoxycinnamate in sunscreen products by HPLC and Raman spectroscopy. Talanta **44**, 1807–1813 (1997)

23. Harvanová, M., Mašek, V., Jírová, D., Kolářová, H.: Raman fabel-free visualisation of titanium dioxide nanoparticles uptake in BJ cell lines. Lékař a Technika **46**(1), 25–28 (2016)

# Development of a Responsive System with Immersion in a Virtual Environment Integrated into a Biaxial Force Platform for Balance Training

Bruno Toshio Gomes Gunji[(✉)] , André Roberto Fernandes da Silva ,
Luan Almeida Moura , Mariana Palma Valério , Silvia Cristina Martini ,
Silvia Regina Matos da Silva Boschi , Terigi Augusto Scardovelli ,
and Alessandro Pereira da Silva

Technology Research Center, University of Mogi das Cruzes (UMC), Av. Dr. Cândido X. de
Almeida E Souza, Mogi das Cruzes, Brazil
`bruno-toshio2013@hotmail.com`, {`silviac,boschi,terigiscardovelli,
alessandrops`}`@umc.br`

**Abstract.** Balance is an essential skill for human motor functioning that causes a sense of security and is indispensable for stability during movement. In this process, the postural orientation provides stimuli responsible for sensations of movement that help in the environmental orientation. The reduction of balance can favor the increase of falls providing some type of injury, which in more serious conditions lead to the need for care with therapies. Currently, one of the strategies for maintaining balance is multisensory training, which consists of series of exercises with sensorized equipment with static and dynamic metrics. One of these equipment is the force platform that measures the fluctuation of the Center of Mass (COM) and Pressure (COP) with load cells. To provide a balance training experience for people who need it, a bipedal force platform with biaxial rotation was used with a virtual environment designed to use the flexibility of chang-ing the inclination of the force platform along with a game element.

**Keywords:** Balance board · Games · Responsive system

## 1 Introduction

Balance is an essential skill for individuals' motor functionality and can be defined as the ability to maintain stable posture despite external disturbances [1]. This skill is essential to provide security and stability during movements [2].

In order for the individual to achieve postural stability, it is necessary to use several sensory systems, which provide information according to the levels of attention [3].

Postural orientation has three main stimuli, exproprioception, which is responsible for the sensation of position and movement of a body part relative to the environment,

represented by the vestibular, auditory, somatosensory and visual systems. Proprioception, responsible for the sense of position and movement of a part of the body relative to another, being represented by the somatosensory and visual system. And finally, exteroception, which is responsible for locating an object in the environment in relation to another, represented by the auditory and visual system [4].

Impaired balance can favor an increase in falls, which consequently can increase the risk of injury [5], in some cases, loss of balance can contribute to more serious conditions that require intensive care [6].

One of the strategies for maintaining postural balance is multisensory training, which consists of a series of exercises with static and dynamic metrics sensorized equipment, which can be performed in any environment [7].

Among the studies collected in the systematic review of [8], there were no games able to directly control the actions of a dynamic force platform. A game capable of performing this control could make the virtual environment more immersive and interactive. Such a system could stimulate the player to continue training, increasing adherence to the protocol [9].

Therefore, the objective of this study is to develop an immersive balance training game capable of controlling the slopes of a biaxial dynamic force platform.

## 2 Material and Methods

### 2.1 Equipment Used

To assist the immersion and stimulate the individual's motor training, a bipedal force platform with biaxial rotation was used (Fig. 1). The equipment is composed of eight load cells (Fig. 1A) distributed in two metal plates (Fig. 1B), they are coupled to an axis controlled by four stepper motors (Fig. 1C), which receive command from the microcontrollers by the speed of 74880 bps (Fig. 1D), which change the slope of the metal plates of the platform.
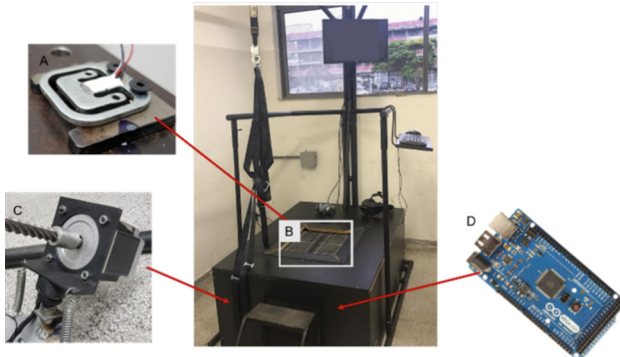


**Fig. 1.** Force *platform with biaxial rotation. A) Load cell, A) Load cell, B) Metal plate, C) Stepper motor, D) Microcontroller.*

## 2.2 Platform/Game Interactivity

The interactions between the platform and the game occur via scripts by the Visual Studio coding software, which establishes serial communication with the microcontrollers, responsible for sending and receiving information from the respective hardware (Fig. 2).
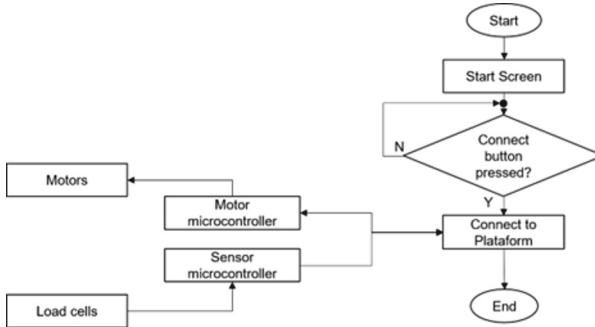


**Fig. 2.** Platform/Game connection

When passing a checkpoint, the game informs the force platform slope value to the motors microcontroller.

The game moves the avatar based on the value received from the equipment's sensors. The posterior sensors are responsible for the frontal movement at the beginning of the game, the anterior sensors are responsible for the decrease in speed, and the lateral-lateral sensors are responsible for the avatar's medio-lateral movements (Fig. 3).
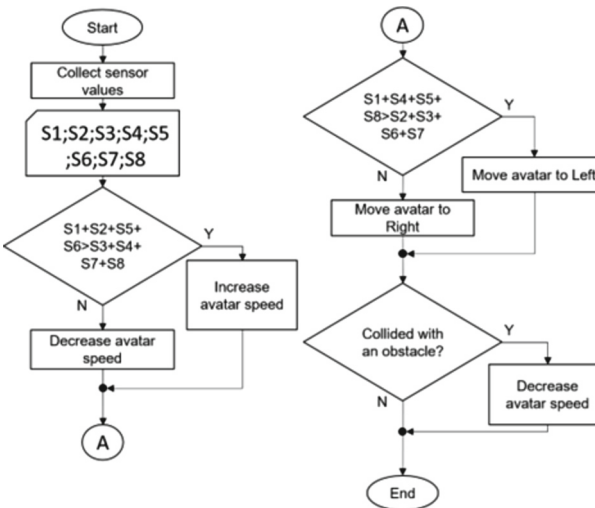


**Fig. 3.** Player movement

The player's avatar has a RigidBody (responsible for gravity physics in the character), a BoxCollider (responsible for collision interactions and platform tilt changes) and a script (which counts the events during the level). In addition to interpreting the values received by the microcontroller for movement. Movement is limited to a radius of 180° (one hundred and eighty degrees) forward.

## 2.3 Game Development

The game was developed by the Unity Graphics Engine software, and modeled in 3D by the Blender software.

It consists of a Ski Runner game and aims to go down a snowy mountain with a pair of skis, it also contains several obstacles that, when reached, the descent speed is reduced, thus causing an increase in the completion time of the game. To control the avatar, it is necessary for the volunteer to move his center of pressure (COP) to the right or to the left, thus generating the medio-lateral movement. The faster the volunteer reaches the end game objective, the better his performance.

At the beginning of the game, the user is presented to the initial screen (Fig. 4A), which allows the player to enter his name, connect with the force platform and finally
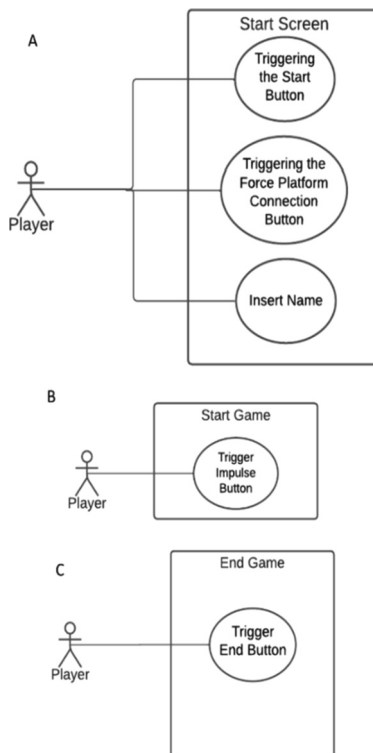


**Fig. 4.** Use case diagram. A) Start screen use case diagram. B) Use case diagram of the beginning of the game. C) Endgame use case diagram.

start the game. After pressing the start button, the screen switches to the avatar view, which contains the match time and a representation of the player's current speed. On this screen, the player can start the course through an "Impulse" (Fig. 4B). At the end of the path, the game ends with the non-player characters (NPCs) celebrating the player's arrival. Finally, the final screen is presented, with the total starting time and the number of collisions with obstacles, on this screen there is a button for the game's initial screen (Fig. 4C).

The track is composed of a set of random terrain segments (prefabricated objects) that have BoxCollider and RigidBody for interaction with the avatar. There is also a script whose function is to randomly spawn objects in a straight line). Obstacles have BoxCollider to count collisions and a script to slow down the avatar. The side barriers contain RigidBody that prevent the player from leaving the specified route.

At the beginning of the game, the system generates random challenges, different in number of obstacles and inclination from one stage to another.

The game has checkpoints (arcs) containing BoxCollider and script that inform the desired inclination for the motors microcontroller that readjusts the inclination of the platform (Fig. 5). The player must adjust their center of mass (COM) to match the slope faced.
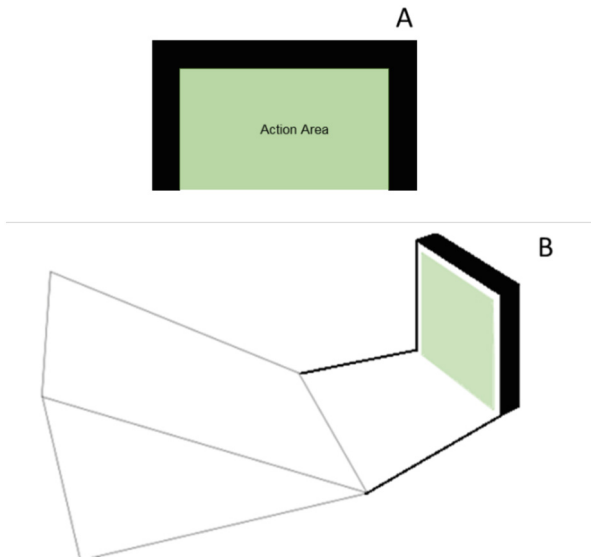


**Fig. 5.** 5 Checkpoint representation. A) Front Checkpoint; B) Checkpoint in three/fourth view.

During the course, the player must dodge obstacles by shifting his center of pressure accordingly. If the player hits an obstacle, his speed will be reduced, increasing the time needed to complete the course.

To validate the game's features, a Black Box test was performed, which consists of testing each software function, observing the behavior and analyzing each response received by the command sent.

The actions of the Home Screen Buttons, the "Impulse" function, the controls of medio-lateral movements, actions of increase and decrease of speed, the expected movements of the responsive system and the collisions of obstacles were analyzed.

At the end of each analysis, the answer given by the software is compared with the programmed answer.

## 3  Results

The game is a Ski Runner and to implement and increase the player's immersion in the game, a realistic scenario was made (Fig. 6), stones (Fig. 7A), trees (Fig. 7B) and logs (Fig. 7C) were implemented.



**Fig. 6.** Scenery.

When running the game, the user is directed to the initial screen (Fig. 8), where the type of control is selected, the player's name is entered and the game starts.

By pressing the start button the player will be directed to the player view (Fig. 9) where he can start the game with a "Impulse" button.

The track is formed by side barriers and some type of obstacle, and at the end of each obstacle sequence a checkpoint area is formed.

The checkpoint change areas (Fig. 10) were made to be visible during the game, thus encouraging the player that from that moment on, the platform inclination and/or speed can be changed.

Finally, when reaching the end of the course (Fig. 11) there is a commemoration that symbolizes the end of the game and presents the information about the occurrence of the game (Fig. 12). The game's rewards vary depending on the player's performance during the level. The variables analyzed to determine the feedback are the time and the total of collisions.

To validate the functionality of the software, the black box test was performed. This test verifies the functions and performance without looking at the internal functionality of the software [10]. The system is segmented into individual actions (seven actions), where parts of the total functionality were disabled to observe individual behavior, comparing the system response with the expected response.
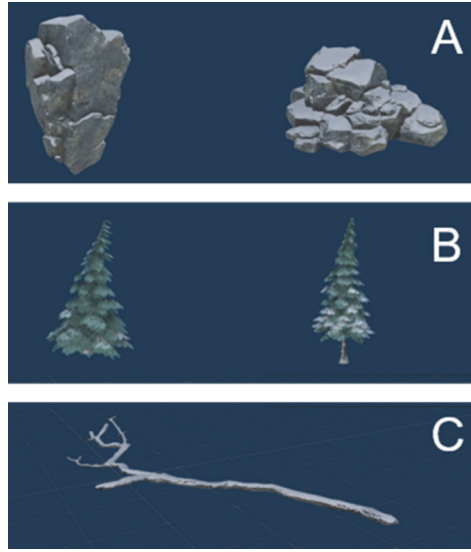
**Fig. 7.** Scene objects. A) Stones; B) Trees; C) Trunk.



**Fig. 8.** Game start screen. A) Control selection; B) Start game button; C) Connect to platform; D) Insertion of name.

The results were organized in the board (Fig. 13) to verify if there is any problem in the execution of the codes and functionalities.

### 3.1 Game Timeline

The game starts with the player's impulse, that is, when leaning forward, the platform's front sensors capture the pressure exerted and send it to the game avatar. Upon passing the first checkpoint, the platform tilts and starts the timer.

With the beginning of the game, the user will be able to perform side-to-side movements that will be captured by the sensors to move the avatar. Forward inclusions increase the avatar's speed, conversely, backwards inclusions decrease the avatar's speed.

**Fig. 9.** Player view. A) Speedometer; B) Stopwatch.
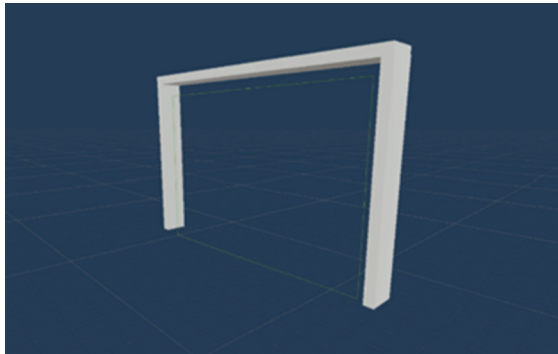


**Fig. 10.** Checkpoint.
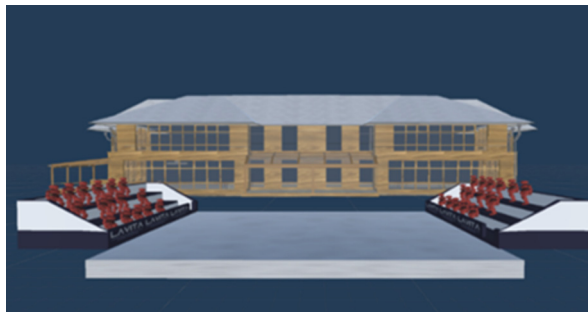


**Fig. 11.** End of game.

The player's objective is to reach the end of the track in the shortest possible time, avoiding obstacles. Checkpoints are points that symbolize increase/decrease in the slope of the track and change the slope of the platform.

**Fig. 12.** End of game screen. A) Number of hits occurred; B) Game end/restart button; C) Total travel time.

| Action | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Pressed Start/End Button | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Pressed Boost Button | X | No | Yes | Yes | Yes | Yes | Yes |
| Individual's Center of Mass left/right | X | X | No | X | X | X | X |
| Individual's Center of Mass backwards | X | X | X | No | X | X | X |
| Individual's Center of Mass foward | X | X | X | X | No | X | X |
| Went through the tilt change | X | X | X | X | X | No | X |
| Collided into na obstacle | X | X | X | X | X | X | No |
| Expected Output | It wasn't directed to the beginning of the game | Not started the game | Character didn't move sideways | Character hasn't slowdown | Character hasn't speed boost | There was no change of slope on the plataform | No character speed reduction |

**Fig. 13.** Black Box functionality test.

At the end of the game, the player celebrates with the NPCs the completion of the phase, which finally shows the time and the total number of collisions that occurred during the game.

# 4   Conclusions

A virtual environment was produced that can control and interact with a biaxial force platform. The purpose of this game is to provide stimuli in real time through the interface and to promote the experimentation of a snow skiing environment, consequently helping in balance training to different audiences (elderly, patients with neurological pathologies or in the process of motor rehabilitation due to sports or traumatic injuries).

**Conflict of Interest.** The authors declare that there is no conflict of interest regarding the publication of this paper.

# References

1. Mann, L., Kleinpaul, J.F., Mota, C., Santos, S.G.: Equilíbrio corporal e exercícios físicos: uma revisão sistemática Corporal balance and physical exercises: a systematic review. Motriz rev. educ. fís. LILACS ID: lil-535220 (2009)
2. Manna, L., et al.: Investigation of the corporal balance in elderly people. Rev. Bras. de Geriatria e Gerontologia **11**(2), 155–165 (2008)
3. Lundy-Ekman, L.: Neurociência: Fundamentos Para a Reabilitação, 2nd ed. Elsevier, Rio de Janeiro (2004)
4. Kleiner, A.F.R., Schlittler, D.X.D.C., Sánchez Arias, M.D.R.: O papel dos sistemas visual, vestibular, somatosensorial e auditivo para o controle postural. Rev. Neurociências **19**(2), 349–357 (2011). https://doi.org/10.34024/rnc.2011.v19.8382
5. Lim, D., et al.: Use of the Microsoft Kinect system to characterize balance ability during balance training. Clin. Interv. Aging **10**, 1077 (2015)
6. Ayed, I., et al.: Feasibility of Kinect-based games for balance rehabilitation: a case study. J. Healthcare Eng. **2018**, 8 (2018)
7. An, M., Shaughnessy, M.: The effects of exercise-based rehabilitation on balance and gait for stroke patients: a systematic review. J. Neurosci. Nurs. **43**(6), 298–307 (2011). https://doi.org/10.1097/JNN.0b013e318234ea24. PMID: 22089406.
8. Schröder, J., et al.L Combining the benefits of tele-rehabilitation and virtual reality-based balance training: a systematic review on feasibility and effectiveness. Disabil. Rehabil. Assist. Technol. **14**(1), 2–11 (2019). https://doi.org/10.1080/17483107.2018.1503738. Epub 2018 Oct 14. PMID: 30318952.
9. Yim, J., Graham, N.: Using games to increase exercise motivation. In: Proceedings of the 2007 Conference on Future Play (Future Play 2007). Association for Computing Machinery, New York, NY, USA, pp. 166–173 (2007)
10. Pressman, R.: Engenharia De Software (2006)

# Dental Enamel Remineralization Following Diode Laser Irradiation

Márcia Regina Cabral Oliveira[1] , Ilka Tiemy Kato[2] ,
Luiz Henrique Cabral Oliveira[2(✉)] , Pedro Henrique Cabral Oliveira[3] ,
Carol Brandt Alves[1] , Carolina Benetti[2] , and Renato Araujo Prates[1]

[1] Uninove/Biophotonics Applied to Health Sciences, São Paulo, Brazil
[2] UFABC/Biomedical Engineering, São Paulo, Brazil
`luiz.cabral@ufabc.edu.br`
[3] Department of Restorative Dentistry, Guarulhos University, São Paulo, Brazil

**Abstract.** Caries is a globally important public health problem that affect teeth and may also cause problems for systemic health. High power lasers have been studied as a adjuvant therapy for caries prevention and remineralization. Furthermore, OCT has been proposed as a diagnostic tool for dental caries due to changes in the optical properties of enamel after demineralization. The aim of this study was to evaluate the effects of diode laser irradiation on enamel remineralization. Bovine enamel blocks were prepared and submitted to caries lesion formation using an *in vitro* microbial model. Samples were incubated with *Streptococcus mutans* suspended in a medium supplemented with sucrose during 4 days and were additionally exposed to 10% sucrose 10 times/day. Then, caries lesion were irradiated with a 808 nm diode laser (1 W, irradiance of 35.7 J/cm$^2$, during 10 s) and were submitted to remineralization process for 7 days. A commercial Optical Coherence Tomography system, with wavelength of 930 nm, was used to analyse the samples after caries formation and after remineralization process. OCT detected enamel mineral loss promoted by caries lesion formation. After induction of remineralization, samples without treatment did not presented alteration in reflectivity data suggesting no increase in mineral content. However, irradiated samples showed a significant decrease in reflectivity results ($P < 0.05$). Our data showed that diode laser irradiation successfully induce enamel remineralization.

**Keywords:** microbial model · dental caries · semiconductor laser · OCT

## 1 Introduction

Caries is an important public health problem with significant influence on the quality of life, and implications for systemic health. About 35% of the world population, approximately 2.4 billion people, are affected by this disease. Untreated caries lesions in permanent dentition are the most prevalent chronic oral disease observed in the Global Burden of Disease Study [1, 2]. This disease presents significant numbers in different age groups, in some countries the prevalence of dental caries varies from 27 to 64% in children aged 12 years and adults up to 26 years 83% [3, 4].

Caries development depends on the presence of cariogenic bacteria adhered to teeth and sugar to cause a progressive loss of the mineral content. When fermentable carbohydrates are ingested, it is metabolized by bacteria resulting in acid production which is able to demineralize tooth structure. Over time, repeated events of mineral loss leads to a subsurface lesion formation, and progression may lead to cavitation [4].

High power lasers have been proposed as a adjuvant therapy for caries prevention. Due to its thermal effect, an increase of the acid resistance of the enamel can be achieved by denaturation of the organic content, carbonate loss, and alteration of inorganic constituents [5]. Recently, lasers have also been investigated for enamel remineralization [6]. High power near infrared diode lasers are compact, portable and low cost compared to other lasers, and have become popular in clinical practice.

Optical coherence tomography (OCT) is a tomographic imaging technique based on interferometry with low coherence light. It was developed in the 90's to obtain high-resolution images of biological tissues in a non-invasive, non-destructive, painless and real-time manner. The basic principle of the technique is based on the interaction of light with the tissues. The light coming from an optical source is divided into two beams; one of them is led to a reference mirror and the other to the sample. During light-tissue interaction, a part of incident light is backscattered by subsurface layers of the tissue and detected [7].

In the imaging process, the OCT shows values proportional to the amount of backscattered light as a function of depth, i.e., a longitudinal scan of the sample, which is called A-scan. By shifting the light beam passing through a line, it is possible to obtain a set of data of a cross section, which is called a B-scan. Near-infrared OCTs are generally used for biological tissues because the light in this spectral region suffers minimal absorption and moderate scattering, allowing resolution to a depth of typically 3 mm [7, 8].

The OCT technique has been studied as a diagnostic tool for dental caries due to changes in the optical properties of enamel after demineralization. Images obtained by OCT can be evaluated qualitatively by identifying structures, dimensions and proportions and quantitatively by analyzing reflectivity or total optical attenuation [8, 9]. Using OCT, demineralized enamel has been shown to promote a strong light backscattering [9].

Thus the aim of this study was to evaluate the effects of diode laser irradiation on enamel remineralization assessed by OCT.

## 2 Methodology

### 2.1 Experimental Design

The experimental units were composed of 20 bovine dental enamel fragments divided into 2 groups (n = 10). The factor under study was laser treatment, and the variable response was the integrated reflectivity. Caries-like lesions were induced by an in vitro microbial model. Caries lesions were irradiated with diode laser and submitted to in vitro remineralization. The samples were evaluated by OCT after the enamel demineralization and after 7 days of in vitro remineralization.

## 2.2  Sample Preparation

Freshly extracted bovine teeth were kept in a 0.1% thymol solution at 4 °C until experiment. Teeth with stains or enamel cracks were discarded. Then the samples were randomly distributed into 2 groups: control and laser groups.

All specimens had the buccal surface sequentially abraded with silicon carbide grinding papers of 150, 250, 400 and 600 grit (Buehler-São Paulo, Brazil) with water in a polisher. Enamel blocks of 7 x 4 x 4 mm were sectioned of the buccal surface with diamond discs at low rotation under refrigeration. Half of each enamel block was covered by an acid resistant nail polish (sound enamel) leaving an area to be exposed to demineralization process (caries lesion). After preparation, the samples were stored in relative humidity. Before caries induction, the samples were washed and then sterilized in an autoclave.

## 2.3  Formation of Caries Lesion

Caries lesions were formed using an in vitro microbial model using the bacteria Streptococcus mutans (ATCC 25175). Prior to bacteria incubation, teeth samples were treated with artificial saliva to prepare enamel surface for bacteria adhesion and biofilm formation [10].

Bacteria was grown in BHI broth for 48 h, at 37 °C, under 5% of CO2. After this period, enamel samples were incubated in a 24-well plate with 2 ml of S. mutans suspension (106 CFU/ml) in BHI broth supplemented with 1% sucrose [11].

During 4 days, the blocks were exposed to 10% sucrose solution 10 times a day, for 3 min (09:00, 10:00, 11:00, 12:00, 13:00, 14:00, 15:00, 16:00, 17:00, 18:00 h) [10]. The media (BHI with 1% sucrose) was replaced every 24 h. After 4 days, the samples were carefully washed, disinfected in 2% chlorhexidine for 30 min, dried at room temperature and stored at relative humidity prior to analysis.

## 2.4  Laser Irradiation

A commercial diode laser (Thera Laser Surgery-DMC-São Carlos-Brasil) with a wavelength of 808 nm was used. Laser was manually applied by scanning enamel surface with a 600 $\mu$m optical fiber, keeping a distance of 1 cm from fiber tip to sample. Samples from laser group were irradiated with an output power of 1 W, irradiance of 35.7 J/cm$^2$, during 10 s.

## 2.5  Remineralization

In order to induce remineralization, enamel samples were immerse in a remineralizing solution composed of H2O, HCL, KOH, CaCl2, (HOCH2)3CNH2, pH 7 (Fórmula&Ação, São Paulo, Brazil) for 7 days.

### 2.6   OCT Analysis

The samples were evaluated by a commercial OCT system (OCT CALLISTO, Thorlabs Inc., New Jersey, USA), operating in the wavelength of 930 nm, 1.2 kHz. Prior to scanning, specimens were removed from relative humidity, washed with distilled water gently dried with absorbent paper to remove excess of water.

OCT imagens were obtained for tested samples (from control and laser groups) and sound enamel of each enamel block before remineralization, i.e. after caries formation, and 7 days after remineralization. These images were constituted by a numerical matrix of 2000 pixels per line and 512 pixels per column. Considering the enamel refraction index of approximately 1.6 [12], we calculated a pixel resolution of 3.0 $\mu$m x 2.5 $\mu$m.

In OCT image of each sample, two regions of 150 columns (width) were selected and a graph of reflectivity intensity as a function of depth (A-scan) was plotted. The first three pixels correspond to a intense reflection due to the interface between the air and the sample and were eliminated from the analysis. The reflectivity was integrated from the surface to a depth of 150 $\mu$m. This data presents a correlation with mineral loss [8].

Then, a variation of the integrated reflectivity was calculated for tested groups (control and laser) using the following equation:

$$\Delta R = (IR_T - IR_S)/IR_S$$

Where $\Delta R$ is the variation of integrated reflectivity, $IR_T$ is the integrated reflectivity of test samples, and $IR_S$ is the integrated reflectivity of sound enamel.

### 2.7   Statistical Analysis

The data is presented as integrated reflectivity. They were analyzed to verify the assumption of normality using Shapiro–Wilk test. Comparisons between groups were made by T-test. For all tests, the significance level was set at 5%.

## 3   Results and Discussion

In OCT analysis, light interacts with matter and photons that were bounced back are detected to provide an image of the irradiated object. Dental enamel is a cristaline heterogeneous structure, and a modification of mineral content alters its interaction with light. For this reason, we examined the ability of OCT to detect changes induced by demineralization and remineralization of incipient caries-like lesions.

Figure 1 shows a OCT image of one enamel block with a sound and a caries region. In the sound enamel, a large subsurface zone in light gray indicates less signal coming from this region. In a carious portion of enamel, there is an increase of reflectivity below enamel surface, seem as white image, due to mineral loss. The signal suddenly drops, and assuming that it occurs in the lesion limit, the depth of the carious lesion was approximately 114 $\mu$m.
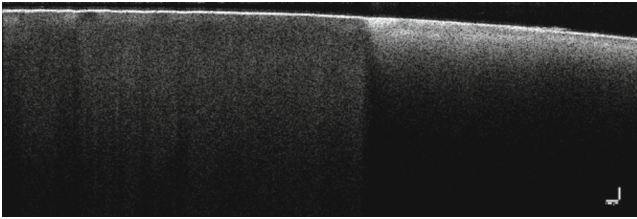
**Fig. 1.** OCT image of sound enamel (left part of the image) and enamel with caries lesion (right part of the image)(Color Figure Online)

Seven days of immersion in remineralizing solution did not alter the variation of integrated reflectivity of carious enamel samples from control group (Fig. 2, P > 0.05). On the other hand, irradiated samples presented a significant decrease of variation of integrated reflectivity after the remineralization process (P < 0.05) indicating that diode laser irradiation successfully induce enamel remineralization.
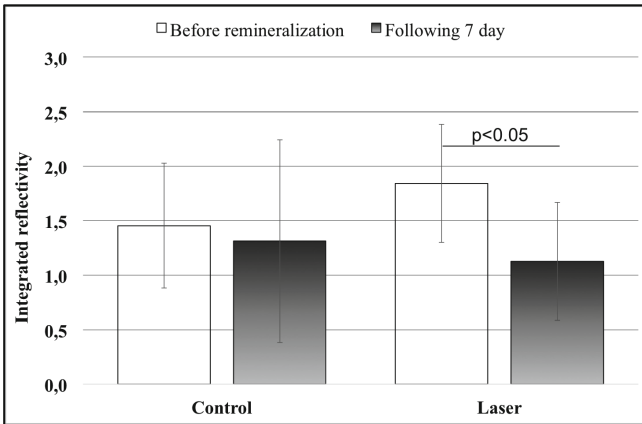


**Fig. 2.** Variation of integrated reflectivity of caries enamel samples of control and laser group before and after remineralization

The use of diode laser is very popular in current dentistry, specially for surgical procedures. A 808 nm diode laser did not present absorption by major enamel constituents, i.e. water and hydroxyapatite, to induce physical modification of enamel structure in a similar manner of Er, Cr:YSGG and $CO_2$ lasers [5]. Nevertheless, the effect observed in our study can be related to denaturation of the organic components, which is achieved at lower temperature compared to alteration of inorganic content [5]. Moreover, infrared diode laser associated to fluoride agents were also reported to enhance enamel remineralization [6].

Enamel image acquisition by OCT is rapid and non destructive, permitting repeated analysis of a sample. In the present study, OCT was applied for quantitative analysis of mineral loss and may be used for monitoring lesion remineralization. We observed that the high power diode laser successfully induce enamel remineralization. Further studies

are been conducting to investigate therapies association and longer remineralization period.

## 4   Conclusion

Our results indicate that irradiation of enamel caries lesion with 808 nm diode laser induced remineralization. Analysis of OCT images using integrated reflectivity was able to quantify enamel mineral changes.

**Conflict of Interest.**   The authors declare that they have no conflict of interest.

## References

1. GBD: Disease and Injury Incidence and Prevalence Collaborators (2016) Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet **388**(10053), 1545–1602 (2015)
2. Mathur, V.P., Dhillon, J.K.: Dental caries: a disease which needs attention. Indian J. Pediatr. **85**(3), 202–206 (2018)
3. Benjamin, R.M.: Oral health: the silent epidemic. Public Health Rep. **125**, 158–159 (2010)
4. Peres, M.A., Macpherson, L.M.D., Weyant, R.J., et al.: Oral diseases: a global public health challenge. Lancet **394**, 249–260 (2019)
5. Al-Maliky, M.A., Frentzen, M., Meister, J.: Laser-assisted prevention of enamel caries: a 10-year review of the literature. Lasers Med. Sci. **35**(1), 13–30 (2020)
6. Xue, V.W., Yin, I.X., Niu, J.Y., et al.: Combined effects of topical fluorides and semiconductor lasers on prevention of enamel caries: a systematic review and meta-analysis. Photobiomodul Photomed. Laser Surg. **40**(6), 378–386 (2022)
7. Monteiro, G.Q., Montes, M.A., Gomes, A.S., et al.: Marginal analysis of resin composite restorative systems using optical coherence tomography. Dent. Mater. **27**, 213–223 (2011)
8. Le, M.H., Darling, C.L., Fried, D.: Automated analysis of lesion depth and integrated reflectivity in PS-OCT scans of tooth demineralization. Lasers Surg. Med. **42**(1), 62–68 (2010)
9. Maia, A.M., de Freitas, A.Z., de L Campello S, et al.: Evaluation of dental enamel caries assessment using quantitative light induced fluorescence and optical coherence tomography. J. Biophotonics **9**(6), 596–602 (2016)
10. Maske, T.T., van de Sande, F.H., Arthur, R.A., et al.: In vitro biofilm models to study dental caries: a systematic review. Biofouling **33**(8), 661–675 (2017)
11. Hayati, F., Okada, A., Kitasako, Y., et al.: An artificial biofilm induced secondary caries model for in vitro studies. Aust. Dent. J. **56**(1), 40–47 (2011)
12. Freitas, A.Z., Zezell, D.M., Vieira, N.D., et al.: Imaging carious human dental tissue with optical coherence tomography. J. Appl. Phys. **99**(2), 024906 (2006)

# Acute Effect of Photobiomodulation with LED in Apparatus on Power in Hip Muscles in Classical Ballerinas

Mariana A. N. Duque[(✉)] [ID], Bruno H. Godoi [ID], Sergio L. Lemos [ID],
Elessandro V. Lima [ID], Fernanda M. G. Gonzaga [ID], and Juliana Ferreira-Strixino [ID]

Photobiology Applied to Health Lab (PhotoBioS), Research and Development Institute (IP&D),
University of Vale do Paraíba (UNIVAP), São José dos Campos-SP, Brazil
mariana.duque@univap.br

**Abstract.** Classical Ballet generates mechanical stress on the musculoskeletal structure, with the hip joint being one of the most requested. This study aimed to evaluate the acute effect of photobiomodulation on the hip abductor muscles in classical dancers using a Light Emitting Diode (LED) device at a wavelength of 850 nm (infrared). Twenty-three classical dancers aged 18 and 30 participated in two days: the control group (CG), without LED irradiation, and the Infrared Group (GI), with one week between collections. The participants evaluated the power in the concentric phase of the abductor and adductor muscles in the Isokinetic Dynamometer, performing 05 repetitions of abduction and adduction of the hip with an amplitude of 30 degrees and angular velocity of 120°/s. Data processing using Matlab software showed significant improvement ($p < 0.05$) in the power of the abductor muscles of the right and left lower limbs with Infrared LED intervention. Most of the literary findings did not recommend a total delivered energy of 17.5 J and an energy density of 4.46 J/cm$^2$, suggesting the reason why there was no increase in power also in the hip adductors. However, research has shown results effective with a density of 5 J/cm$^2$ but in muscle groups with a smaller area. It was concluded that even with a significant result with infrared LED on muscle power, further studies are necessary with variation in parameters for irradiation, such as the area of coverage of the LED apparatus or an increase in photobiomodulation sessions.

**Keywords:** Hip lesions · Ballet · Low intensity light therapy · Photobiomodulation

## 1 Introduction

The activities developed in Classical Ballet highly demand the lower limbs, and injuries to the hip, knee, and ankle affecting cartilage, ligaments, tendons, or bursa can lead dancers to withdraw from their activities [1].

About 17.2% of injuries in dancers occur in the hip [2] due to the extreme level of activity to which they are subjected, incorrect injury diagnosis, or incomplete recovery [3].

It is necessary to analyze the strength of the hip abductor muscles of dancers, as it is considered relevant for the prevention and recovery of the hip, knee, or ankle injuries since Classical Ballet is an activity that requires excellent mechanical effort. In addition, it suffers an aggravation when considering the use of pointe shoes by the female gender, which completely alters the distribution of weight and the mechanics of movement and gait [4].

Intervention with photobiomodulation can help muscle recovery, increase strength and consequently prevent injuries [5, 6], an essential factor for this public. The Light Emitting Diode (LED) is a light source with a photobiomodulation action with positive effects that can enhance muscle function and improve exercise performance.

This preliminary study describes part of the development and validation of the prototype of apparatus for irradiation, which aimed to evaluate the acute effect of photobiomodulation on the abductor and adductor muscles of the hip using LED in the region of 850 nm.

## 2 Materials and Methods

### 2.1 Materials

The Biodex Multi-joint System 3 Isokinetic Dynamometer from BIODEX MEDICAL SYSTEMS INC was used to collect data on the power and total work of the hip abductor and adductor muscles. In addition, to confirm the hip's abduction and adduction angle, a Goniometer was used, which confirmed the angular amplitude electronically established by the equipment.

For photobiomodulation intervention, a LED apparatus was developed, with dimensions of 30 cm high by 60 cm long, with a total of 20 LEDs, controlled by a microcontroller under the registered software BR512021001498-7, which required the use of protective eyewear for the researcher and participants to reduce eye risks due to infrared irradiation. Furthermore, the time of application of photobiomodulation was controlled by the stopwatch of the cell phone (model Samsung S20 Plus).

The apparatus has two main programs, irradiation with 660 nm LEDs (Red) and irradiation with 850 nm (Infrared). The equipment calibration was carried out in partnership with the company Biopdi-Equipment for Testing Materials, which provided access to the Equipment Console of the PM100D Energy Meter coupled to the ThorLabs S350C sensor.

Measurements were obtained from the quantification of LEDs individually. The fluence was calculated for each LED independently since the irradiation points are independent, and there is contact with the irradiated location. The parameters of the Infrared LED used were: power - 3 W; voltage - 1.6 to 2.0 V; current – 700 mA; wavelength - 850 nm $\pm$ 5; Infrared LED number – 10 LED (Table 1).

**Table 1.** Properties of the LEDs apparatus

| LED Power | 3 W |
|---|---|
| Wavelength | 850 nm $\pm$ 5 |
| Photobiomodulation Apparatus Power | 30 W |
| Energy Density of Each LED | 2 J/cm$^2$ |
| Total Energy | 17.5 J |
| Spot Size of the LED | 1 cm$^2$ |
| Irradiation Time | 10 s |

The Infrared LED was applied for 10 s, delivering 17.5 J of total energy. The choice to work with an energy density close to 20 J/cm$^2$ is due to literary findings on the applicability of photobiomodulation to improve muscle performance [7]. On the power set display, it is possible to set the brightness, which at the same time changes the delivered power density.

**Methods.** The study was characterized as a descriptive cross-sectional analysis of the power of the Hip Abductor Muscles of Classical Ballet practitioners.

This research protocol was submitted for analysis by the Research Ethics Committee (CEP) by Plataforma Brasil and was approved under the CAAE protocol: 40119520.3.0000.5503, being conducted per resolution nº 466/2012 of the National Health Council (CNS). All participants were previously informed and oriented about the procedures to which they would be submitted. After full acceptance, they signed the Free and Informed Consent Term, being also instructed that they could stop participating in the research anytime and without any burden.

This research protocol was carried out at the Biodynamics and Photobiology Applied to Health Laboratories (PHOTOBIOS), belonging to the Research and Development Institute (IP&D) of the University of Vale do Paraíba (Univap).

Classical Ballerinas aged 18 and 30 years were recruited. The choice of age is justified by the changes that occur in the body over the years due to the aging process, such as the reduction in the cross-section of the muscle, with more significant intramuscular fat and collagen content; decreased muscle strength; loss of mineral salts; among others [8–10].

Twenty-three women from Cia de Dança and ballet academies in São José dos Campos – SP, Brazil, participated in this study. The research participants were submitted to two collection moments on two different days, with a one-week interval between them. One collection was considered a Control Group (CG), with the protocol performed without LED application, and one was considered a Group with Infrared LED Application (GI).

*Procedures.* Measurements were performed by the same researcher who instructed the participants not to practice physical exercise on the day of data collection. A specific warm-up was performed for the abductor muscles (Fig. 1), consisting of 25 repetitions of standing abduction with a miniband elastic. The specific warm-up was used with a comfortable number of repetitions and intensity, and movement and amplitude similar

**Fig. 1.** Specific physical warm-up using a mini band

to what would be performed in the test [11], thus not exerting influence on the acquired results, only fulfilling the objective of preparation for the main exercise, preventing possible injuries [12].

After warming up, for data recording, they were positioned on the examination table in lateral decubitus, with the limb to be tested facing up, trunk, and contralateral lower

limb secured by a belt. The dynamometer's axis was aligned with the intersection of two straight lines: one medial to the Greater Trochanter of the Femur and the lever arm medial to the patella. The limb to be tested was set at 0°, referring to the participant's hip-width (Fig. 2) measured with the goniometer. From there, a range of 30° (Fig. 3) of hip abduction was applied, calculated by the dynamometer. Although initially, the test was performed on the dominant and contralateral sides to instruct and adapt the patient to the equipment, the first evaluation may not reflect the individual's natural capacity, which is called the learning effect [13]. Thus, the first test was excluded but used as a specific warm-up.
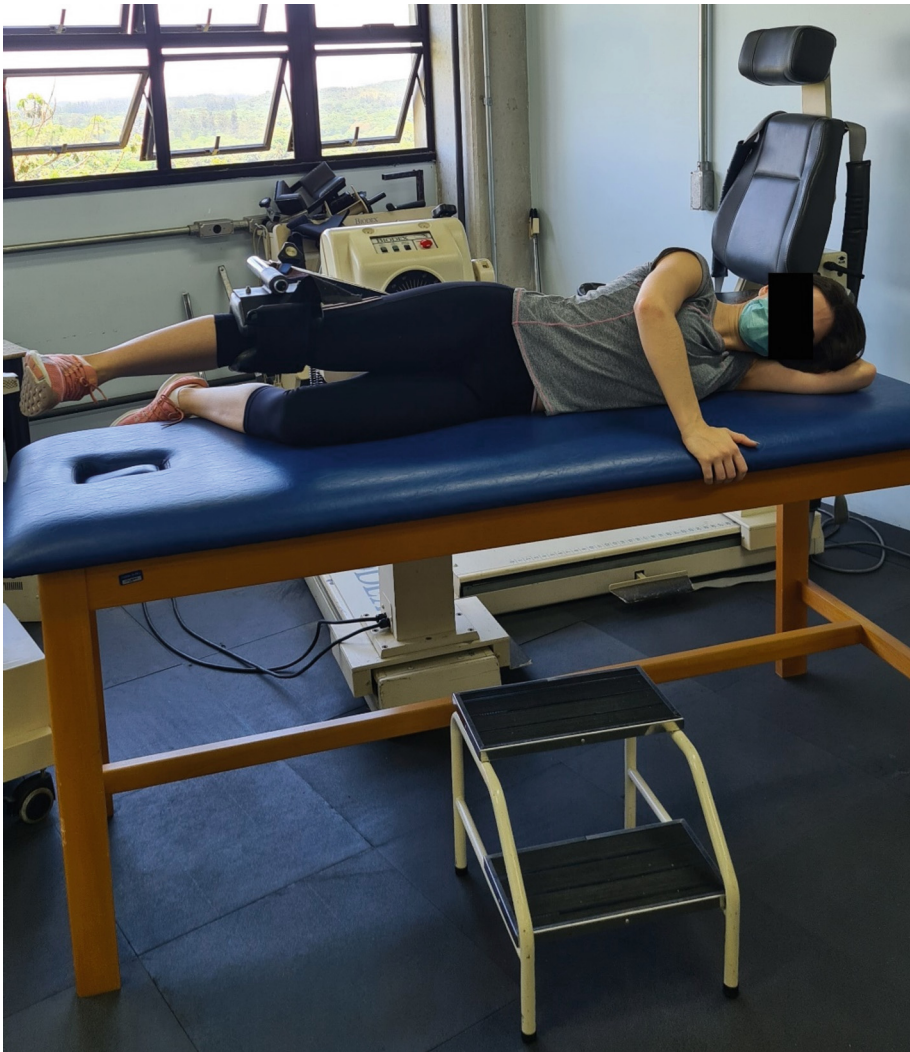


**Fig. 2.** Initial angle of hip abduction
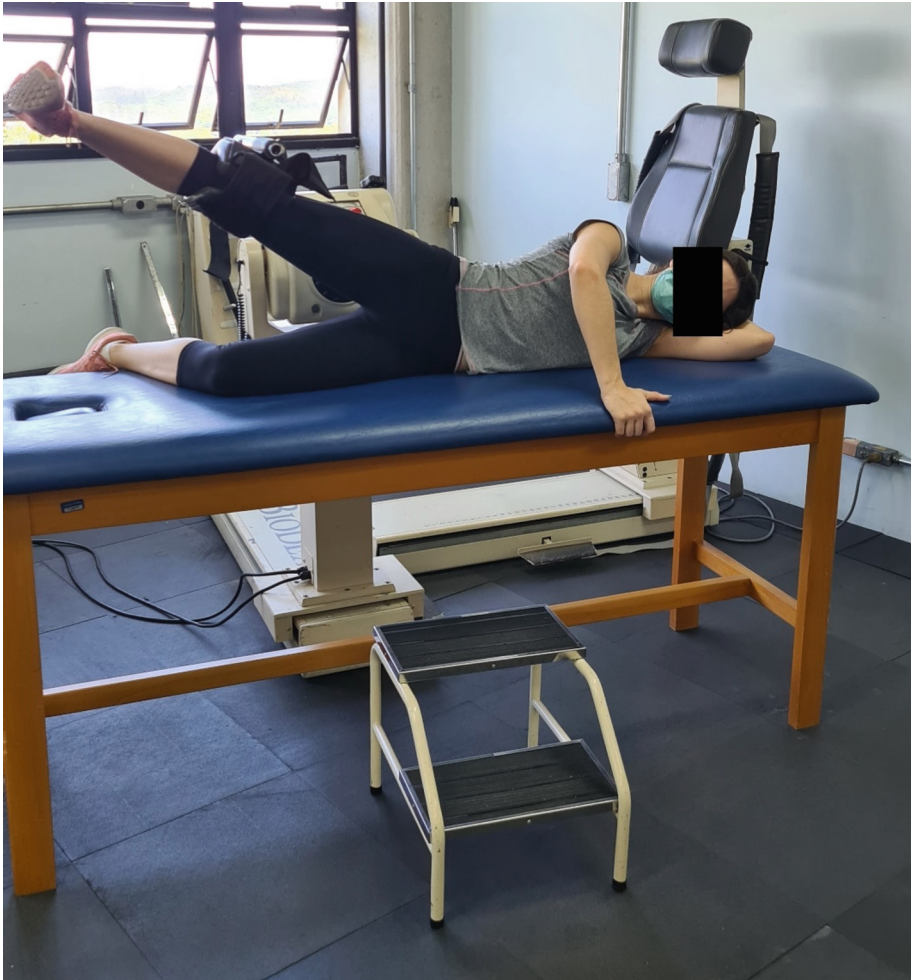
**Fig. 3.** Final angle (30°) of hip abduction

Then, the intervention with LED was performed (Fig. 4) or a 2-min rest in the case of the session without its application, followed by the test with 05 repetitions at an angular speed of 120°/s [14]. The repetitions reproduce the hip abduction and adduction movement, with the limb fixed to the Dynamometer equipment. The Dynamometer software recorded the concentric phase of hip abduction and adduction movements. The LED blanket irradiated 10 points distributed over the area of the Abductor, and Adductor Muscles of the Hip, with an Infrared, LED delivering 17.5 J of total energy for 10 s at a wavelength of 850 nm.

The data obtained in the collections with the isokinetic dynamometer in the txt format were transported in spreadsheets in Microsoft Excel. Later the normalization and the statistical treatment were carried out with the help of the MATLAB software. The analyzes for Potency verified the normality of the data by the Shapiro-Wilk test, and the unpaired one-way ANOVA test was applied to standard data and confirmed by the Holm-Sidak test. Finally, the p-value was considered for the significance level $<0.1$.



**Fig. 4.** Lower member overlapping the LED apparatus

## 3   Results

The results showed a significant difference ($p < 0.1$) for the Power in the hip abductors with the application of infrared LED (GI) when compared to the Control Group (CG), without the application of LED (Fig. 5), in the execution of the lower limb left side (PABD-e).

In the comparison of Power in the hip adductors, there was a significant difference (P $> 0.1$) between the control group (CG) and the group irradiated with infrared LED (GI), both in the execution of the lower limb left side and for the right lower limb (PABD-e and PABD-d).
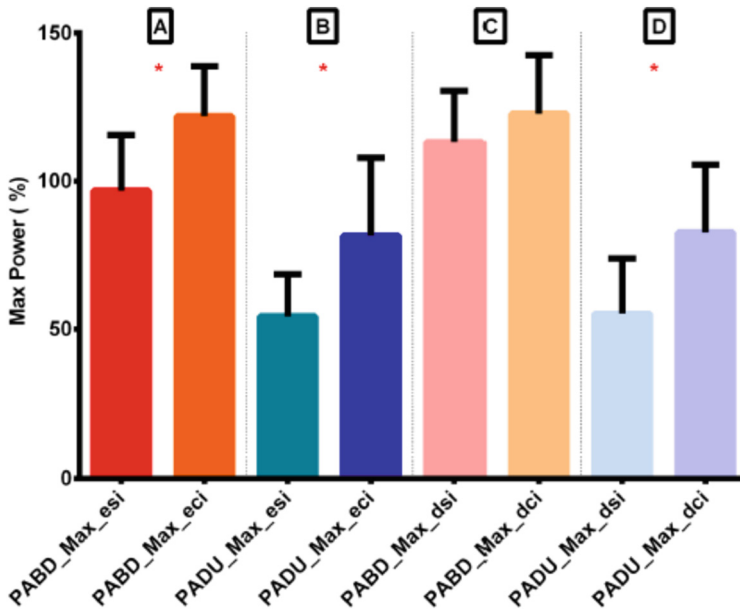
**Fig. 5.** Power analysis between group without irradiation and with Infrared LED irradiation **Subtitle. A** - abduction without LED versus with LED on the left side; **B** - adduction without LED versus with LED on the left side; **C** – abduction without LED versus with LED on the right side; **D** - adduction without LED versus with LED on the right side; **PABD_Max** – maximum abduction power; **PADU_Max** - maximum adduction power; esi – left without irradiation; eci - left with irradiation; dsi - right without irradiation; dci - right with irradiation; *Significant Difference p < 0.1.

## 4   Discussion

Muscular power is the product of strength by speed over time, that is, the ability to produce strength quickly, being one of the most critical physical valences for sports, such as Classical Ballet [15, 16].

Studies to verify the influence of photobiomodulation using laser or LED on the performance of the musculoskeletal system, such as muscle power, are frequent. Some of these studies showed significant results in athletic performance or physiological responses at the cellular level [17–19]. However, others did not show any benefit in muscle performance in their results [20–22]. Divergent results may occur due to manipulating the different methodological variables (type of equipment used, irradiation density applied, research sample, gender, age group, training level, cell type used).

Irradiated cells absorb light through the respiratory chain, activating the Cytochrome C Oxidase enzyme in the mitochondria and cytoplasmic membrane, triggering various effects after light absorption, such as increased concentration of adenosine triphosphate (ATP). Given the fundamental role of mitochondria in ATP production, one of the effects is its increase and modulations in the rates of DNA and RNA synthesis, which affect cell proliferation and gene expression of various cellular pathways (mitosis, apoptosis,

inflammation, and cell metabolism, mitochondrial energy), which may be related to the benefits of treatment with photobiomodulation in skeletal muscle during exercise and also in regenerative actions [5].

The few studies on the effect of photobiomodulation on hip abduction action used an energy density of 20 J/cm$^2$ for the analyzed area [7], higher than the one used in this study (4.6 J/cm$^2$, approximately four times lower density energy, or 17.5 J of total area). Still, a significant difference was found in the study.

According to Alves, Furlan, and Motta [7], the energy density delivered for irradiation of the biceps, rectus femoris, and soleus muscle, with significant results, was 5 J/cm$^2$ or 20 J of total energy. These results corroborate those found in this study, with energy density similar to that applied to the abductor muscles (4.46 J/cm$^2$ for infrared LED).

The significant results in the research mentioned above were obtained through photobiomodulation in only one muscle (biceps, rectus femoris, or soleus) since this research used similar energy density to obtain the same results in 8 muscles in the case of the abductors (larger area irradiated by the LED blanket). The same authors also cite different irradiation densities, such as 60 J of total energy for the biceps muscle, which had a significant result with 20 J of total energy. Therefore, it seems necessary to recalculate the energy density variables to higher values to obtain positive torque and total work results, not analyzed in this work, for the muscles responsible for hip abduction and significant results for the hip adductors. They were perhaps using higher energy density since the preliminary results need a statistical treatment and a greater probability of significance (p-value) to understand the effects of irradiation.

## 5   Conclusions

The results showed a significant difference in the increase in Muscle Power of the hip abductors on the left side and the adduction movement on both sides with Infrared LED irradiation compared to the condition without irradiation. Therefore, further research should involve the hip joint using different Infrared LED irradiation densities.

**Conflict of Interest.**  The authors declare that they have no conflict of interest.

## References

1. Nilsson, C.: The injury panorama in a Swedish professional ballet company. Knee Surg. Sports Traumatol. Arthrosc. **57**, 241–246 (2001). https://doi.org/10.1007/s001670100195
2. Trentacosta, N., Sugimoto, D., Micheli, L.J.: Hip and groin injuries in dancers: a systematic review. Sports Health **9**, 422–427 (2017). https://doi.org/10.1177/1941738117724159
3. Reid, D.C.: Prevention of hip and knee injuries in ballet dancers. Sports Med. **6**, 295–307 (1988). https://doi.org/10.2165/00007256-198806050-00005
4. Allen, N., et al.: Ballet injuries: injury incidence and severity over 1 year. J. Orthop. Sports Phys. Ther. **42**, 781–790 (2012). https://doi.org/10.2519/jospt.2012.3893

5. Tsuk, S., et al.: Does photobiomodulation therapy enhance maximal muscle strength and muscle recovery? J. Hum. Kinet. **73**, 135–144 (2020). https://doi.org/10.2478/hukin-2019-0138

6. Almeida, J.N., et al.: Effects of photobiomodulation on muscle strength in post-menopausal women submitted to a resistance training program. Lasers Med. Sci. **35**, 355–363 (2020). https://doi.org/10.1007/s10103-019-02822-4

7. Alves, V.M.N., Furlan, R.M.M.M., Motta, A.R.: Immediate effects of photobiomodulation with low-level laser therapy on muscle performance: an integrative literature review. Rev. Cefac. **21**, 1–9 (2019). https://doi.org/10.1590/1982-0216/201921412019

8. Montopoli, M., et al.: Aging and osteoarticular system for healthy aging on behalf of integrative medicine research group (IMRG). Eur. Rev. Med. Pharmacol. Sci. 2348–2360 (2021). https://doi.org/10.26355/eurrev_202103_25274

9. Weineck, J.: Biologia do Esporte, Manole, São Paulo (1991)

10. Gallahue, D.L., Ozmun, J.C.: Understanding Motor Development: Infants, Children, Adolescents, Adults: Infants, Children, Adolescents, Adults, 8th edn. Jones & Bartlett Publishers, Boston (2005)

11. Simão, R., et al.: Influence of different protocols of warm-up for the capacity of developing maximum load on the 1RM test. Fit Perf. J. **3**, 263–265 (2004). https://doi.org/10.3900/fpj.3.5.261.e

12. Fermino, R.C., et al.: Influence of the specific warm-up and stretching in the performance of the muscular force in 10 maximum repetitions. Rev. Bras. Cienc. e Mov. **13**, 25–32 (2005)

13. Fillipin, N.T., Vieira, W.H.B., Costa, P.H.L.: Repetibilidade de medidas isocinéticas em quadril. Rev. Bras. Educ. Fís. Esp. **20**, 131–139 (2006)

14. Dvir, Z.: Isokinetics: Muscle Testing, Interpretation and Clinical Applications. Churchill Livingstone, New York (1995)

15. Macaluso, A., De Vito, G.: Muscle strength, power and adaptations to resistance training in older people. Eur. J. Appl. Physiol. **91**, 450–472 (2004). https://doi.org/10.1007/s00421-003-0991-3

16. Kraemer, W.J., et al.: Strength and power training: physiological mechanisms of adaptation. Exerc. Sport Sci. Rev. **24**, 363–397 (1996). https://doi.org/10.1249/00003677-199600240-00014

17. Lima, E.V., Pacheco-Soares, C., Silva, N.S.: Photobiomodulation assay of muscle cells C2C12 after irradiation with LED device. Res. Soc. Dev. **11**, 1–11 (2022). https://doi.org/10.33448/rsd-v11i6.28884

18. Toma, R.L., et al.: Photobiomodulation (PBM) therapy at 904 nm mitigates effects of exercise-induced skeletal muscle fatigue in young women. Lasers Med. Sci. **31**, 1473–1480 (2016). https://doi.org/10.1007/s10103-018-2454-4

19. Vassão, P.G., et al.: Effects of photobiomodulation on the fatigue level in elderly women: an isokinetic dynamometry evaluation. Lasers Med. Sci. **31**, 275–282 (2015). https://doi.org/10.1007/s10103-015-1858-7

20. Azuma, R.H.E., et al.: Photobiomodulation therapy at 808 nm does not improve biceps brachii performance to exhaustion and delayed-onset muscle soreness in young adult women: a randomized, controlled. Crossover Trial. Front. Physiol. **12**, 1–9 (2021). https://doi.org/10.3389/fphys.2021.664582

21. Malta, E.S., et al.: Acute LED irradiation does not change the anaerobic capacity and time to exhaustion during a high-intensity running effort: a double-blind, crossover, and placebo-controlled study. Lasers Med. Sci. **31**, 1473–1480 (2016). https://doi.org/10.1007/s10103-016-2011-y

22. Dutra, Y.M., et al.: Acute photobiomodulation by LED does not alter muscle fatigue and cycling performance. Med. Sci. Sports Exerc. 2448–2458 (2020). https://doi.org/10.1249/MSS.0000000000002394

# Proposal of a Computerized System Based on Gametherapy to Reduce Postural Sway

Luan de Almeida Moura[(✉)] , Paulo César dos Reis,
André Roberto Fenandes da Silva , Rodrigo Parra do Prado,
Mariana da Palma Valério , Silvia Regina Matos da Silva Boschi ,
Silvia Cristina Martini , Terigi Augusto Scardovelli ,
and Alessandro Pereira da Silva

University of Mogi das Cruzes, Mogi das Cruzes, SP 08780-911, Brazil
luanm213@gmail.com

**Abstract.** For balance training, one of the methods used is gametherapy, which consists in the use of games for a better motivation of the patients. Thus, the objective of this work was to develop three games for balance training that explore laterolateral and anteroposterior movements, as well as the combination of both. The games were developed and tested with a force platform with controlled inclination, which is used to control the game avatars. To test the platform's functionality, weights were placed on the equipment, and the response of the sensors was verified. The games were evaluated by two distinct groups of professionals: the first composed of 5 system/game developers, and the second by 5 physiotherapists. To this end, two questionnaires were developed based on regulatory standards for software development and quality and on the System Usability Scale (SUS). In the functionality test, the sensors responded by increasing their values according to the region of the platform being pressed. The responses obtained in the SUS questionnaire presented a score of 81 in the evaluation of the physiotherapists and 83 in the evaluation of the system developers. Both values classify the system as good, according to the SUS scale. Games can be a very useful tool with great potential for use in balance training and rehabilitation. The developed system may bring a significant improvement in treatments, as well as help the health professionals by increasing the possibilities of applied techniques, thus reaching a longer permanence of the patient during treatment, with the motivation to use the games.

**Keywords:** Balance · Software tests · Gametherapy · Unity

## 1 Introduction

Postural control can be defined as a complex motor skill arising from the interaction of multiple sensorimotor processes [1]. The neuromuscular responses aim to ensure that the body's center of gravity (CG) is maintained within the support base (area formed by the lower limbs) [2, 3]. Thus, over time, postural training methods have been developed aiming at improving orthostatic balance [4]. For balance training, unstable platforms

such as the wobble board are commonly used [5]. Either by strengthening the lower limb stabilizing muscles, by improving proprioception, or by improving the muscle activation capacity [6].

One of the resources to improve balance training is the implementation of virtual reality. This system is capable of simulating balance training activities, increasing mobility and reducing the fear of falls [7]. On this way, [8, 9] define that balance training based on computerized games can improve patient motivation, raising the training capacity.

In the [10] study a wii balance board was used to control the games. However, on the wii balance board the patient cannot be tested in tilting situations, which is more indicated [9]. In another study [11] a Kinect controlled game was developed along with a force platform. However, in this work the force platform is only used for data collection and not for controlling the game. Other authors [9] made a manually tilted platform with three games. However, the tilt of this platform is manual and limited to angles multiple of 5.

Therefore, a game system controlled by a force platform with biaxial rotation could better adapt to the exercises and the needs of each patient. This way, the training can be done in several angles, generating a personalization with different degrees of difficulty. Thus, the objective of this work was to develop a gametherapy proposal based on the application of three virtual games, where each one of them allows the training of a different type of movement, being the first focused on the laterolateral movement, the second focused on the anteroposterior movement, and the third focused on the combination of both movements. The games developed have an interaction interface through the movement of the player on a force platform.

## 2 Material and Methods

### 2.1 Equipment

To control the avatars of the developed games, a biaxial force platform with controlled inclination is used [12]. It has the ability to change its inclination through four step motors controlled by a microcontroller, allowing its use in various inclinations (Fig. 1a). This platform has eight load cells fixed on its surface (Fig. 1b). Its sensors allow the movement or control of the game as soon as they are pressed.

In order to test the functionality of the force platform, we first used two alters with eight kilograms each. They were placed in the front region of the platform, one between sensors 1 and 2, and the other between sensors 5 and 6. Then it was observed if these sensors increased their value read by the microcontroller's analog ports, in order to test the avatar's response, either moving forward or upward depending on the game to be played. The same was done by placing the loads between sensors 3 and 4, 7 and 8. For the right side, the loads were placed between the four sensors, being 5, 6, 7 and 8. Finally, the same test was performed for the left side with the loads between sensors 1, 2, 3 and 4, thus registering their values. By increasing the values of each sensor depending on the region pressed, it is possible to determine the control of the avatars. Through this, the patient can be trained by inducing him/her to perform laterolateral or anteroposterior movements. The following flowchart shows how the avatar control is performed (Fig. 2).
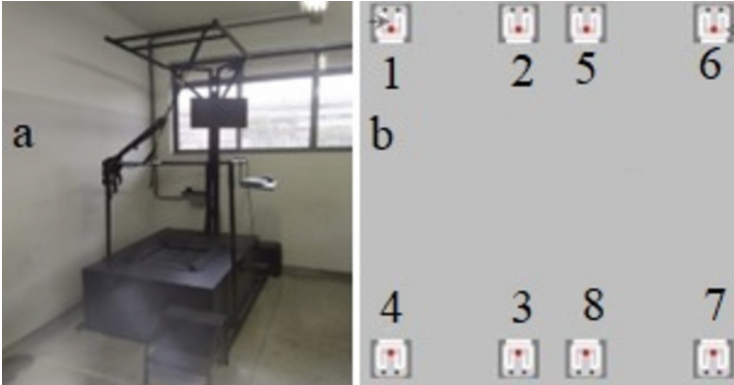
**Fig. 1.** a) Biaxial force platform with controlled inclination; b) Positioning of the load cells that allows the control of the avatars.
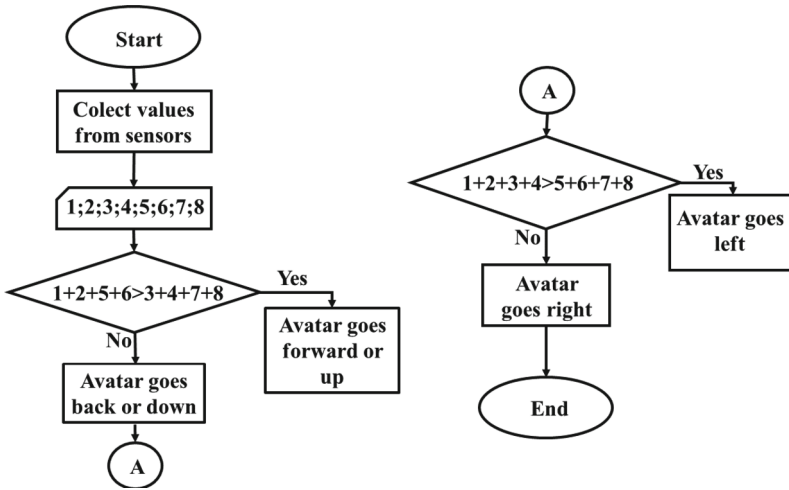


**Fig. 2.** Flowchart showing the difference between the sensors for moving avatars.

## 2.2 Game Development

For this study, the Unity multiplatform 3D game engine was used. Through this tool, three games were developed, each responsible for the training of different movements: Balance Block Breaker (Fig. 3a), for training laterolateral movements; Balance Bird (Fig. 3b), for training anteroposterior movements; and Balance Maze (Fig. 3c), for training laterolateral and anteroposterior movements.

## 2.3 Game 1 – Balance Block Breaker

In Balance Block Breaker, the objective is to bounce a ball off a series of blocks located at the top of the screen using a board at the bottom. The board is controlled by a laterolateral movement on a force platform.
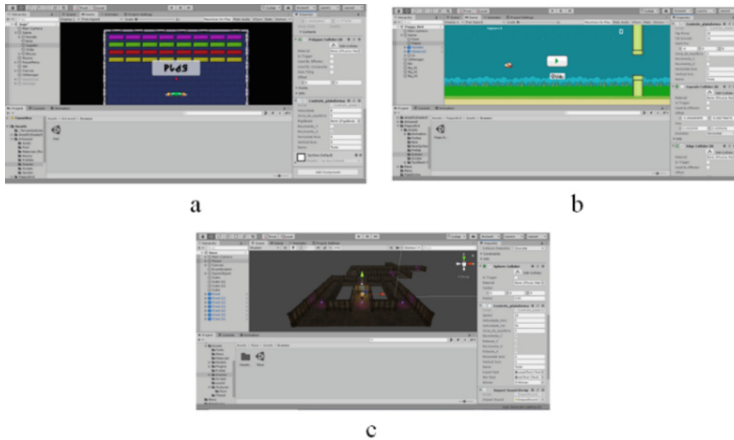
**Fig. 3.** Unity game development screens; a) Balance Block Breaker game development screen; b) Balance Bird game development screen; c) Balance Maze game development screen

### 2.4 Game 2 – Balance Bird

The goal of Balance Bird is to control a bird that moves at a constant speed, in order to pass through obstacles located in front of it, the more obstacles it passes, the more points it receives. The control is done through anteroposterior movement.

### 2.5 Game 3 – Balance Maze

In the game Balance Maze, an avatar in the ball form is controlled on a 3D plane inside a maze. The goal of the game is to collect crystals scattered throughout the labyrinth, for each crystal collected, the player gets one point. When all available crystals are collected, the game is finished and the score and game time are recorded in the player's profile. The movements of the ball are free for all axes, so the player must combine anteroposterior and laterolateral movements to control the ball and complete the maze.

### 2.6 Game Validation

According to the literature [13], to test the quality of a computerized system 5 evaluators are necessary, because about 85% of the errors are listed and after this number the same problems will tend to be observed. In this sense, to evaluate the quality of the system proposed in this study, 10 evaluators were invited, divided into 2 groups according to their backgrounds. Group 1 was composed of 5 physiotherapists, aiming at evaluating whether the system proposes movements that may contribute to balance improvement. Group 2 was composed of 5 developers aiming to evaluate the system's usability. This multidisciplinary approach aimed to obtain opinions from several angles of the system application.

This evaluation was done using specific questionnaires based on the training of the evaluators. Following the NBR ISO/IEC 9126-1 [14] to validate the games, the

principles of using questionnaires proposed by John Brooke [15], called SUS (System Usability Scale), were adopted. SUS presents excellent consistency of results for tests with relatively small sample sizes [16].

However, due to the difficulty of performing the tests in person, we recorded a video of the system working for each game. For this, four different views of the force platform were filmed (Fig. 4a). Camera 1 filmed the front view (Fig. 4b), camera 2 filmed the side view (Fig. 4c), camera 3 filmed the oblique view (Fig. 4d), and camera 4 filmed a rear view (Fig. 4e). The videos were then edited to present all four views at the same time and sent to the experts along with the questionnaire.
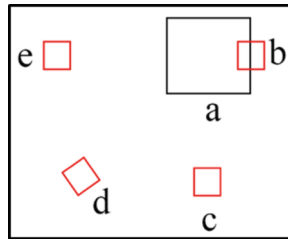


**Fig. 4.** Concept map of feature recording. a) Biaxial force platform; b) Frontal camera; c) Side camera; d) Oblique camera; e) Rear camera.

## 3   Results

The developed games establish a serial communication with the biaxial force platform. When starting the game, the platform is tilted to the desired angle, which can be between 0 and 20 degrees. This tilt can be forward, backward, right, or left. Through this, balance training is conducted with games.

The avatars are controlled according to where the force platform is pressed. The Table 1 shows the values of the admensional reading of the microcontroller's analog ports, obtained in the platform functionality test.

**Table 1.**  Value of the sensors when pressed

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |  |
|---|---|---|---|---|---|---|---|---|---|
| No load | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Avatar standing still |
| load between 1;2;3;4 | 2 | 2 | 0 | 0 | 2 | 2 | 0 | 0 | Avatar goes left |
| load between 5;6;7;8 | 0 | 0 | 2 | 2 | 0 | 0 | 2 | 2 | Avatar goes to the right |
| load between 1 and 2; 5 and 6 | 0 | 0 | 0 | 0 | 2 | 2 | 3 | 3 | Avatar goes up or forward |
| load between 3 and 4; 7 and 8 | 2 | 2 | 2 | 4 | 0 | 0 | 0 | 0 | Avatar goes down or goes back |

In the game Balance Block Breaker (Fig. 5a), for the control of its avatar, the movement in the X axis is considered, inducing the player's movement in laterolateral. In the

Balance Bird game (Fig. 5b), for the control of its avatar, the movement in the Y axis is considered, inducing the player's movement in anteroposterior. In the Balance Maze game (Fig. 5c), to control the avatar, both axes are combined, leaving the player more free, combining anteroposterior and laterolateral training.
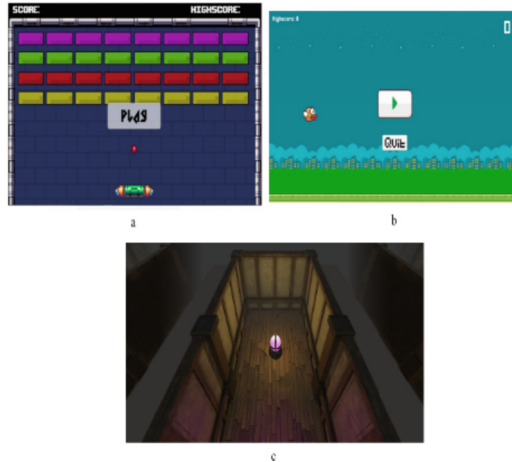


**Fig. 5.** Developed balance training games. a) Balance Block Breaker; b) Balance Bird; c) Balance Maze

Due to the pandemic state, the device validation was done through videos (Fig. 6). In line with the Ministry of Health and in order to avoid contamination with the COVID-19 virus, everyone involved was properly protected.

Through the SUS scale, the following results were obtained and are showed in Table 2:

The results obtained from the evaluation of the physical therapists reached an average of 81 points on the SUS scale. And the results obtained by the System Developers evaluation reached an average of 83 points by the SUS scale. According to the literature [17], the results of both groups are classified as a B. The literature [18] further states that the average SUS scale is 70.14 points, and that tests with scores above 70 but below 90 can be considered good. Thus, one can legitimize the evaluation of professionals in the correlated areas of study.

## 4 Discussion

In this study, 3 digital games were developed using the Unity game engine [19]. The games were associated with a force platform, and offer conditions for the individual to experience obstacles and difficulties of the real world in a safe way. Likewise, the study [20] used Unity in the development of its game for balance training in the elderly and the work [21] used a force platform to train balance in the elderly in a controlled and safe way.
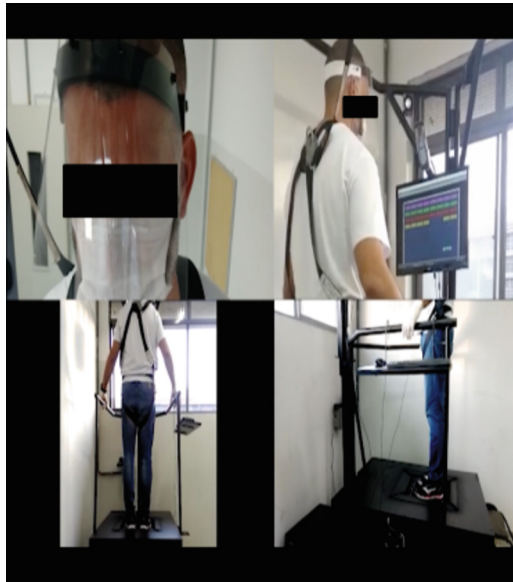
**Fig. 6.** Balance Block Breaker game recording

Games can be considered a tool for balance training, helping health professionals [20]. They can bring more concentration to the individuals who are using them, providing a playful and more attractive environment when compared to conventional training methods, also being used for rehabilitation [11].

The present study developed games for balance training, and they were combined with a force platform with controlled inclination that is used to control the Avatars of the games. Thus, several protocols can be performed, using the inclined platform at different angles, inducing the patient to maintain his balance by performing anteroposterior and laterolateral movements.

These movements can bring results of improvement to the individual, since for this the sensorial and motor system is trained. One of the factors that influence postural balance is the stabilization of the ankle joint [21]. Through it, the movement of plantar flexion and dorsiflexion is performed, and it also influences the anteroposterior and laterolateral. Some muscles are also part of this system, such as the medial gastrocnemius and anterior tibial [22]. All of this can influence postural balance [23].

As the individual needs to move to control the games, this makes them dynamic, and may also train compensatory postural adjustment and anticipatory postural adjustment. The fact that the games have a scoring system and a time register allows the evaluation of the individual's evolution during the gametherapy sessions.

Each game is designed to suit a specific type of movement, unlike commercial games that cannot be changed and are inaccessible to some participants [24].

During the SUS evaluation, a combination of specialists in software development and in physiotherapy brought about a broad evaluation of the system from different points of view. The results showed that it was simple, userfriendly, easy to use and totally safe,

**Table 2.** Average per questions in the SUS assessment questionnaire

| Item | SUS assessment results for physiotherapy experts | |
|---|---|---|
| | Question | Average |
| 1 | The movements made in the proposed games areappropriate for balance training | 4,8 |
| 2 | The games weren't able to motivate the user's interest for a use | 2,0 |
| 3 | The games presented in any way may harm or damage the player's health | 1,6 |
| 4 | Gametherapy is not able to provide an improvement in balance training | 1,0 |
| 5 | They would often use the proposed gametherapy system in their therapies | 4,8 |
| 6 | The use of game therapy is not able to motivate balance training | 1,2 |
| 7 | You were satisfied with the functioning of the proposed gametherapy system | 4,6 |
| 8 | Wouldn't need technical help to use the proposed system | 3,0 |
| 9 | You were satisfied with the application of the safety structure to prevent accidents | 4,8 |
| 10 | You would not subject your patient to training using the proposed system | 1,0 |
| | SUS assessment results for development experts | |
| Item | Question | Average |
| 1 | The game is able to attract user interest to interact | 5,0 |
| 2 | The game is not able to provide easy operation and control to the user | 2,2 |
| 3 | The game is able to provide the results accurately and effects correctly as intended | 4,0 |
| 4 | The game lacks the ability to perform functions to accomplish tasks and goals | 1,0 |
| 5 | The game is able to prevent software or interface malfunctions | 3,4 |
| 6 | The game is not able to recover data if there is a crash while running the software | 2,6 |
| 7 | The game is easy to collect data | 5,0 |
| 8 | The game is unable to process and respond while performing its basic functions | 1,2 |
| 9 | The game has stability, due to possible changes in the software | 3,8 |
| 10 | Without the need for other actions, the game does not have adaptability to different environments | 1,8 |

without the need for great interference in usability. According to the literature [25], the SUS scale is a robust means of evaluation, being used in a variety of systems and/or devices.

## 5    Conclusions

According to the proposed objectives, the presented project can offer a useful tool to the health area, especially to the rehabilitation area, helping professionals who deal with balance in the therapeutic processes.

**Conflict of Interest.**    The authors declare that there is no conflict of interest regarding the publication of this paper.

## References

1. Fernie, G.R., Gryfe, C.I., Holliday, P.J., Llewellyn, A.: The relationship of postural sway in standing to the incidence of falls in geriatric subjects. Age Ageing **11**(1), 11–16 (1982). https://doi.org/10.1093/ageing/11.1.11
2. Zatsiorsky, V.M.: Kinetics of Human Motion (2002). http://books.google.com/books?id=wp3zt7oF8a0C&pgis=1
3. Winter, D.A.: Biomechanics and Motor Control of Human Movement, 4th edn (2009). https://doi.org/10.1002/9780470549148
4. Chiaramonte, R., Bonfiglio, M., Leonforte, P., Coltraro, G.L., Guerrera, C.S., Vecchio, M.: Proprioceptive and dual-task training: the key of stroke rehabilitation, a systematic review. J. Funct. Morphol. Kinesiol. **7**(3), 53 (2022). https://doi.org/10.3390/jfmk7030053. PMID: 35893327; PMCID: PMC9326539
5. da Silva, A.R.F., Lazzareschi, L., de Paula Bischof, K., et al.: Distribution of mass and center of pressure in dynamic balance presented by an instrumented wobble board. Res. Biomed. Eng. **36**, 501–510 (2020). https://doi.org/10.1007/s42600-020-00092-6
6. DiStefano, L.J., Clark, M.A., Padua, D.A.: Evidence supporting balance training in healthy individuals: a brief review. J. Cardiopulm. Rehabil. Prev. **30**(2), 139–140 (2010). https://doi.org/10.1097/HCR.0b013e3181d6c875
7. Bateni, H.: Changes in balance in older adults based on use of physical therapy vs the Wii Fit gaming system: a preliminary study. Physiotherapy (UK) **98**(3), 211–216 (2012). https://doi.org/10.1016/j.physio.2011.02.004
8. ACMS. ACSM's Guidelines for Exercise Testing and Prescription (2010)
9. Amritha, N., Mahima, M.M., Namitha, K., et al.: Design and development of balance training platform and games for people with balance impairments. In: 2016 International Conference on Advance Computer Communication Informatics (ICACCI 2016), pp. 960–966 (2016). https://doi.org/10.1109/ICACCI.2016.7732169
10. Lange, B., Flynn, S., Proffitt, R., Chang, C.Y., Rizzo, A.: Development of an interactive game-based rehabilitation tool for dynamic balance training. Top. Stroke Rehabil. **17**(5), 345–352 (2010). https://doi.org/10.1310/tsr1705-345
11. Postolache, O., Girão, P.S., López, A., Ferrero, F.J., Dias Pereira, J.M., Postolache, G.: Postural balance analysis using force platform for K-theragame users. In: Proceedings of the 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA 2016) (2016). https://doi.org/10.1109/MeMeA.2016.7533705

12. Ishizaki, M.T., et al.: Development of a dynamic balance analysis and training system through posturography in a bipodal force platform with biaxial rotation. Res. Biomed. Eng. 1–17 (2022)
13. Nielsen, J.: Why You Only Need to Test with 5 Users. Jakob Nielsens Alertbox. http://www.useit.com/alertbox/20000319.html
14. ABNT. NBR ISO/IEC 14598-1 Tecnologia de informação - Avaliação de produto de software - Parte 1: Visão geral. 1–14 (2001)
15. Brooke, J.: This document: more Project information and further documents: SUS - a quick and dirty usability scale. Psychol. Mark. **27**(10), 1–8 (1986)
16. Tullis, T.S., Stetson, J.N.: A comparison of questionnaires for assessing website usability. Abstract: Introduction. In: Usability Professionals Association Conference, pp. 1–12 (2004). http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf
17. Bangor, A., Staff, T., Kortum, P., Miller, J., Staff, T.: Determining what individual SUS scores mean: adding an adjective rating scale. J. Usabil. Stud. **4**(3), 114–123 (2009)
18. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Int. J. Hum. Comput. Interact. **24**(6), 574–594 (2008). https://doi.org/10.1080/10447310802205776
19. UNITY3D. (2021). https://store.unity.com/products/unity-pro?gclid=CIm178878gCFciBkQodLhMMIQ%3E. Accessed 25 Feb 2021
20. Eggenberger, P., Wolf, M., Schumann, M., de Bruin, E.D.: Exergame and balance training modulate prefrontal brain activity during walking and enhance executive function in older adults. Front. Aging Neurosci. **8**, 1–16 (2016). https://doi.org/10.3389/fnagi.2016.00066
21. Terada, M., Harkey, M.S., Wells, A.M., Pietrosimone, B.G., Gribble, P.A.: The influence of ankle dorsiflexion and self-reported patient outcomes on dynamic postural control in participants with chronic ankle instability. Gait. Posture **40**(1), 193–197 (2014)
22. Galdino, T.R., Kneubuehler, P.A.: Análise comparativa do movimento de extensão de joelho na cadeira extensora, associado à dorsiflexão ou flexão plantar do tornozelo durante treinamento de hipertrofia em mulheres. RBPFEX-Rev. Brasil. Pres. E Fisiol. Exer. **12**(73), 130–139 (2018)
23. Spink, M.J., Fotoohabadi, M.R., Menz, H.B.: Foot and ankle strength assessment using hand-held dynamometry: reliability and age-related differences. Gerontology **56**(6), 525–532 (2010)
24. Whyatt, C., Merriman, N.A., Young, W.R., Newell, F.N., Craig, C.: A wii bit of fun: a novel platform to deliver effective balance training to older adults. Games Health J. **4**(6), 423–433 (2015). https://doi.org/10.1089/g4h.2015.0006
25. Katsanos, C., Tselios, N., Xenos, M.: Perceived usability evaluation of learning management systems: a first step towards standardization of the system usability scale in Greek. In: Proceedings of the 2012 16th Panhellenic Conference Informatics (PCI 2012), pp. 302–307 (2012). https://doi.org/10.1109/PCi.2012.38

# Automatic Morphological Evaluation of Endothelial Cells Using Different Classification Methods

Miriela Escobedo-Nicot[1](✉) , Wilkie Delgado-Font[1] ,
Elisângela Monteiro-Pereira[2] , and Ligia Ferreira-Gomes[3] 

[1] Universidad de Oriente, Patricio Lumumba No. 507, Santiago de Cuba, Cuba
`miri@uo.edu.cu`
[2] Universidade Federal de Alfenas, Rua Gabriel Monteiro da Silva 700, Minas Gerais, Brasil
[3] Universidade de São Paulo, R. da Reitoria, 374 Cidade Universitária, Butantã, São Paulo, Brasil

**Abstract.** The cell morphology analysis is applicable to pathophysiology studies in biological samples. In this work, digital images of Human Umbilical Vascular Endothelial Cells (HUVEC) were classified according to their morphological properties, to help the detection of functional and/or structural anomalies for the study of angiogenesis, a process by which new capillaries are formed from pre-existing capillaries. The automatic classification was produced by the algorithms: support vector machine (SVM), k-Nearest Neighbors (k-NN), and decision trees (DT), with three classes: circular, elongated deformed (elongated), and slightly elongated deformed (others deformations). The processes of cell migration and proliferation could be correlated with this classification. The sensitivity values for all three methods exceed 95%. The highest accuracy value, 98.89%, was reached by SVM method. Results shows that it is feasible to use these three methods for the classification of HUVEC.

**Keywords:** HUVEC · endothelial cell · angiogenesis · classification · morphology analysis

## 1 Introduction

Human Umbilical Vascular Endothelial Cells, known as HUVEC are employed for the study of angiogenesis, which is the process of forming new capillaries from existing ones [1]. This process includes events like growth, migration, and differentiation of cells. Its study is of great importance as it is a process closely related to a large number of pathological events, including the transition from the harmless vascular growth phase to the potentially fatal tumor stage [2]. During angiogenesis, cells develop different shapes depending on the function they will later assume, which can be correlated with the appearance of tip and stalk cell phenotypes: some take elongated forms and characterize the vascular growth path, and these are associated with the tip cells; others take forms close to the circular ones guaranteeing the organization of the lumen in the new vascular formation, these are stalk cells [3].

HUVEC sample studies use the microscope, which is a great effort for the specialist, who must observe, sometimes for hours in the microscope, the visual fields to issue a criterion. In these cases, fatigue, the position of observation from the specialist, and even the variability of criteria among specialists affect the quality of the obtained results and cause the effort to achieve them to be high to guarantee the accuracy of the samples.

In vitro test enables, from microscopy images, the study of endothelial cell migration and proliferation processes, and two-dimensional or three-dimensional organization of capillary structures, which allows in many cases to identify the set of possible effects of a type of intervention (proteins, cells, pharmacological agents) or molecular and/or cellular events in angiogenesis [4–7].

Concerning HUVEC, automatic quantitative analysis is relatively a new topic. Studies of in vitro angiogenesis have been developed from isolated endothelial cells cultured in a micro or macrovascular form. So far, these studies focus more on the analysis of the tubular structures of the cellular networks formed in the cultures, using different computational platforms [8], so that angiogenic activity can be assessed by quantifying the tubules, the junctions between them, and other structures of interest, the area, the length of the tubules, among others.

Wiseman et al. in [9] presented a high-content HUVEC phenotyping platform using morphology, vascular endothelial cadherin staining, and analysis of NOTCH activation. The authors collected a total of 47 features e.g. cell and nuclei area, roundness, width/length ratio together with an array of symmetry, threshold compactness, and axial or radial features. Other characteristics extracted are analysis of junctions, population stratification, and quantification of size for clusters of NOTCH positive cells. With these characteristics it was possible to study the phenotypes of HUVEC under different experimental conditions.

In [10], the authors presented two algorithms to study the endothelial tube formation assay (ETFA) and fibrin bead assay (FBA) in HUVEC. Image analysis in both cases was using a program developed for ImageJ software. This plugin is an extension of the "Angiogenesis Analyzer" for ImageJ. In FBA, the analysis included sphere detection, tree detection, and tree structure analysis of junctions and extremities. On other hand, the analysis in ETFA included meshed detection and isolated elements detection.

In [8] the authors used a combination of machine learning aided image segmentation with ImageJ software and CellProfiler software, to extract the phenotype of a single EC (HUVEC, HAoEC, HPMEC) in monolayers. To measure the cell morphology, they used the area, perimeter, shape descriptors, and cell neighbors.

On the other hand, some authors considered the study of individual cells using the morphological changes in these samples. In [11], the Angiogenesis Analyzer plugin for ImageJ performed the cell shape analysis for HUVEC-C3 differentiation, the inhibition of HUVEC-C3 differentiation with target agents, and the analysis of the Matrigel differentiation assay. The parameters used were the form factor (FF), tubular networks, number of junctions, number of tubules (number of branches plus the number of segments), and tubules length (total branching lengths).

Escobedo et al. [3] used the mathematical functions generalized support, a variant of it called weighted integrated generalized support, and the Crofton descriptor, in the analysis of cell deformations in 2D HUVEC samples. The authors detected cell edges using

active contour methods and classified them using the k-NN algorithm with a 5x1 cross-validation process for error estimation. The supervised classification contemplates three classes: circular cells, elongated deformed (elongated), or slightly elongated deformed (other deformations).

At early times of the incubation process, endothelial cell differentiation can express different levels of deformation due to the differential activation of the cell migration and proliferation processes, and the consequent appearance of the corresponding tip and stalk phenotypes [3]. Considering these results, and the feasibility of testing the classification performance of HUVECs using different morphological features and other widely used classification methods (which have shown excellent results in other types of samples), this work aims to compare the results obtained with three classifiers: support vector machine (SVM), k-NN and decision trees (DT), considering the three classes proposed in [3]: circular, elongated deformed and slightly elongated deformed (other deformation).

The paper is organized as follows. After this introduction, in Sect. 2, it is presented all the information related to sample preparation and image acquisition, besides the explanation of the methodology proposed using the three classification methods, and the experiments performed. The experimental results are described in Sect. 3 and the discussion in Sect. 4. Finally, concluding remarks and future work are presented in Sect. 5.

## 2   Materials and Method

### 2.1   Sample Preparation and Image Acquisition

Cultures of endothelial cells in the concentration of $2 \times 10^4$ cells/ml were placed in RPMI 1640 medium supplemented with fetal bovine serum (FBS) to 10% in 24-well plates, on glass coverslips of 13 mm diameter, previously sterilized. Each of the wells containing the samples was added at a concentration equal to 30 µg/ml of β2-glycoprotein. Cells were stained with May Grunwald – Giemsa [3].

Images were taken using a Point Grey camera (GS3-U3-1455M) with a 3.2 × magnification objective. The images obtained have $114 \times 114$ pixels and 72 pixels per inch. Obtaining a total of 360 single images, 120 for each class, Fig. 1.

### 2.2   Methodology

Figure 2 shows the methodology steps based on the following phases: image segmentation, feature extraction, and classification.

First, we binarized the original image. Let $f(x, y)$, such that the pixels belonging to the objects and the background have intensity values grouped in two dominant modes. The thresholded image $g(x, y)$ is defined as follows [12]:

$$g(x, y) = \begin{cases} 1, & if (x, y) \geq T \\ 0, & if (x, y) < T \end{cases} \quad (1)$$

where pixels labeled 1 correspond to the objects, 0 to the background and T is the threshold value, using Otsu's threshold. As a result of this process, we obtain an image with a black background and white objects.
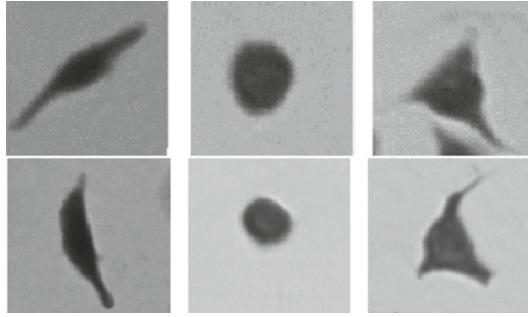
**Fig. 1.** Cell images of circular, elongated deformed (elongated), and slightly elongated deformed (other deformation) class of HUVEC.
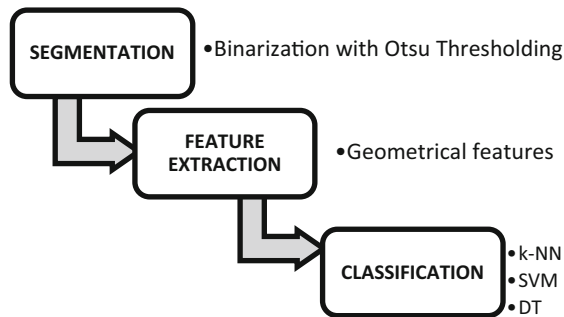


**Fig. 2.** Images processing steps from HUVEC automatic classification.

For the feature extraction, we selected 11 shape features to facilitate the classification of the different cell types. Shape features provide measures of the cells based on their geometrical properties. Table 1 shows the description of each of the characteristics assessed.

The classification methods used were SVM, k-NN, and DT. The SVMs are linear classifiers that allow finding the optimal way to classify between classes from the support vectors. Given these support vectors, the optimal classification is performed by finding the hyperplane that maximizes the margin of separation between classes. This hyperplane is obtained from the solution of an optimization problem: the maximum distance between the hyperplanes containing the support vectors of the classes (maximum margin) [13]. The approximation provided by SVMs is only useful for classifying sets that are linearly separable and without overlaps between classes. But considering that this situation is quite frequent in real cases, then, in this case, to be efficient in these environments, there are two alternatives: a certain number of class overlaps are allowed by relaxing the margin condition, and on the other hand, the model is extended so that it can work on sets that are not linearly separable, by transforming the feature space into another of higher dimension, which is known as kernel trick.

On the other hand, the k-NN algorithm given the set of previously labeled objects (training set), assigns the analyzed object to the class of the nearest neighbors in the

**Table 1.** Morphological measures used for the classification of HUVEC in this work.

| Morphological measures | Description |
| --- | --- |
| Perimeter | Distance around the boundary of the region |
| Area | The number of pixels in the region |
| Eccentricity | The eccentricity of the ellipse with the same second-moments as the region |
| Equivalent Diameter | Diameter of a circle with the same area as the region |
| Major Axis Length | Length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region |
| Minor Axis Length | Length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region |
| Solidity | The proportion of the pixels in the region that are also in the convex hull |
| Extent | The ratio of pixels in the region to pixels in the total bounding box |
| Orientation | The angle between the $x$-axis and the major axis of the ellipse that has the same second moments as the region |
| CSF | Represents deviation from a circle |
| ESF | Expresses the extent of elongation |

training set, corresponding to a similarity measure in the feature space [14]. Generally, the metric used is Euclidean. This algorithm has the disadvantage that it requires prior knowledge of the value of k to determine the k nearest neighbors, in addition to the fact that, with different values of k, different results are also obtained.

In the case of DT, these are sequential models, which logically combine a sequence of simple tests; each test compares a numeric attribute against a threshold value or a nominal attribute against a set of possible values [15]. There are two major phases of the decision tree induction process: the growth phase and the pruning phase.

## 2.3 Experimental Setup

The tests were carried out on a PC with an AMD A8-6410 APU with AMD Radeon R5 Graphics at 2.00 GHz and 12 GB of RAM. The implementation of all algorithms was performed using Matlab®. The experiment considered the supervised classification of circular (C), elongated deformed (ED), slightly elongated deformed (SED) HUVEC, using the following settings:

a) SVM:
    kernel = medium Gaussian
    boxConstraint = 5
    kernel scale = 3.3
    multiclass method: One-vs-All

b) k-NN:

    k = number of neighbors (3)

    distance = city block

    rule = nearest

c) DT:

    preset: Fine Tree

    max_num_Splits=100

    split criterion: Gini diversity index

    surrogate decision split: find all

The confusion matrix, with the following measures: sensitivity (TPR), specificity (TNR), precision (P), and F1-score (F1) for each class. As a measure of the performance of the complete process, accuracy (Acc) was used. These are described below [16]:

$$TPR = \frac{TP}{TP+FN} \tag{2}$$

$$TNR = \frac{TN}{FP+TN} \tag{3}$$

$$P = \frac{TP}{TP+FP} \tag{4}$$

$$F1 = \frac{2(PTPR)}{P+TPR} \tag{5}$$

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

where:

- TP corresponds to true positives.
- FP corresponds to false positives.
- FN corresponds to false negatives.
- TN corresponds to true negatives.

The classification validation is performed by cross-validation using the k-fold method with k = 5.

## 3  Results

Figure 3 shows the result of applying the method of Otsu with Thresholding Binarization for the image segmentation. Table 2 shows the confusion matrix of the classification of HUVEC considering the three classes and the classifiers SVM, k-NN, and DT.

The values obtained in relation to the performance measures assessed are also presented.

Table 3 shows the confidence interval with which the sensitivity to detection of each of the defined classes is calculated. For this purpose, we work with a conservative confidence interval for a proportion, which is given by the expression:

$$\hat{p} - h * \sqrt{\frac{\hat{p}*\hat{q}}{n}} \leq p \leq \hat{p} - h * \sqrt{\frac{\hat{p}*\hat{q}}{n}} \tag{7}$$

where:
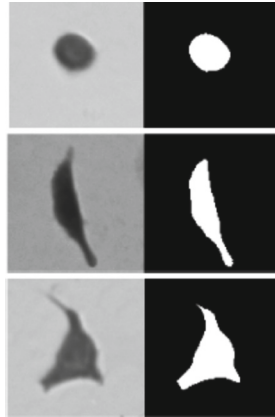
**Fig. 3.** Right column original images and left column segmented images: black color represents the image background and bank color the cells.

**Table 2.** Performance measures for the classification of HUVEC.

|  |  | C | DE | SED | TPR | P | TNR | F1 | Acc |
|---|---|---|---|---|---|---|---|---|---|
| SVM | C | 120 | 0 | 0 | 100 | 99.17 | 99.58 | 99.59 | 98.89 |
|  | DE | 0 | 119 | 1 | 99.17 | 98.35 | 99.16 | 98.76 |  |
|  | SED | 1 | 2 | 117 | 97.50 | 99.15 | 99.58 | 98.32 |  |
| k-NN | C | 120 | 0 | 0 | 100 | 98.36 | 99.17 | 99.17 | 98.06 |
|  | DE | 0 | 118 | 2 | 98.33 | 97.52 | 98.75 | 97.92 |  |
|  | SED | 2 | 3 | 115 | 95.83 | 98.29 | 99.17 | 97.04 |  |
| DT | C | 120 | 0 | 0 | 100 | 100 | 100 | 100 | 97.50 |
|  | DE | 0 | 117 | 3 | 97.50 | 95.12 | 97.50 | 96.30 |  |
|  | SED | 0 | 6 | 114 | 95.00 | 97.44 | 98.75 | 96.20 |  |

- $\hat{p}$ is the estimated sensitivity of the classifier.
- $\hat{q}$ is 1- the classifier sensitivity.
- $h = 1 - \frac{\alpha}{2}$ is the percentile of $N(0, 1)$, as the reliability level of the 95%, $h = 1.96$.

## 4 Discussion

As can be seen in Table 1, the three methods classified all the circular cells were as belonging to their class, which is the appropriate behavior, because it is the class with objects that are best differentiated from the rest of the objects in the other two classes. In the case of elongated cells, only one, two and three cells were classified as slightly elongated deformed for the SVM, k-NN and DT methods respectively, none was classified as circular. This behavior was not the same with the slightly deformed cells. There

**Table 3.** $IC_I$ and $IC_S$ correspond with the inferior and superior confidence interval for the sensitivity.

| Methods | Class | Sensibility | $IC_I$ (%) | $IC_S$ (%) |
|---------|-------|-------------|------------|------------|
| SVM | C | 100 | 100 | 100 |
| | DE | 99.17 | 97.54 | 100 |
| | SED | 97.50 | 94.70 | 100 |
| k-NN | C | 100 | 100 | 100 |
| | DE | 98.33 | 96.03 | 100 |
| | SED | 95.83 | 92.25 | 99.40 |
| DT | C | 100 | 100 | 100 |
| | DE | 97.50 | 94.70 | 100 |
| | SED | 95.00 | 91.10 | 98.89 |

were cells of this type that were classified as elongated for all methods and in the case of the SVM and k-NN methods they were also classified as circular cells.

The sensitivity reached values of 100% in the circular class and of more than 97% in the elongated deformed class. In the specific case of the low elongated deformed class, which has objects with greater shape variation, the classification showed lower sensitivities results, although they can be considered good, with above 95% for the k-NN and DT methods and 97.50% for the SVM method. Nevertheless, the precision and specificity of this class remain high for all three methods with values above 97%, which means that, although compared to the circular class, they defined fewer objects as belonging to this class than the total contained in it; those they defined were indeed belonging to the class they were classified to be in.

The highest precision values were found for the circular class for all three methods. The specificity values for the circular class exceeded 99% and for the other two classes, the values were higher than 97%. With the F-measure, the behavior was not similar in all methods; the highest results were in the circular class (100% for the DT method) and the most conservative in the slightly elongated deformed class (96.20% for the DT method). The SVM method outperformed the others by achieving an accuracy of 98.89%, but it should be pointed out that for the k-NN and DT methods, the accuracy achieved represent a good result, with a value of 98.06% for k-NN and 97.50% for the DT. To indicate the computational cost of our algorithm, the time taken for training these methods, under the experimental conditions defined were the followings:

- SVM: 0.70s
- k-NN: 0.36s
- DT: 0.33s

The results in Table 3 show the values corresponding to the sensitivity confidence intervals, presenting the values that these can take, taking into account each class in the classifiers evaluated. The results obtained in this work can be considered superior to those obtained in [3], in which images were used under the same conditions, but with a smaller number, only 300 images. The sensitivities reported in [3] for the best method, the Crofton descriptor, were 99%, 91%, and 86%, for the circular, elongated deformed, and slightly elongated deformed classes, with an accuracy of 92.67%.

## 5  Conclusions

This work presents the results of applying the SVM, k-NN, and DT methods in the morphological classification of HUVEC considering three classes, which is of great importance because it could be correlated with the appearance of tip and stalk cell phenotypes from the early stages of cell development. The methods were tested considering 360 images of individual cells, 120 for each defined class, and 11 morphological characteristics. The highest accuracy. 98.89%, was reached by SVM method. The sensitivity values in all cases exceed 95% and the class with the lower results in this section was the slightly elongated deformed class, considering that the shape of HUVECs in this class was relatively similar to the elongated deformed cell class. To check the confidence interval with which the sensitivity of the classification in each class was calculated, it was considered a conservative confidence interval for a proportion.

Further work in perspective, the method will be applied to investigate angiogenesis in endothelial cell challenging conditions, by analyzing cell morphology in microscopy images and moving towards clinically interpretable approaches. The evaluation of other features such as textures, and color, among others, as well as the performance of other classifiers is also of interest.

## References

1. Jin, S., et al.: Conditioned medium derived from FGF-2-modified GMSCs enhances migration and angiogenesis of human umbilical vein endothelial cells. Stem Cell Res. Ther. **11**(1), 1–12 (2020)
2. Pijuan, J., et al.: In vitro cell migration, invasion, and adhesion assays: from cell imaging to data analysis. Front. Cell Dev. Biol. (7), 07 (2019). https://doi.org/10.3389/fcell.2019.00107
3. Escobedo, M., Herold, S., Ferreira, L., Machado, C., Monteiro, E., Delgado, W.: Morphological analysis of HUVEC samples with integral geometry based functions. Ciencias Matemáticas **30**(2), 79–86 (2016)

4. Xu, L., Willumeit-Römer, R., Luthringer-Feyerabend, B.: Effect of magnesium-degradation products and hypoxia on the angiogenesis of human umbilical vein endothelial cells. Acta Biomater. **98**, 269–283 (2019)
5. Amann, A., et al.: Development of a 3D angiogenesis model to study tumour endothelial cell interactions and the effects of anti-angiogenic drugs. Sci. Rep. **7**(1), 1–13 (2017)
6. Bang, H., Yoon, E., Kim, S., Ahn, M.: Antiangiogenic and apoptotic effects of benzyl caffeate on human umbilical vein endothelial cells (HUVECs) and chick embryo chorioallantoic membrane (CAM): in vitro and in vivo models. J. Funct. Foods **93**, 105079 (2022). https://doi.org/10.1016/j.jff.2022.105079
7. Sarani, M., et al.: Study of in vitro cytotoxic performance of biosyn-thesized α-Bi2O3 NPs, Mn-doped and Zn-doped Bi2O3 NPs against MCF-7 and HUVEC cell lines. J. Mater. Res. Technol. **19**, 140–150 (2022)
8. Chesnais, F., et al.: High-content image analysis to study phenotypic heterogeneity in endothelial cell monolayers. J. Cell Sci. **135**(2), jcs259104 (2022)
9. Wiseman, E., et al.: Integrated multiparametric high-content profiling of endothelial cells. SLAS Discov. **24**(3), 264–273 (2019)
10. Carpentier, G., et al.: Angiogenesis analyzer for ImageJ—A comparative morphometric analysis of "Endothelial Tube Formation Assay" and "Fibrin Bead Assay." Sci. Rep. **10**(1), 1–13 (2020)
11. Chiew, Y., Fu, A., Perng, L.K., Qian, L.K.: Physical supports from liver cancer cells are essential for differentiation and remodeling of endothelial cells in a HepG2-HUVEC co-culture model. Sci. Rep. **5**(1), 1–16 (2015)
12. Tensmeyer, C., Martinez, T.: Historical document image binarization: a review. SN Comput. Sci **1**(3), 1–26 (2020)
13. Namdeo, A., Singh, D.: Challenges in evolutionary algorithm to find optimal parameters of SVM: a review. Mater. Today Proc. (2021). https://doi.org/10.1016/j.matpr.2021.03.288
14. Mezquita, Y., Alonso, R.S., Casado-Vara, R., Prieto, J., Corchado, J.M.: A review of k-nn algorithm based on classical and quantum machine learning. In: Rodríguez González, S., et al. (eds.) DCAI 2020. AISC, vol. 1242, pp. 189–198. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-53829-3_20
15. Charbuty, B., Abdulazeez, A.: Classification based on decision tree algorithm for machine learning. J. Appl. Sci. Technol. Trends **2**(1), 20–28 (2021)
16. Font, W.D., et al.: Classification of red blood cell shapes using a sequential learning algorithm. In: Bastos-Filho, T.F., de Oliveira Caldeira, E.M., Frizera-Neto, A. (eds.) CBEB 2020. IP, vol. 83, pp. 2059–2065. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-70601-2_301

# Synthesis and Internalization of MCHC-Chlorin Photosensitizers on *Trichophyton Rubrum*- A Preliminary Study

M. W. M. Lopes[1,2(✉)] 🆔, H. C. G. Veiga[1,2] 🆔, and A. F. Uchoa[1,2] 🆔

[1] Universidade Anhembi Morumbi (UAM), São José dos Campos, SP, Brazil
mariane.wml@gmail.com

[2] Centro de Inovação, Tecnologia e Educação (CITE), São José dos Campos, SP, Brazil

**Abstract.** It appears that there is widespread difficulty in treating onychomycosis. Fungi are difficult to eradicate and the very dense keratin that participates in the structure of the nail, and there is still little vascularization, making it difficult for drugs to penetrate topically or systemically. Conventional treatment methods, in addition to therapeutic failures, cause undesirable side effects in most patients. The prolonged time of use reflects in the low adherence to the treatment and in the resistance of the fungus to the drug. In this sense, Photodynamic Therapy is a promising alternative for solving most of these problems. However, the technique is dependent on the photosensitization potential. The Diels Alder reaction between protoporphyrin IX dimethyl ester and maleimides provides endo adducts that completely prevent the self-aggregation of chlorins. Fluorescence resonant light scattering (RLS) and ^{1}H NMR experiments, as well as crystallographic X-rays have demonstrated this configuration. The arrangement of synthesized chlorins avoids π stacking in the interactions between macrocycles, indicating that it is a non-aggregating photosensitizer with high quantum yield of singlet oxygen ($\Phi\Delta$) and fluorescence ($\Phi f$). Preliminary results demonstrate that this type of synthetic strategy can provide the lead for a new generation of non-aggregative chlorine photosensitizers (MCHC-Chlorin) for PDT.

**Keywords:** Photodynamic Therapy · onychomycosis · non-aggeregative photosensitizers · chlorines

## 1 Introduction

Onychomycosis is a fungal infection that affects the nail plate, caused by dermatophytes, with *Trichophyton rubrum* or *Trichophyton mentagrophytes* identified as the most frequent causative pathogens. Yeasts are responsible in 5% of cases and non-dermatophyte fungi, responsible in 3% of cases. Characterized by making the nails brittle, thick, with excess or lack of pigmentation, it is considered one of the most frequent dermatoses, representing around 30% to 50% of all nail infections [1].

Due to its high incidence and prevalence, it is classified as a public health problem in Brazil. It is a disease multifactorial, a fact that often goes unnoticed and deserves

attention on several social and health fronts. We can mention as contributing factors for the development of the pathology the characteristics of our climate, demographic and social data.

Another problem is those with Acquired Immunodeficiency Syndrome, Diabetes, peripheral arterial diseases, and the elderly population, which are groups more prone to the development of the pathology and, mainly, to present severe forms of the disease in addition to difficulties and restrictions on treatment. Onychomycosis causes considerable physical and psychological effects to the patient, resulting in several implications and deficiencies [2–4].

Its treatment is a challenge in clinical practices due to the particularities of the pathology and the affected region. The slow growth of nails, especially on the feet, and because it is a non-vascularized region, makes it difficult for drugs to penetrate drug therapy. Conventional treatments consist of the use of topical or systemic antifungal agents, depending on the proliferation of the pathology. Systemic treatment can result in a number of adverse effects, such as: erythema, headache, hypersensitivity, possible drug interactions and a greater propensity to trigger serious liver diseases due to the extended period of treatment. In addition to the adverse effects, the long time required and the high cost of treatment contribute to low patient adherence and therapeutic failures are very common due to the fungal resistance to the drug used. Because of these implications, topical antifungals are the most used by patients, but they have low efficacy precisely because they do not have the ability to completely eliminate the infection. All these problems show that conventional treatments are increasingly failing, making necessary new therapeutic modalities that are safe and effective, reducing the disadvantages and high rate of recurrence of the disease, observed with commonly used antifungals [1, 5].

In turn, photodynamic therapy (PDT) emerges as a promising resource for the treatment and diagnosis of various pathologies. PDT occurs in the presence of 3 basic elements: photosensitizing drug; light source and molecular oxygen. The infected tissue is pre-treated with the photosensitizer (Fs) and subsequently irradiated locally with light of the appropriate wavelength. The interaction of these factors leads Fs to the triplet excited state interacting with the cells of the diseased tissue, which, in the presence of circulating cellular oxygen, produce reactive oxygen species (ROS) [5, 6].

Photodynamic therapy has been used in the treatment of onychomycosis with conventional photosensitizers such as methylene blue and protoporphyrin IX. The photodynamic effect is related to characteristics such as: the biodistribution of photosensitiers, interactions with specific biomolecules and organelles and with their phototophysical properties. However, this treatment is limited by aggregation factors. Compounds that have a high aggregation index present unfavorable photophysics, with low triplet quantum yields ($\tau$) and singlet oxygen ($\Phi\Delta$). Conventional photosensitizers present aromatic rings with a large $\pi$ cloud and extensive flatness, all of which are induced to self-aggregate. MCHC-Chlorin is a new photosensitizer composed of two orthogonal macrocycles joined by an $sp^3$ carbon. Such conformation, confers steric impediment and, consequently, prevents the formation of the aggregate state. Thus, such photosensitizer presents its photophysics maximized with high singlet oxygen quantum yield ($\Phi\Delta$) 0.78 and fluorescence quantum yield ($\Phi f$) of 0.16. These results enable MCHC-Chlorin as an excellent photopharmaceutical for photodynamic therapy and diagnostic treatments.

Thus, opening a new window for the study of PDT with the first photosensitizer free of self-aggregation. In this way, this photosensitizer may offer a more effective treatment for onychomycosis [7–9]. In this work, A MCHC-Chlorin was synthesized and characterized, in order to determine its photodynamic activity in *Trichophyton rubrum.*

## 2  Materials and Methods

A. *The Synthesis of 1-Phenyl-1H-Pyrrole-2,5-Dione(Phenylmaleimide)*

   To obtain phenylmaleimide, 300 mg (3.6 mmol) of maleic anhydride in 4.0 ml of ethyl ether, immersed in an ice bath, were added with 3.6 mmol of aniline. The precipitate formed, the amic acid, was vacuum filtered and washed with ethyl ether. A solution of 240 mg (2.84 mmol) of anhydrous sodium acetate and 3 ml of acetic anhydride was prepared and left in an oil bath ($\approx$70 °C) then the amic acid obtained was added. The reaction was followed, precipitated in ammonia/ice solution, extracted with chloroform and the solvent was removed by rotary evaporation under reduced pressure. After complete elimination of the solvent, the product mass determined and resuspended and purified on column chromatography, using silica gel (230–70 mesh) as support and dichloromethane with 1% methanol as mobile phase. After chromatography the maleimide was crystallized and characterized by $^1$H and $^{13}$C NMR and high resolution mass.

B. *The MCHC-Chlorin Synthesis*

   MCHC-Chlorin was synthesized by cycloaddition reaction between phenylmaleimide (0.5 mmol) with protoporphyrin IX (0.1 mmol) in toluene. The reaction system was kept at 120 °C, in an inert atmosphere, for a period of 12 h in a sealed tube. After this period, the end of the reaction was determined by ultraviolet/visible spectrum, and by TLC plates.

C. *Fungal Strain and Culture Media*

   To determine the antifungal activity of the photosensitizers, the strain of the species *Trichophyton rubrum* ATCC (American Type Culture Collection) 28188 was used.

   For cultivation, Sabouraud-Dextrose Agar (ASD) and Soy Triptecase in (TSB) liquid medium were used.

D. *Cultivation of Trichopyton Rubrum*

   The *Trichophyton rubrum* strain was cultivated in ASD medium, incubated at 28 °C for seven days, for colonies to grow. The mycelium was removed from this culture and plated in TSB liquid medium.

E. *Preparation of Trichopyton Rubrum for In Vitro Assays*

   After culturing *T. rubrum* in TSB medium for 72 h, the medium was centrifuged (4,000 rpm for 5 min) to wash the fungal cells. The supernatant was discarded and the precipitated cell material was resuspended in sterilized NaCl solution (0.5%) and centrifuged again (4,000 rpm for 5 min). This procedure was repeated 5 times to ensure asepsis of the sample. After the 5th wash, the conidia were resuspended in a sterile test tube containing 1 mL of saline solution and vortexed for 1 min. This suspension was diluted in sterile NaCl solution (0.5%) and homogenized by vortexing.

F. *Fluorescence Microscopy*

　　*T. rubrum cultures* in TSB medium obtained after 6 hours' incubation at 28 °C with MCHC-Chlorin were used for the study. After incubation, the sample was centrifuged (150 rpm) and then aliquots of the suspension were removed and deposited on glass slides and covered with covers lips for observation under a fluorescence microscope, (ZEISS AXIOVERT 200), using a 40× objective. The samples were excited in blue (410 ± 20 nm), and in green (546 ± 12 nm). The emission was recorded from $\lambda > 515$ nm and 590 nm respectively. The images were recorded on a Canon Power Shot G10 camera.

# 3　Results

A. *Synthesis of 1-Phenyl-1H-pyrrole-2,5-Dione (Phenylmaleimide)*

　　Phenylmaleimide was synthesized with an overall yield of 74%, and unequivocally characterized by $^1$H and $^{13}$C nuclear magnetic resonance.

　　The reaction scheme is shown in Fig. 1.
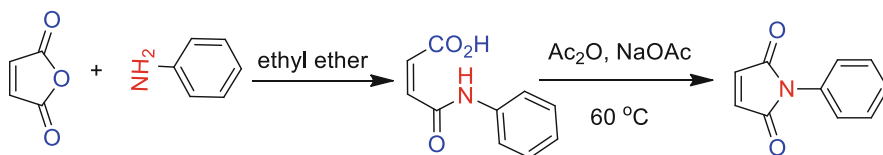


**Fig. 1.**　Synthesis of 1-phenyl-1H-pyrrole-2,5-dione

$^1$H NMR spectrum (CDCl$_3$, 500 MHz), δ(ppm): 6.84. (sbr, H-2, H-3 vinyl group) 7.33-7.38(m, H-7, H-8 and H-9)7.45–7.48 (m, H-6 and H-10). $^{13}$C NMR (CDCl$_3$, 125.77MHz)(ppm): 126.03(2C, C-6 and C-10); 127.93(C-8); 129.11(2C, C-7 and C-9); 131.17(C-5); 134.16(2C, C-2 and C-3) and 169.47(2C, C-1 and C-4) HPLC-MS,m/zfor C10H7NO2 calculated 174.0550 was determined to be174.0582. The base peak at 206.0827 refers to the opening of the imide ring, and the addition of methanol during the analysis process.

*The Synthesis of MCHC-Chlorin*

It was obtained through the addition cycle (4 + 2) - Diels–Alder reaction - a reaction already known to obtain chlorins. Phenylchlorin (ring A chlorine) and (B ring chlorine) were obtained by the process illustrated in Fig. 2, purified by column chromatography using silica gel as a support. The isomers were separated by preparative TLC. The technique offers a perfect separation of the chlorins of Ring A and B. In both techniques, chloroform/ethyl acetate in the propor-tions of 25:1 was used as mobile phase .as shown in Fig. 2.

Phenylchlorin A was isolated in the lower spot, with an Rf of 0.38. The structural characterization was unequivocally performed by 1D ($^1$H, $^{13}$C and 135 DPET) and 2D
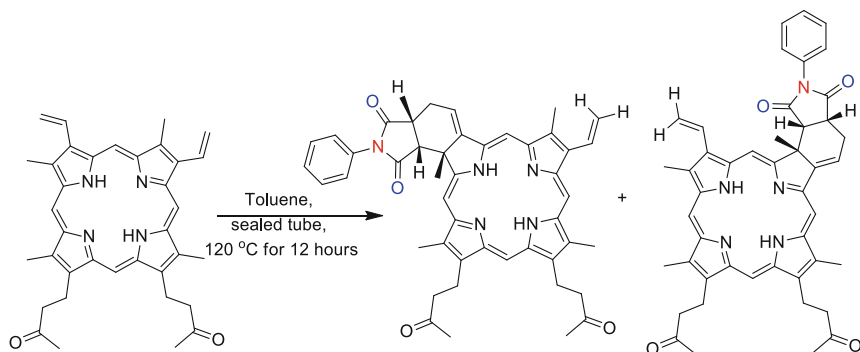
**Fig. 2.** Synthesis of phenylchlorin

(COSY, NOESY, HMBC and HSQC) NMR and high resolution mass, HPLC-mass. The $^{1}$H and $^{13}$C spectra are shown in Fig. 3a, for unequivocal characterization, the spectra of gCosy, gnoesy, gHSQC, gHMBC, and DEPT 135 were also determined. Phenylchlorin B was isolated in the upper spotand shows Rf of 0.54. The structural characterization was performed in a similar way to phenylchlorin A, by 1D ($^{1}$H, $^{13}$C and 135 DPET) and 2D (COSY, NOESY, HMBC and HSQC) NMR and high resolution mass, HPLC-mass. The $^{1}$H and $^{13}$C spectra are shown in Fig. 3b. The other spectra gCosy, gnoesy, gHSQC, gHMBC, DEP T135 were also determined. For characterization, the emission spectrum at 1270 nm obtained. Quantification performed with temporal resolution,and comparison of the are as below the curvature of the reference compound, phanalenone. The photophysical characterization performed with both isomers, being confirmed that they present equal quantum yield of singlet oxygen. The biological assays were performed only with the A isomer.

C. *Internalization of the Compound in the Fungal Cell Membrane*

Incorporation of MCHC-Chlorin in *T. rubrum* was ob-served by fluorescence microscopy, observed in red ($\approx$670 nm), referring to the emission of Chlorine. This emission was observed in hyphae and conidia, being perfectly differentiated from natural fluorescence of the species, which emits 450 nm when irradiated in ultraviolet. Figure 4 shows the fluorescence by emission of endogenous fluorophores and MCHC-Chlorin incorporate in *T. rubrum* hyphae.
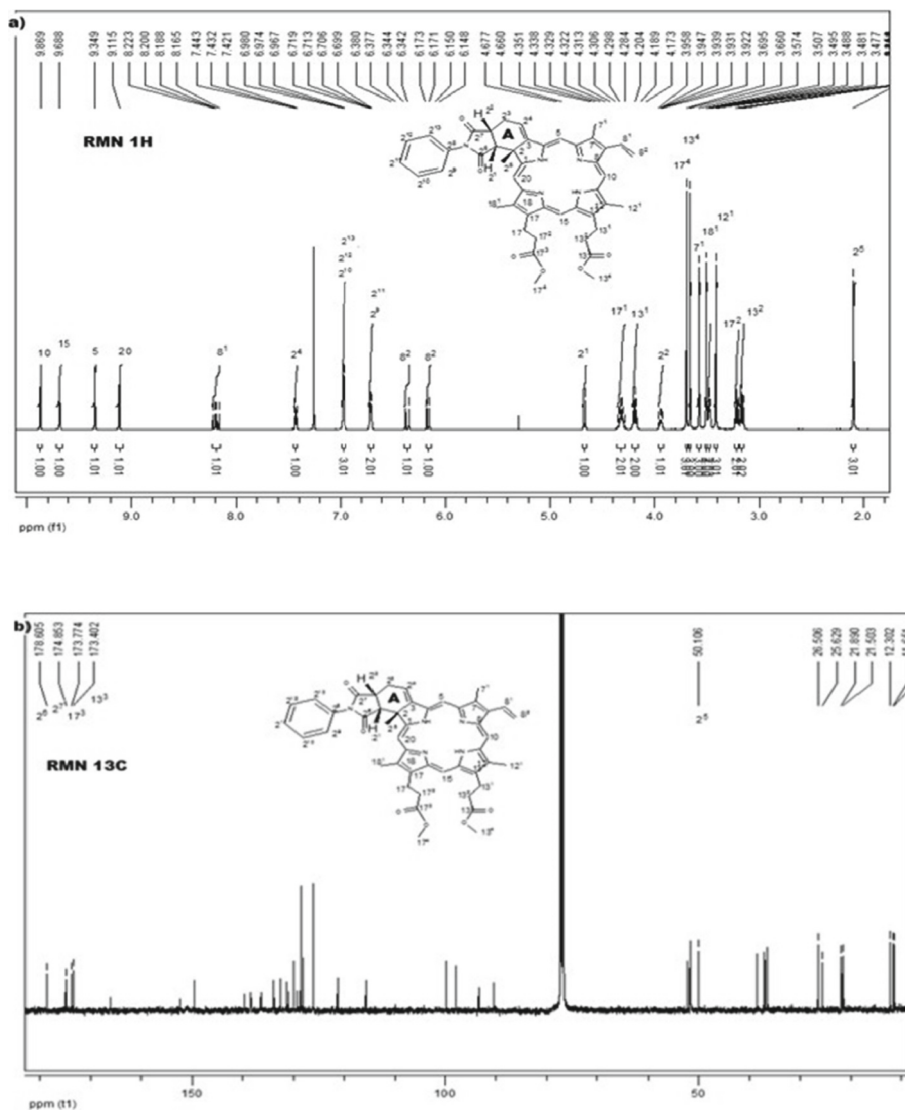
**Fig. 3.** CDCl$_3$ NMR spectra for phenylchlorin A a) $^1$H at 500.13 MHz in CDCl$_3$ and B b) $^{13}$C at 125.77 MHz
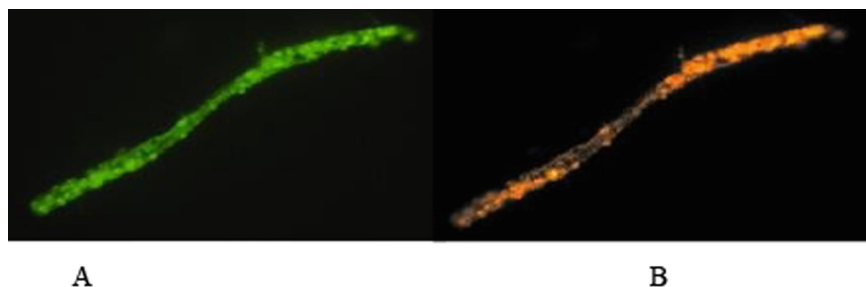
**Fig. 4.** A - fluorescence of the fungus excited at 340 nm and emission at 450 nm; B fluorescence through the incorporation of the MCHC-Chlorin photosensitizer, excited at 400 nm, emission at 670 nm. Images with a 100× objective.

## 4   Discussion

In the last decade, PDT has gone from an investigational method to a reality indermatological therapy. However, there are still some points that need to be improved and better understood. One of the fundamental lines of research for the improvement and popularization of PDT is the development of new photosensitizers, since the incorporation of the photosensitizer in the diseased tissue is as important as the formation of singlet oxygen. An efficient FS depends on a good interaction with biological systems, especially membranes. Producing assets that promote this interaction is an important factor in the development of new, more effective Fs. We still have some more essential characteristics for a compound to act as a good Fs: 1) Absorption in the phototherapeutic window (600–800 nm); 2) favorable photophysical characteristics: high quantum yield of singlet oxygen triplet formation and small suppression constant; 3) Low cytotoxicity in the dark; 4). Favorable pharmacokinetics and easy elimination by the body, a voiding prolonged photosensitivity. 5) Defined formulation, allowing to establish the relation between thestructure and the activity; 6) Solubility in a physiological medium, facilitating the application of the drug; 7) Easy to obtain on an industrial scale, with good reproducibility and reduced costs; 8) Skin permeability and selectivity; higher affinity in diseased skin compared to healthy tissue [5, 8]. One of the ways to solve these problems is the synthesis of a new class of Fs that are not self-aggregating, since the interaction with cell membranes can concentrate Fs and shift the monometer/aggregate balance due to the strong interaction between π systems of polyaromatic macrocycles, especially in the biological environment, where there is greater difficulty in controlling such events. Thus, the most effective strategy would be to avoid aggregation directly in the target skin through the use of molecules whose own molecular structure prevents aggregation, since their formation deactivates the excited state of the molecules, also leading to a consequent reduction the ability to generate singlet oxygen, triplet states and quantum fluorescence yields, compromising the efficiency of the photosensitization process [8–11].

With this objective, several authors have proposed the use of bulky peripheral groups in order to reduce and optimize the interaction between solvent and solute.

The Diels Alder reaction described here, between the dimethyl ester of protoporphyrin IX and phenylmaleimides, has been used to induce an endoaddition by secondary orbital stabilization, with the difference that these photosensitizers have bulkier substituents and a smaller nucleus, allowing the synthesis of new non-planar chlorine molecules, which have been shown to have the ability to completely prevent aggregation, even at high concentrations, and ensure affinity and internalization in the fungal cell membrane. This fact was confirmed by the high solubility of the compounds in solvents of different polarities. Preliminary biological assays demonstrate excellent permeability of MCHC-Chorin in T. rubrum [7, 12–15].

The fluorescence at 670 nm proves the incorporation and internalization of MCHC-Chorin in *T. rubrum* photodynamic feasibility of therapy with this new photosensitizing treatment. Similar processes have already been obtained with other photosensitizers, however, this new compound has maximized photophysics, with a quantum yield of singlet oxygen $(\Phi\Delta) = 0.71$, and a quantum yield of fluorescence $(\Phi f) = 0.16$.

## 5  Conclusions

Obtaining MCHC-Chlorin, a chlorin-free photosensitizer of self-aggregation, opens a new window for the development of PDT and Photodiagnostics. The absence of self-aggregation presents several gains, such as better solubility by maximizing the solvent/solute interaction, maximized photo-physics, with excellent $\Phi\Delta$ and $\Phi f$. And excellent permeability in the hyphae of Trichophyton rubrum which was proven by fluorescence microscopy.

**Conflict of Interest.**   The authors declare that they have no conflict of interest.

## References

1. Alberdi, E., Gómez, C.: Efficiency of methylene bluemediated photodynamic therapy versus intense pulsed light in thetreatment of onychomycosis in the toenails. Photodermatol. Photoimmunol. Photomed. (2018). https://doi.org/10.1111/phpp.12420
2. Martins, E.A., et al.: Onicomicose: estudo clínico, epidemiológico e micológico no município de São José do Rio Preto Onychomycosis: clinical, epiemiological and mycological study in the municipal ity of São José do Rio Preto. Rev. Socied. Brasil. Med. Trop. **40**(5), 596–598 (2007)
3. Finch, J.J., Warshaw, E.M.: Toenail onychomycosis: current and future treatment options. Dermatol. Therapy **20**(16), 31–46 (2007)
4. Ruiz, B.R.L., Chiacchio, N.: Manual de conduta nas onicomicoses. In: Diagnóstico e Tratamento. Sociedade Brasileira de Dermatologia |Departamento de Cabelos e Unhas (2015)
5. Christenson, J.K., et al.: Challenges and opportunities in the management of onychomycosis. J. Fungi (Basel). **4**(3), 87 (2018). https://doi.org/10.3390/jof4030087

6. Ribeiro, M.S., Silva, D.F.T., Nunez, S.C., Zezell, D.M.: Laser em baixa intensidade. In: de Maio, M. (Org.). Tratado de Medicina Estética, 2nd edn. Editora Roca, São Paulo (2011). https://www.ipen.br/biblioteca/2011/17988.pdf

7. Uchoa, A.F., de Oliveira, K.T., Baptista, M.S., Bortoluzzi, A.J., Iamamoto, Y., Serra, O.A.: Chlorin photosensitizers sterically designed to prevent self - aggregation. J. Org. Chem. **76**(21), 8824–8832 (2011). https://doi.org/10.1021/jo201568

8. Simplicio, F.I., Maionchi, F., Hioka, N.: Terapiafoto - dinâmica: aspectos farmacológicos, aplicações e avanços recentes no desenvolvimento de medicamentos. Quim. Nova **25**(5), 801–807 (2002). https://doi.org/10.1590/s0100-40422002000500016

9. Linares, I.A.P., Martinelli, L.P., Moritz, M.N.O., de Araujo, H.S.S., de Oliveira, K.T., Rodrigues Perussi, J.: Cytotoxicity of structurally-modifiedchlorins aimed for photodynamic therapy applications. J. Photochem. Photobiol. A: Chem. **425** (2022). https://doi.org/10.1016/j.jphotochem.2021.113647

10. Bonnett, R., Martínez, B.: Photobleaching of sensitisers used in photodynamic therapy. Tetrahedron **57**(47), 9513–9547 (2001). https://doi.org/10.1016/s0040-4020(01)00952-8

11. Macdonald, I.J., Doughuerty, T.J.: Basic principles of photodynamic therapy. J. Porphyrins Phthalocyan. **05**(02), 105–129 (2001). https://doi.org/10.1002/jpp.328

12. Escudero, C., Crusats, J., Díez-Pérez, I., El-Hachemi, Z., Ribó, J.M.: Folding and hydro dynamic forces in J-aggregates of 5-phenyl - 10,15,20 – tris (4-sulfophenyl) porphyrin. Angew. Chem. Int. Edn. **45**(47), 8032–8035 (2006). https://doi.org/10.1002/anie.20060312

13. Gabrielli, D., Belisle, E., Severino, D., Kowaltowski, A.J., Baptista, M.S.: Binding, aggregation and photochemical properties of methylene blue in mitochondrial suspensions. Photochem. Photobiol. **79**(3), 227–232 (2004). https://doi.org/10.1562/be-03-27.1. PMID: 15115294

14. Kano, K., Fukuda, K., Wakami, H., Nishiyabu, R., Pasternack, R.F.: Factors influencing self-aggregation tendencies of cationic porphyrins in aqueous solution. J. Am. Chem. Soc. **122**(31), 7494–7502 (2000). https://doi.org/10.1021/ja000738g

15. Rossi, L.M., Silva, P.R., Vono, L.L., Fernandes, A.U., Tada, D.B., Baptista, M.S.: Protoporphyrin IX nanoparticle carrier: preparation, optical properties, and singlet oxygen generation. Langmuir **24**(21), 12534–12538 (2008). https://doi.org/10.1021/la800840k

# Pain Assessment and Autonomic Profile in Patients Undergoing Laparoscopic and Open Cholecystectomies: A Study of Respective Effects on Postoperative Pain

C. F. S. Guimarães[1,2] ![iD], C. M. C. Scassola[1] ![iD], B. M. Silva[2] ![iD], S. A. Miyahira[2] ![iD], O. H. M. Hypolito[2] ![iD], T. S. Cunha[1] ![iD], and K. R. Casali[1(✉)] ![iD]

[1] Department of Anesthesiology, Surgery Center and Intensive Care, Dr. José de Carvalho Florence Municipal Hospital, São José dos Campos, SP, Brazil
`rabello.casali@unifesp.br`

[2] Institute of Science and Technology, Federal University of São Paulo, São José dos Campos, SP, Brazil

**Abstract.** Cholecystectomy is very common in hospitals, being performed by open (OC) or laparoscopic (LC) procedures. Surgical and postoperative pain may affect the recovery and its assessment can provide important information about the patient's clinical condition, avoiding future complications and public health costs. This prospective observational study investigated the differences between OC and LC procedure effects on postoperative pain, by using pain scales and autonomic parameters. We evaluated the pain level and autonomic parameters in 26 adult patients undergoing cholecystectomy. The pain level was measured using the visual analog scale (VAS) and the autonomic parameters were assessed by a 10-min continuous record of the heart rate variability and blood pressure. Data was collected in three moments: before the patient entered the surgery center (M1), on the arrival at the postanesthetic care unit (M2) and three hours after M2 (M3). Data from 26 patients were included in the final analysis: LC (N = 19) or OP (N = 7). Our results suggested that LC is not advantageous over OC, considering pain management. The postoperative pain reported by LC patients was higher at M2, and they also presented an increased sympathetic activation at M3. Unexpectedly, higher pain scores, associated with sympathetic activation, were observed after laparoscopic surgery. Although LC is a minimally invasive procedure, our results showed lower pain levels in OP patients, probably due to epidural anesthesia. Our results highlight the need for reassessment of surgical and postoperative analgesic interventions in LC surgeries for a better recovery and shorter hospitalization of patients.

**Keywords:** Pain · Cholecystectomy · Heart Rate Variability · Laparoscopic Procedure · Laparotomic Procedure

# 1 Introduction

Cholecystectomy is the surgical removal of the gallbladder and is largely performed around the world [1]. Open cholecystectomy (OC), is the classical method, performed by extensive abdominal incision [2]. Laparoscopic Cholecystectomy (LC) is a minimally invasive procedure, performed using a video camera [3], requiring a specific training surgeon [4]. OC is the standard procedure, used in high complexity cases, patient's own choice, or when LC does not reach total success, due to technical difficulties or serious complications [5]. LC allows for lower morbidity and shorter length of hospital stay and recovery time [6–11]. Therefore, it has often been used in cholecystectomy procedures. However, a recurrent complaint reported by patients after LC is postoperative pain [12, 13], a relevant factor for delayed recovery [13].

Surgical and postoperative pain leads to increased heart rate and blood pressure, reduced local tissue blood flow, changes in immune response, hyperglycemia, lipolysis, increased metabolic activity and oxygen consumption [15] and may have important implications for adverse cardiovascular outcomes after non-cardiac surgeries [16]. Pain can also affect the emotional state and physical recovery of hospitalized patients, also contributing to prolonged hospitalization [17]. The use of postoperative strong injectable analgesics corroborates the need for improvement in postoperative pain management [1, 18]. Thus, pain assessment can provide important information about a patient's clinical condition, to avoid future complications and to reduce public health costs [19].

There are several methods used for pain level assessment, such as scales [20] and questionnaires [21–23]. However, the gold standard method is still a topic of debate [24, 25]. Some studies propose a pain index based on the assessment of the autonomic responses to noxious stimuli indicating analgesia/nociception balance [26], such as pupillometry [27]. In addition, evaluation of heart rate variability (HRV), related to cardiac autonomic regulation, showed a great correlation to pain level in several situations [28–31].

In this study, we used a visual analog scale (VAS) and autonomic parameters, to assess postoperative pain level, after LC or OC. Our primary hypothesis was that we could observe significant differences regarding pain level and cardiovascular autonomic control, between OC and LC procedures in the postoperative period, with increased sympathetic activation in the early period after the surgery, especially in the OC group. Our secondary hypothesis consisted in finding a possible correlation between the pain level indicated by using the VAS, and autonomic parameters, including HRV.

# 2 Methods

The protocol was approved by the Ethics Committees from Federal University of São Paulo and Dr. José de Carvalho Florence Municipal Hospital (CAAE: 53144021.1.3001.5451) and followed the ethical standards of the responsible committee on human experimentation, in accordance with the Helsinki Declaration. All patients signed an informed consent form, prior to entering the surgical area.

## 2.1  Sample

Twenty-six patients, between 18 and 60 years old, were enrolled in this study. Recruitment and selection were done at Dr. José de Carvalho Florence Municipal Hospital and the inclusion criteria were patients submitted to cholecystectomy, both sexes, physical status ASA I and II, body mass index between 18.5 and 30 kg/m$^2$, and normal renal function. The exclusion criteria were patients with prevalent hypertension, diabetes, or neurological disorders. Eventual procedure complications that could interfere with the results also resulted in exclusion from the study.

Regarding the choice of type of surgery, the study followed the procedure already adopted by the hospital, which assigns the indication to the doctor in charge, considering clinical parameters and the consent of the patient. The pharmacological and anesthetic protocol followed the institutional standard, regardless of research participation. LC patients received only general anesthesia, while OC patients received both, general and epidural anesthesia (local anesthetic and morphine), according to the hospital's current protocol.

## 2.2  Protocol

Patients were evaluated in the preoperative room at the surgical unit. Data were collected in three moments: just before the patient entered the surgery center (M1), on the arrival at the postanesthetic care unit (M2), and three hours after M2 (M3). Data included pain level, blood pressure measurement and HRV.

**Pain Assessment.**  Pain level was assessed by VAS, a straight line with equally spaced marks ranging from 0 to 10. The 0 endpoint defines the extreme limit as "no pain at all", whilst the 10 endpoint defines the extreme "the worst pain possible". The patients were asked to point their pain intensity on the line, and the indicated number defined pain level.

**Autonomic Control Assessment.**  Data from systolic blood pressure (SBP) and diastolic blood pressure (DBP) were obtained by a non-invasive blood pressure monitor. Cardiac autonomic control was obtained by a continuous and non-invasive heart rate recording signal, collected using a Polar V800 heart rate monitor, during 10 min, in the supine position. Heart rate time series (tachogram) were obtained by the intervals between consecutive RR peaks. Stationary sequences, 200–300 consecutive beats, were selected within the recorded signal, for every three moments [32].

In heart rate variability (HRV) analysis, time-domain indices quantify the successive beets intervals variability. Heart rate variance (VarHR), standard deviation (StdHR) and mean (MeanHR) represent the ECG signal features, quantified in beats per minute (bpm). Root mean square error (RMSE) was calculated to the differences between successive RR peak intervals.

Frequency-domain analysis emphasized spectral density and was performed by autoregressive algorithms. HRV spectrum consists of three bands: very low frequency (VLF), 0.00–0.04 Hz; low frequency (LF), 0.04–0.15 Hz; and high frequency (HF), 0.15–0.40 Hz. These components were expressed in absolute (abs) and normalized units (nu). The LF band represents sympathetic modulation, whilst the HF band represents parasympathetic modulation. LF/HF ratio was used to quantify cardiac sympathovagal balance [33, 34].

Data were presented as mean ± standard deviation. Statistical analysis was performed using two-way analysis of variance (ANOVA) for repeated measures, followed by Tukey honestly significant difference (HSD) multiple comparison post-hoc test. When analyzing the data from LC or OC patients, separately, statistical analysis was performed using one-way analysis of variance (ANOVA) for repeated measures, followed by Tukey honestly significant difference (HSD) multiple comparison post-hoc test. Pearson's correlation was used to assess associations among variables. $P < 0.05$ was considered significant. The statistical analysis was performed by SigmaPlot 11.0 software.

## 3   Results

Results show that during the period of this study, 54 eligible cholecystectomy procedures were performed at the Dr. José de Carvalho Florence Municipal Hospital in São José dos Campos. From those, 19 patients were excluded: 10 were older than 60 years; 1 was younger than 18 years; 3 did not show up; 2 were blind, 1 was at the intensive care unit and 1 presented chronic non-progressive encephalopathy. Recruitment has been done among the remaining 36 patients, and all patients agreed to participate. Due to surgery logistics and/or complications, 28 patients were recorded, and data from 2 of them were excluded from the final analysis due to cardiovascular signal interference. Therefore, 26 patients, between 19 and 60 years, were enrolled in the study. Statistical analysis showed no significant differences in autonomic parameters regarding sex or age. Patients were allocated into two groups, according to the surgical technique: LC (N = 19) or OC (N = 7). LC patients received only general anesthesia, while OC patients received both, general and epidural anesthesia (local anesthetic and morphine), according to the hospital's current protocol.

### 3.1   Pain Assessment

Pain level indicated by LC patients in M2 ($p < 0.001$) and M3 ($p < 0.001$) was significantly higher compared to M1, whilst OC patients reported a significantly higher pain level only in M3, compared to M1 ($p = 0.037$) (see Fig. 1).
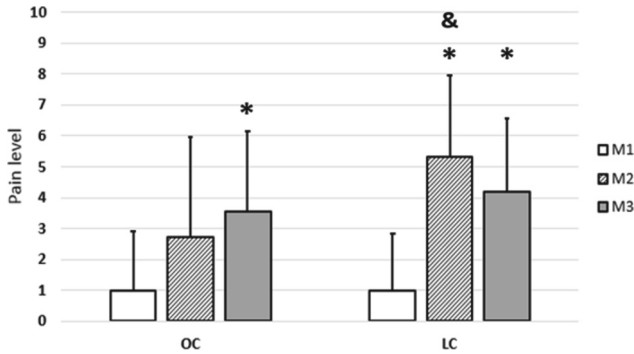
**Fig. 1.** Pain level assessment in OC and LC patients in each of the three moments recorded. Statistically significant difference compared to moment M1 from the same type of surgery (*). Statistically significant difference compared to the same moment from the other type of surgery (&).

### 3.2 Autonomic Control Assessment

Regarding time-domain analysis, pain level, VarHR, StdHR, SBP and DBP, showed significant differences among the three moments and the two groups (see Table 1). Patients from the LC group presented lower SBP ($p < 0.001$; $p < 0.001$) and DBP ($p = 0.006$; $p = 0.005$) in M2 and M3, as compared to M1. On the other hand, only SBP was lower in M2 ($p = 0.006$) and M3 ($p = 0.007$), as compared with M1, in the OC group (see Fig. 2). Also, LC patients presented higher StdHR ($p = 0.046$) in M3 compared to M1, whilst OC patients presented lower VarHR ($p = 0.024$; $p = 0.010$) and StdHR ($p = 0.003$; $p = 0.005$) in M2 and M3 compared to M1 (see Fig. 3).

### 3.3 Autonomic Control Assessment

Regarding time-domain analysis, pain level, VarHR, StdHR, SBP and DBP, showed significant differences among the three moments and the two groups (see Table 1). Patients from the LC group presented lower SBP ($p < 0.001$; $p < 0.001$) and DBP ($p = 0.006$; $p = 0.005$) in M2 and M3, as compared to M1. On the other hand, only SBP was lower in M2 ($p = 0.006$) and M3 ($p = 0.007$), as compared with M1, in the OC group (see Fig. 2). Also, LC patients presented higher StdHR ($p = 0.046$) in M3 compared to M1, whilst OC patients presented lower VarHR ($p = 0.024$; $p = 0.010$) and StdHR ($p = 0.003$; $p = 0.005$) in M2 and M3 compared to M1 (see Fig. 3).

### 3.4 Comparison Between Cholecystectomy Techniques

In M2, LC patients presented higher pain levels ($p = 0.017$) and SBP ($p = 0.007$) compared to OC patients. In M3, the VarHR ($p = 0.018$) and StdHR ($p = 0.026$) were lower in LC group, whilst SBP ($p = 0.003$) was higher, compared to OC. In frequency domain analysis, no significant differences between the techniques were detected.

**Table 1.** Autonomic cardiovascular control

| | LAPAROTOMIC CHOLECYSTECTOMY (OC) | | | LAPAROSCOPIC CHOLECYSTECTOMY (LC) | | |
|---|---|---|---|---|---|---|
| | M1 | M2 | M3 | M1 | M2 | M3 |
| Pain Level | 1.000 ± 1.915 | 2.714 ± 3.251 | **3.571 ± 2.573*** | **1.000 ± 1.826*** | **5.316 ± 2.647*&** | 4.211 ± 2.371 |
| MeanHR (bpm) | 81.167 ± 10.412 | 10.303 ± 6.477 | 74.420 ± 5.131 | 81.865 ± 13.029 | 80.209 ± 11.527 | 82.003 ± 16.228 |
| StdHR (bpm) | 3.826 ± 0.626 | 2.647 ± 1.710 | **4.090 ± 1.861#** | **3.881 ± 1.219*** | **2.748 ± 1.310*** | **2.694 ± 1.472&** |
| VarHR (bpm) | 14.976 ± 4.964 | 9.547 ± 11.227 | 19.991 ± 16.032 | **16.584 ± 10.151*** | **9.183 ± 8.221*** | **9.320 ± 9.193&** |
| SBP (mmHg) | 136.86 ± 11.481 | **102.57 ± 15.704*** | **102.14 ± 11.276*** | **134.16 ± 19.074*** | **121.37 ± 13.969*&** | **123.05 ± 13.786&** |
| DBP (mmHg) | 79.571 ± 12.15 | **63.714 ± 14.139*** | **61.429 ± 10.784*** | 77.105 ± 10.806 | 69.842 ± 14.155 | 70.316 ± 11.638 |
| LFnu | 58.934 ± 15.685 | 47.079 ± 26.679 | 57.864 ± 26.236 | 64.886 ± 14.580 | 64.241 ± 28.287 | 57.193 ± 17.939 |
| HFnu | 41.066 ± 15.685 | 52.921 ± 26.679 | 42.136 ± 26.236 | 35.114 ± 14.580 | 35.759 ± 28.827 | 42.807 ± 17.939 |
| LF/HF | 1.769 ± 1.053 | 1.367 ± 1.194 | 4.266 ± 6.899 | 2.376 ± 1.528 | 5.830 ± 9.984 | 1.833 ± 1.404 |

Data presented as mean ± standard deviation. Significant differences were tested by one-Way ANOVA for repeated measurements, followed by Tukey test. Statistically significant difference compared to M1 (*) and M2 (#), in the same group. Statistically significant difference comparing the same moment, between the two groups (&). Statistically significant differences in bold. Heart rates mean (MeanHR), heart rate standard deviation (StdHR), heart rate variance (VarHR), systolic blood pressure (SBP), diastolic blood pressure (DBP), low frequency components normalized values (LF nu), high frequency components normalized values (HF nu) and low and high frequency components ratio (LF/HF). LC (N = 19) and OC (N = 7)
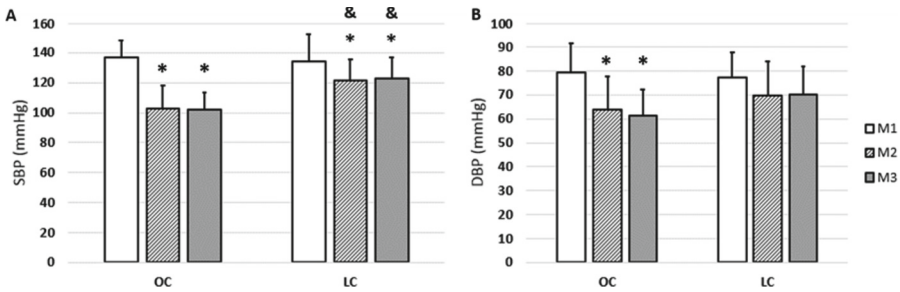


**Fig. 2.** Systolic (A) and Diastolic (B) Blood Pressure measures in OC and LC patients in each of the three moments recorded. Statistically significant difference compared to moment M1 from the same type of surgery (*). Statistically significant difference compared to the same moment from the other type of surgery (&).
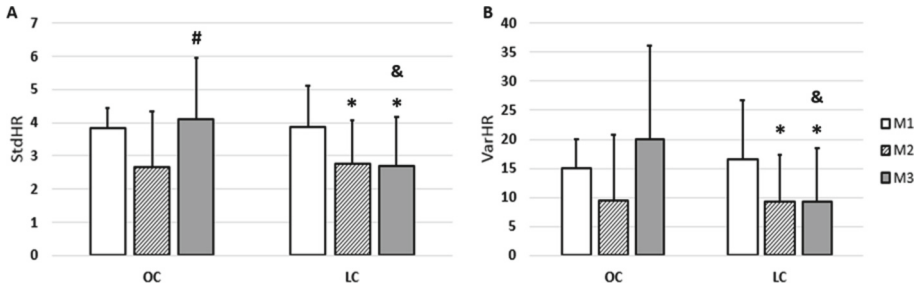
**Fig. 3.** Heart Rate standard deviation (A) and variance (B) measures in OC and LC patients in each of the three moments recorded. Statistically significant difference compared to moment M1 (*) and M2 (#) from the same type of surgery (*). Statistically significant difference compared to the same moment from the other type of surgery (&)

**Table 2.** Autonomic cardiovascular control

|  | ALL | OC | LC |
|---|---|---|---|
| RMSSD (ms) | **−0.276 (p = 0.0144)** | −0.00249 (p = 0.991) | **−0.349 (p = 0.00776)** |
| StdHR (bpm) | **−0.269 (p = 0.0174)** | 0.00646 (p = 0.978) | **−0.356 (p = 0.0078)** |
| VarHR (bpm) | **−0.251 (p = 0.0265)** | 0.032 (p = 0.886) | **−0.346 (p = 0.0084)** |
| LF abs ($mmHg^2$) | **−0.270 (p = 0.0168)** | −0.168 (p = 0.466) | **−0.399 (p = 0.00992)** |
| HF abs ($mmHg^2$) | −0.185 (p = 0.105) | 0.0615 (p = 0.791) | **−0.277 (p = 0.0371)** |

Pearson's correlation between HRV and pain level in: all patients (ALL), OC (open cholecystectomy) patients and LC (laparoscopic cholecystectomy) patients. Statistically significant differences in bold. Root mean square error of successive difference (RMSE), heart rate standard deviation (StdHR), heart rate variance (VarHR), absolute values of low frequency components (LF abs) and absolute values of high frequency components (HF abs).

### 3.5 Correlation Between Indexes

Pain level was inversely correlated to RMSE (p = 0.00776), VarHR (p = 0.0078), StdHR (p = 0.0084), LF abs (p = 0.00992) and HF abs (p = 0.0371) in LC group, whilst no correlation between the variables was detected in OC group (see Table 2).

## 4  Discussion

Our results suggest that LC is not advantageous over OC concerning pain management, regarding protocols used in this research. Even though LC is a minimally invasive procedure, postoperative pain reported by these patients was higher than OC patients report. This phenomenon was elicited not only by pain data, but also by autonomic parameters assessment.

In this study, the pain level reported by LC patients in M2 and M3 was significantly higher than in M1, whilst the OC patients reported a significantly higher pain level only

in M3 compared to M1. These results confirmed higher pain levels shortly after the LC procedure, compared to OC, as also reported by other authors [35, 36]. M3 pain was similar in both groups.

This difference can be attributed to procedure intrinsic factors, such as the visceral and somatic components present in LC postoperative pain, as well as to epidural anesthesia performed only in the OP group. The neuroendocrine stress response is a major event during laparoscopy, as tissue manipulation and pneumoperitoneum result in intense stimulation to the nervous system, leading to corticosteroid and catecholamine release [37]. Some components are related to shoulder pain and diaphragmatic irritation caused by residual gas that is used to perform pneumo-peritoneum in laparoscopic procedures [38]. In this way, some studies had demonstrated that intraperitoneal application of anesthetic drugs (such as morphine, ropivacaine or bupivacaine) during LC, can significantly reduce abdominal pain scores in the first postoperative hours, when visceral pain incidence is higher [39, 40]. However, this option does not completely treat postoperative pain [36]. Besides that, laparoscopic postoperative pain is due to other influences, such as gas humidity and temperature, residual gas volume and pneumo-peritoneum intraperitoneal pressure level [41]. So, differences between techniques and their respective postoperative pain pathophysiological mechanisms, make it difficult to accurately compare them.

Regarding autonomic control, HRV (VarHR and StdHR) was lower at M2 and M3 compared to M1, only in LC patients, and it is related to sympathetic activation [33], corroborated by higher SBP values. Besides that, sympathetic activation, shown by M3 LC values, was higher compared to M3 OC values.

Studies have shown that painful conditions are related to sympathetic activation. In fact, our results demonstrated a negative correlation between HRV data analysis and pain level, only in the LC group. The most important finding in our study was the higher pain level reported by LC patients, and respective autonomic features, in the early postoperative hours, indicating sympathetic activation.

We admit as an important limitation of this study, the unevenness regarding sex and age among the patients. Since the risk factors for developing gallstones include being female, prior pregnancy, age over 40 years and being overweight [42], the predominance of patients with these characteristics was already expected. Almost 90% of the patients were women (N = 23) and more than 60% were over 40 years old (N = 16). In addition, our analysis could be more consistent if we had a larger sample size. Besides that, additional studies are needed to evaluate the association between autonomic parameters and pain in patients with cardiological disorders or under treatments that may interfere with autonomic responses to surgical stress.

## 5  Conclusions

This study compared postoperative pain level and autonomic parameters features in patients submitted to cholecystectomy performed by laparoscopy and open techniques. Unexpectedly, higher pain scores, associated with sympathetic cardiac activation, were observed after laparoscopic surgery. Although LC is a minimally invasive procedure, our results showed lower pain level in OP patients, probably due to epidural anesthesia.

Data showing beneficial physiological effects, obtained by epidural anesthesia on cardio-vascular, respiratory and gastrointestinal systems are consistent. Surgical stress patho-physiology is derived from different factors that impact the recovery process. Regarding pain and autonomic system management, epidural anesthesia remains a powerful available tool aiming at reduced postoperative dysfunctions and enhanced rehabilitation. Our results highlight that perioperative pain management in laparoscopic surgeries needs reassessment, improving recovery conditions and hospitalization time.

As limitations of this study, we point out the sample size and the imbalance between the two groups. This difference is due to the priority given to less invasive surgical methods.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. Singh, R.R., Bansal, D., Singh, M., et al.: Randomized controlled study of effectiveness and timings of bupivacaine at port site in managing pain after laparoscopic cholecystectomy. Int. Sur. J. **3**(4), 1837–1843 (2016). https://doi.org/10.18203/2349-2902.isj20162888
2. Mimica, Ž, Biočić, M., Banović, I., et al.: Laparoscopic and laparotomic cholecystectomy: a randomized trial comparing postoperative respiratory function. Respiration **67**, 153–158 (2000). https://doi.org/10.1159/000029479
3. Reynolds, W., Jr.: The first laparoscopic cholecystectomy. JSLS **5**, 89–94 (2001)
4. Fielding, G.A.: Laparoscopic cholecystectomy, open cholecystectomy and cholecystostomy. In: Clavien, P.A., Sarr, M.G., Fong, Y., Georgiev, P. (eds.) Atlas of Upper Gastrointestinal and Hepato-Pancreato-Biliary Surgery. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-68866-2_59
5. Goyal, V., Nagpal, N., Gupta, M., et al.: A prospective study to predict the preoperative risk factors for conversion of laparoscopic to open cholecystectomy. J. Contemp. Med. Surg. Radiol. **2**(4), 148–152 (2017)
6. Lujan, J.A., Parrilla, P., Robles, R., et al.: Laparoscopic cholecystectomy vs open cholecystectomy in the treatment of acute cholecystitis: a prospective study. Arch. Surg. **133**, 173–175 (1998). https://doi.org/10.1001/archsurg.133.2.173
7. Sauerland, S., Agresta, F., Bergamaschi, R., et al.: Laparoscopy for abdominal emergencies. Surg. Endosc. **20**, 14–29 (2006). https://doi.org/10.1007/s00464-005-0564-0
8. Comitalo, J.B.: Laparoscopic cholecystectomy and newer techniques of gallbladder removal. JSLS **16**, 406–412 (2012). https://doi.org/10.4293/108680812X13427982377184
9. Coccolini, F., Catena, F., Pisano, M., et al.: Open versus laparoscopic cholecystectomy in acute cholecystitis. Systematic review and meta-analysis. Int. J. Surg. **18**, 196–204 (2015). https://doi.org/10.1016/j.ijsu.2015.04.083
10. Kurtulus, I., Culcu, O.D., Degerli, M.S.: Which is more effective: laparoscopic or open partial cholecystectomy? J. Laparoendosc. Adv. Surg. Tech. A **32**(5), 476–484 (2022). https://doi.org/10.1089/lap.2021.0300

11. Song, J., Chen, J., Lin, C.: Therapeutic effect of laparoscopic cholecystectomy on patients with cholecystolithiasis complicated with chronic cholecystitis and postoperative quality of life. Evid. Based Complement Alternat. Med. **2022**, 6813756 (2022). https://doi.org/10.1155/2022/6813756

12. McMahon, A.J., Russell, I.T., Ramsay, G., et al.: Laparoscopic and minilaparotomy cholecystectomy: a randomized trial comparing postoperative pain and pulmonary function. Surgery **115**(5), 533–539 (1994)

13. Ure, B.M., Troidl, H., Spangenberger, W., et al.: Pain after laparoscopic cholecystectomy. Surg. Endosc. **8**, 90–96 (1994). https://doi.org/10.1007/BF00316616

14. Callesen, T., Klarskov, B., Mogensen, T.S., et al.: Ambulatory laparoscopic cholecystectomy Feasibility and convalescence. Ugeskrift for Laeger **160**(14), 2095–2100 (1998)

15. Chang, L.H., Ma, T.C., Tsay, S.L., et al.: Relationships between pain intensity and heart rate variability in patients after abdominal surgery: a pilot study. Chin. Med. J. **125**(11), 1964–2196 (2012)

16. So, V., Klar, G., Leitch, J., McGillion, M., et al.: Association between postsurgical pain and heart rate variability: protocol for a scoping review. BMJ Open **11**(4), e044949 (2021). https://doi.org/10.1136/bmjopen-2020-044949

17. Pavlin, D.J., Chen, C., Penaloza, D.A., et al.: Pain as a factor complicating recovery and discharge after ambulatory surgery. Anesth. Analg. **95**, 627–634 (2002). https://doi.org/10.1097/00000539-200209000-00025

18. Bisgaard, T., Klarskov, B., Rosenberg, J., et al.: Characteristics and prediction of early pain after laparoscopic cholecystectomy. Pain **90**, 261–269 (2001). https://doi.org/10.1016/S0304-3959(00)00406-1

19. Kreling, M.C.G.D., Cruz, D.A.L.M., Pimenta, C.A.M.: Prevalência de dor crônica em adultos. Rev. Bras. Enferm. **59**(4), 509–513 (2006). https://doi.org/10.1590/S0034-71672006000400007

20. Pereira, L.V., Sousa, F.A.E.F.: Mensuração e avaliação da dor pós-operatória: uma breve revisão. Rev. Latino-Am. Enf. **6**(3), 77–84 (1998). https://doi.org/10.1590/S0104-11691998000300010

21. Melzack, R.: The McGill pain questionnaire: major properties and scoring methods. Pain **1**(3), 277–299 (1975). https://doi.org/10.1016/0304-3959(75)90044-5

22. Daut, R.L., Cleeland, C.S., Flanery, R.C.: Development of the Wisconsin brief pain questionnaire to assess pain in cancer and other diseases. Pain **17**, 197–210 (1983). https://doi.org/10.1016/0304-3959(83)90143-4

23. Fishman, B., Pasternak, S., Wallenstein, S.L., et al.: The memorial pain assessment card. Cancer **60**, 1151–1158 (1987). https://doi.org/10.1002/1097-0142(19870901)60:5%3c1151::aid-cncr2820600538%3e3.0.co;2-g

24. Boselli, E., Daniela-Ionescu, M., Bégou, G., et al.: Prospective observational study of the non-invasive assessment of immediate postoperative pain using the analgesia/nociception index (ANI). Br. J. Anaesth. **111**(3), 453–459 (2013). https://doi.org/10.1093/bja/aet110

25. Boselli, E., Bouvet, L., Bégou, G., et al.: Prediction of immediate postoperative pain using the analgesia/nociception index: a prospective observational study. Br. J. Anaesth. **112**(4), 715–721 (2014). https://doi.org/10.1093/bja/aet407

26. Anderson, T.A., Segaran, J.R., Toda, C., et al.: High-frequency heart rate variability index: a prospective, observational trial assessing utility as a marker for the balance between analgesia and nociception under general anesthesia. Anesth. Analg. **130**(4), 1045–1053 (2019). https://doi.org/10.1213/ANE.0000000000004180

27. Larson, M.D., Beherenas, M.: Portable infrared pupillometry: a review. Anesth. Analg. **120**(6), 1242–1253 (2015). https://doi.org/10.1213/ane.0000000000000314

28. Appelhans, B.M., Luecken, L.J.: Heart rate variability and pain: associations of two interrelated homeostatic processes. Biol. Psychol. **77**(2), 174–182 (2008). https://doi.org/10.1016/j.biopsycho.2007.10.004

29. Koenig, J., Jarczok, M.N., Ellis, R.J., et al.: Heart rate variability and experimentally induced pain in healthy adults: a systematic review. Eur. J. Pain **18**(3), 301–314 (2014). https://doi.org/10.1002/j.1532-2149.2013.00379.x

30. Broucqsault-Dédrie, C., Jonckheere, J.D., Jeanne, M., et al.: Measurement of heart rate variability to assess pain in sedated critically ill patients: a prospective observational study. PLoS ONE **11**(1), e0147720 (2016). https://doi.org/10.1371/journal.pone.0147720

31. Koenig, J., Loerbroks, A., Jarczok, M.N., et al.: Chronic pain and heart rate variability in a cross-sectional occupational sample. Clin. J. Pain **32**(3), 218–225 (2016). https://doi.org/10.1097/AJP.0000000000000242

32. Porta, A., Addio, G.D., Guzzetti, S., et al.: Testing the presence of nonstationarities in short heart rate variability series. Comput. Cardiol. **31**, 645–648 (2004). https://doi.org/10.1109/CIC.2004.1443021

33. Malliani, A., Pagani, M., Lombardi, F., et al.: Cardiovascular neural regulation explored in the frequency domain. Circulation **84**(2), 482–492 (1991). https://doi.org/10.1161/01.cir.84.2.482

34. Montano, N., Porta, A., Cogliati, C., et al.: Heart rate variability explored in the frequency domain: a tool to investigate the link between heart and behavior. Neurosci. Biobehav. Rev. **33**, 71–80 (2009). https://doi.org/10.1016/j.neubiorev.2008.07.006

35. Ekstein, P., Szold, A., Sagie, B., et al.: Laparoscopic surgery may be associated with severe pain and high analgesia requirements in the immediate postoperative period. Ann. Surg. **243**(1), 41–46 (2006). https://doi.org/10.1097/01.sla.0000193806.81428.6f

36. Szental, J.A., Webb, A., Weeraratne, C., et al.: Postoperative pain after laparoscopic cholecystectomy is not reduced by intraoperative analgesia guided by analgesia/nociception index (ANI®) monitoring: a randomized clinical trial. Br. J. Anaesth. **114**(4), 640–645 (2015). https://doi.org/10.1093/bja/aeu411

37. Das, W., Bhattacharya, S., Ghosh, S., et al.: Comparison between general anesthesia and spinal anesthesia in attenuation of stress response in laparoscopic cholecystectomy: a randomized prospective trial. Saudi J. Anaesth. **9**(2), 184–188 (2015). https://doi.org/10.4103/1658-354X.152881

38. Bisgaard, T.: Analgesic treatment after laparoscopic cholecystectomy: a critical assessment of the evidence. Anesthesiology **104**(4), 835–846 (2006). https://doi.org/10.1097/00000542-200604000-00030

39. Albuquerque, T.L.C., Bezerra, M.F., Schots, C.C.P.Z., et al.: Evaluation of postoperative analgesia with intraperitoneal ropivacaine instillation in video laparoscopic cholecystectomy. Rev. Dor. **17**(2), 117–120 (2016). https://doi.org/10.5935/1806-0013.20160027

40. Hernández-Palazón, J., Tortosa, J.A., Nuño de la Rosa, V., et al.: Intraperitoneal application of bupivacaine plus morphine for pain relief after laparoscopic cholecystectomy. Eur. J. Anaesth. **20**(11), 891–896 (2003). https://doi.org/10.1017/s0265021503001431

41. Mouton, W.G., Bessel, J.R., Otten, K.T., et al.: Pain after laparoscopy. Surg. Endosc. **13**(5), 445–448 (1999). https://doi.org/10.1186/s12905-017-0397-8

42. Thesbjerg, S.E., Harboe, K.M., Bardram, L., et al.: Sex differences in laparoscopic cholecystectomy. Surg. Endosc. **24**(12), 3068–3072 (2010). https://doi.org/10.1007/s00464-010-1091-1

# Physiological and Mobility Monitoring System for Patients with Lower Limb Amputation Based on a Serious Virtual Reality Game with an Instrumented Trike

Jonathan Gallego-Londoño[✉] , Julián Pineda-Escobar ,
Manuela Gómez-Correa , and Sofía Agudelo-Zapata

Bioinstrumentation and Clinical Engineering Research Group - GIBIC,
Bioengineering Department, Engineering Faculty, Universidad de Antioquia UdeA,
Calle 70 No. 52-21, Medellín, Colombia
jonathan.gallelgo@udea.edu.co

**Abstract.** This project developed a functional training system for patients with lower limb prostheses. The system uses a virtual reality serious game controlled by a self-instrumented recumbent trike, and a self-designed physiological monitoring module to monitor respiratory rate, heart rate, and pressure distribution. The data collected is stored in a database and processed by a central control system using a Raspberry-Pi, which allows parallel visualization in a web interface. The system was designed to be used as a diagnostic tool to generate useful data for future research and for the adaptation period of people with prostheses to complement rehabilitation. To test the system, a conceptual test was carried out with 12 volunteers, which involved adjusting the trike for users, collecting physiological and mechanical variables, collecting user data, playing the game in tutorial and competitive modes, and evaluating the usability and immersion of the system. The test revealed that the system has the characteristics of allowing an adequate and natural simulation of reality, which makes it easy to learn for different types of audiences. The system's use of virtual reality and physiological monitoring allows amputee patients to engage in high-intensity physical activities and improve performance in daily life and work. This, in turn, increases adherence to the use of prostheses and helps reduce the risk of heart disease and other chronic conditions. Overall, the functional training system has the potential to be a valuable tool in rehabilitation and research for amputee patients.

**Keywords:** lower limb prosthesis rehabilitation · integral rehabilitation plan · serious virtual reality game · instrumented recumbent trike · physiological monitoring

# 1 Introduction

Lower limb prostheses seek to reduce the radical impact on the lifestyle of a person with an amputation, helping the users to return to their daily activities, social reintegration, and avoid health problems due to the load supported on a single member [1,2]. However, lack of adherence to the use of prostheses is common [3,4] and its abandonment can cause serious health issues due to problems such as physical inactivity and obesity [5]. This is one of the reasons why integral rehabilitation plans where patients learn to use the prosthesis are created. These plans should contain activities of daily living and work, motivational and psychosocial interventions [6], and high-intensity physical activities for physical and cardiopulmonary strengthening since they have been associated with less athletic and functional restriction, and greater satisfaction with the rehabilitation process [3,7].

A method used as a rehabilitation tool is the implementation of video games to reproduce real life experiences in a controlled environment [8], providing motivational benefits and allowing patients to reproduce more repetitions of the exercise before reporting fatigue [9,10].

With the idea of complementing rehabilitation plans to achieve greater integrality and help amputee patients to adhere better to prostheses, the following question has arisen: How could a system be built to help with prosthetic training, provide solutions to specialists, and contribute to research on rehabilitation progress? This article shows the design process of a multipurpose virtual simulation tool that complements physical training during the rehabilitation of patients with transtibial or transfemoral prostheses through the integration of several systems.

These systems are composed of a recumbent trike that is statically fixed to a roller allowing free pedaling without displacement and instrumented with multiple sensors that capture speed, cadence, steering angle, and braking. It also has a self-manufactured physiological monitoring subsystem designed to measure heart rate, respiratory rate, and pressure distribution of the person's weight on the seat of the trike. In addition, the system has a central control device that processes and sends the measured information to a 3D virtual reality video game allowing the user to navigate within different scenarios, and it has an interface that allows the remote and wireless visualization of all the variables and the adjustment of parameters from any device that has access to the same wireless network. Additionally, a database was designed to store personal user information, kinetic and physiological variables gathered during the game, configuration parameters, and the score recorded.

Finally, a proof of concept with 12 participants was carried out to verify the operation, usability, and usefulness of the system as a possible tool in a rehabilitation plan.

## 2   Methodology

### 2.1   Recumbent Trike

A Hangar 99 TM recumbent trike was used for this project because it increases the user's safety since it is more stable than a conventional bicycle and might allow the progressive improvement of functional movements such as getting up or climbing stairs [11]. It has a 350W electric motor that assists pedaling, a 48.8V lithium-ion battery that is used as the power source of the components involved in the system, disc brakes, a back tire with 29-inch wheels, two front tires with 24-inch wheels and an LCD screen that allows the visualization of the speed and the percentage of battery charge (see Fig. 1). Additionally, a Prodalca S.A.S® roller support was installed to fix the trike statically while allowing it to roll freely.
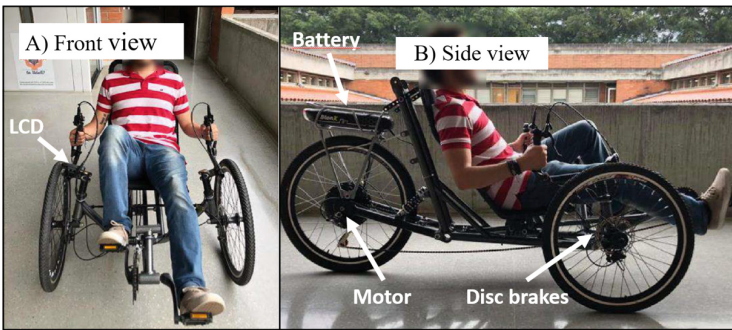


**Fig. 1.** Recumbent trike with a 350W electric motor

### 2.2   Central Control System

A Raspberry-Pi 4 with 4GB of RAM and a quad-core processor at 1.5GHz speed was chosen as the central control system. It was programmed using Node-RED to receive data collected by all the physiological and kinetic sensors. The central control system processes and cleans the signals, exports data to the virtual video game, hosts a web page for control and visualization, and stores user's information and game score inside a dedicated database.

### 2.3   Database

All the information collected by the central control system is stored in an SQLite database. This database consists of 3 tables; the first contains the access credentials, encrypted with the bcrypt protocol, that allows an administrator to load and store information of the users that perform the tests. The second table

contains the user's ID and a column that stores data in JSON format where personal information data are specified, such as name, type of prosthesis, age, and other relevant information for specialists. The third table contains a column with the user's ID, the date and name of the test, and a JSON type column with the results of the tests. In this field, all data related to the use of the trike and game are stored, including the physiological, kinetic, and configuration values. Multiple tests can be associated with the same user (see Fig. 2).
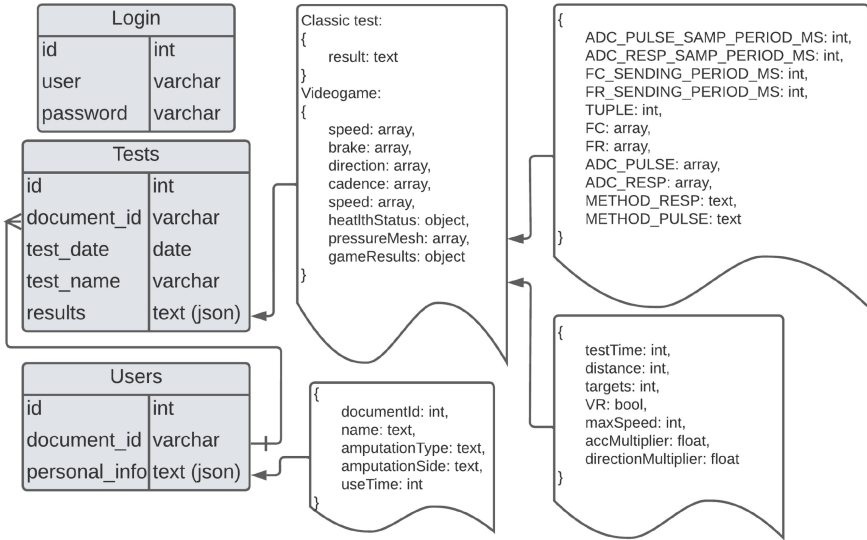


**Fig. 2.** Database composition

## 2.4  Web Page

A web page was designed using Node-RED and hosted inside the Raspberry-Pi. A user administrator can log in to this web page to create or upload patients and store new test data. Once the patient is loaded, the administrator will be able to run a manual or a video-game-controlled test. In the end, the captured data are stored in the database. All the kinetic and physiological variables collected by the sensors can be visualized in real-time graphs facilitating the comprehension of what happens during tests. Furthermore, it has a configuration interface where the administrator can adjust communication and tuning parameters for the calculation of system variables. Another option in the interface allows the user to log in to create or upload patients and store new test data.

## 2.5  Kinetic Variables

*Speed and cadence:* Two TCRT 5000 photo-reflector sensors composed of an infrared LED and a phototransistor were chosen to measure the rotational speed

of the rear wheel and the pedaling cadence. One was installed on the frame of the trike very near to the back wheel and another one was attached to the frame of the trike at the level of the pedals to detect a marker located on the left pedal crank. It was necessary to use 3D printed coupling systems to fix them. They were connected to the digital pins of the Raspberry-Pi to compute the speed and cadence. To find the speed ($v$) in real-time is calculated the relationship between the distance ($d$) of some markers located equidistantly on the rim and the average of the time ($t$) that the sensor takes to detect them. The speed is considered $0\,\mathrm{m/s}$ if the time between detections exceeds 500 milliseconds (1). The calculation of the cadence ($c$) is done by averaging the time ($t$) that the sensor takes to detect an interruption. If the time is greater than $3.5\,\mathrm{s}$, it is considered a cadence of 0 rpm (2). Those limits can be adjusted from the interface based on the number of markers.

$$v[m/s] = \begin{cases} \frac{d}{\Delta t}, \ t < 0.5\,s \\ 0, \ t \geq 0.5\,s \end{cases} \tag{1}$$

$$c[rpm] = \begin{cases} \frac{60}{\Delta t}, \ t < 3.5\,s \\ 0, \ t \geq 3.5\,s \end{cases} \tag{2}$$

*Steering angle:* A mechanism consisting of a mobile bar was designed, 3D printed, and then joined to a precision potentiometer at the pivot of said bar. The mechanism was fixed to the frame and to the steering rod of the trike to measure the steering angle. Then, the output of the potentiometer was connected to an ADS115 16-bit analog-to-digital converter and then to the Raspberry-Pi using the I2C communication module. The relationship between the voltage ($V_{pot}$) and the angle of the potentiometer ($g_{pot}$) was completely linear and its slope ($m$) and intercept ($b$) can be adjusted in real time using the interface (3). However, the relationship found after a calibration process between the angle of the potentiometer ($g_{pot}$) and the steering angle of the trike ($g_{bar}$) was not linear and followed the behavior shown in (4) (see Fig. 3).

$$g_{pot}[^{o}] = m \cdot V_{pot} + b \tag{3}$$
$$g_{bar}[^{o}] = 0,002 \cdot g_{pot}^{3} + 0,005 \cdot g_{pot}^{2} + 0,0845 \cdot g_{pot} \tag{4}$$

*Brake:* To detect the braking action, a piezoresistive flex sensor was placed on the right brake lever by making a bridge with the handlebar. It was connected to a $3.3\,\mathrm{V}$ source and an $82\,\mathrm{k\Omega}$ resistor to generate a voltage divider. The output voltage was digitalized by the ADS1115 analog-to-digital converter and then connected via I2C to the Raspberry-Pi. The braking action ($F$) was calculated, considering the hysteresis, as a percentage of the ratio between the voltage obtained in real time ($V_r$) and the difference between the maximum voltage ($V_M$) when the brake is fully pressed and the minimum voltage ($V_m$) when the brake is released (5).

$$F[\%] = \begin{cases} \left(\frac{V_r - V_m}{V_M - V_m}\right), V_m \leq V_r \leq V_M \\ 0, V_r < V_m \\ 100, V_r > V_M \end{cases} \tag{5}$$
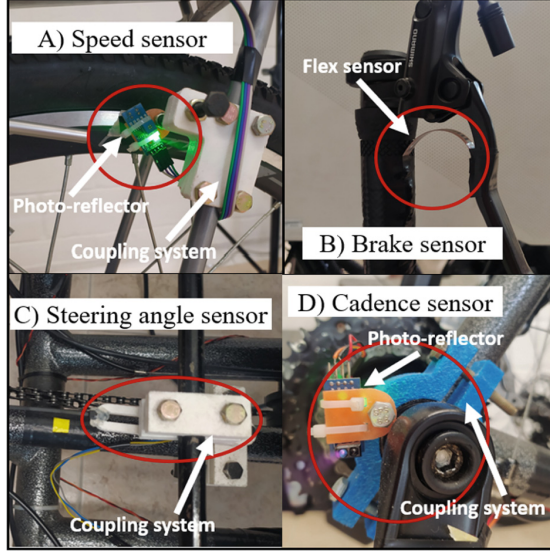


**Fig. 3.** Sensor's distributions for kinetic variables

## 2.6   Physiological Variables

*Plethysmographic signal and heart rate:* Analog and digital signal conditioning of the physiological variables as well as the communication to the central control system was done using a CY8C5888LTI-LP097 PSoC 5 with internal configurable analog and digital peripherals. It was implemented a PulseSensor® from World Famous Electronics IIc located on the thumb of the patient's left hand to obtain the plethysmographic signal. This signal was captured and filtered with an internal high-pass module with a cutoff frequency of $0.4\,\mathrm{Hz}$ and then digitally converted with a 12-bit SAR ADC module. The digital signal was smoothed with a complementary digital filter. Also, to calculate the smoothed values $(S_i)$, the raw value $(x_i)$ is multiplied by a weight constant $(a)$ found experimentally and then added to the complementary value $(1 - a)$ multiplied by the previous raw value $(x_{i-1})$ as shown in (6). A threshold level $(U)$ was implemented to calculate the heart rate from the plethysmographic signal. This threshold can be automatically set by the calculation of an average of $(S_i)$ for $3\,\mathrm{s}$ (7) or through

an external comparison value obtained from a potentiometer. Every time event at which the signal $(S_i)$ exceeds this threshold is recorded. This event occurs only once and is considered the period $(p)$. The average period $(\Delta p)$ is measured by means of a moving average window of 10 records achieving better precision. The heart rate $(f_h)$ will be the inverse of this value multiplied by $60\,\mathrm{s}$ to report the magnitude in beats per minute as shown in (8). The smoothed signal $(S_i)$ and the heart rate $(f_h)$ are sent to the Raspberry-Pi at 57600 bauds through an UART connection for visualization in real time. It is important to clarify that the communication speed is different from the sending frequency which can be modified from the interface.

$$S_i = a \cdot x_i + (1 - a) \cdot x_{i-1} \tag{6}$$

$$U = \frac{\sum_{i=0}^{n} S_i}{n} \textit{for all } t < 3\,s \tag{7}$$

$$f_h = \frac{1}{\Delta p} \cdot 60 \tag{8}$$

*Breathing frequency:* A non-invasive method for capturing volumetric chest changes during breathing was self-designed which consists of a belt placed around the patient's chest equipped with another flex sensor, a spring, a preconditioning circuit, and the PSoC. When the patient breaths, a volumetric signal is generated and then filtered and smoothed following the same procedure implemented with the plethysmographic signal as in (6). The threshold is similarly calculated as in (7) but for $6\,\mathrm{s}$. The breathing frequency $(f_b)$ is also calculated as in (8) with a moving average window but only using 3 periods of the respiratory cycle since the signal is much slower.

*Weight distribution:* A flexible mesh of $13 \times 14$ pressure sensors was fabricated following the methodology proposed by Diego A. Valle et al. [12] and was fixed to the tricycle seat. An ESP32 microcontroller was used for recording and sending the data due to its high processing speed. In this microcontroller, the raw values go through an analog to digital conversion of 8 bits to send less information but faster. The data collected from all sensors that make up the cushion are sent to the Raspberry-Pi through an UART communication at 115200 bauds. The process described above takes around $0.6\,\mathrm{s}$ (see Fig. 4).

## 2.7   Virtual Reality Video Game

The main objective of the designed virtual reality video game was to simulate the real-life activity of riding a recumbent trike. The use of the VR headset allowed the immersion of the user in the game but was not mandatory, allowing the user to continue playing if presented dizziness. The game was developed with Unity TM, and the minimums requirements for functioning were Windows 10, at least 12 GB of RAM, Intel Core i5-7500 or Ryzen 5 1600 processor, GTX 1060 or RX 580 graphics card with 6 GB of VRAM and an HTC Vive, Oculus Rift or Oculus Quest virtual reality headset.
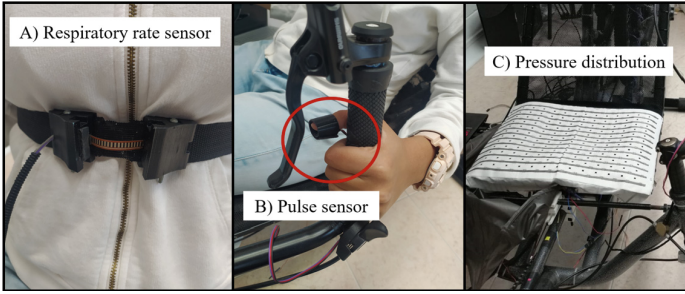
**Fig. 4.** Sensors for physiological variables

The video game is made up of 3 tests: a tutorial, a free mode for practice, and the main test, which consists of driving along a track while collecting different elements. In the latter, the game stores the number of elements obtained, the distance traveled and the average speed over 3 min, a time suggested by specialists, but that can be modified for future experimentation. These data might be used to analyze its correlation with standard mobility tests such as the Time Up and Go and the 2-minute walk test and thus determine a way to calculate an index that could indicate the degree of mobility of a user. All tracks and environments were acquired from the Unity Package Store under a free license. The Vehicle Tools package from Unity Technologies was used for the tutorial. The Lake Race Track by AndreiNi was used in the competitive test and the Mountain Race Tracks by AndreiNi and Shahbaz were used for the free mode.

An anti-sickness system that is often used in the VR industry was implemented in all game modes, proving its effectiveness experimentally. This consists of deactivating the visualization in the event of shocks and restarting the position of the vehicle deleting the discrepancy between what is seen and felt [13]. From the main menu, it is possible to specify the duration of the test, activate virtual reality, set the tunning parameters of speed, brake, and turn to improve the driving experience, and specify communication parameters such as port and baud rate.

For the interaction with the Raspberry-Pi, an FT232 USB UART communication module was used (see Fig. 5). Once the game has started, it initiates the connection and sends a specific message allowing the Raspberry-Pi to start sending the kinetic data of the trike every 20 milliseconds so that the user can control the game with the movement of the trike. Once the test is started, the video game sends a key message to temporarily store data. At the end of the session another key message is sent, and the user can confirm or decline the permanent storage. In case of a permanent storage confirmation, the game sends the test results and a password to the central control system and then the temporarily data and game statistics are permanently saved in the database.
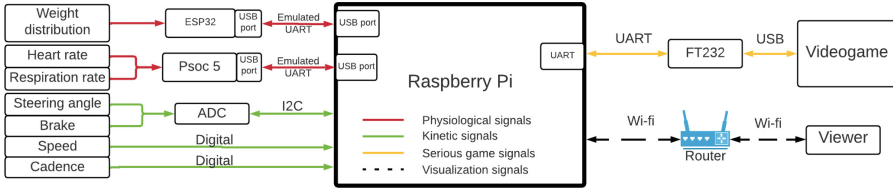
**Fig. 5.** Block diagram of the whole designed system

## 2.8   Concept Test

Once the design of the trike and the virtual reality video game were validated by specialists from the Mahavir Kmina Corporation to test the performance of the system, a usability and immersion test was performed on 12 volunteers (4 women and 8 men) at the corporation located in La Estrella, Antioquia. The age of the study participants ranged from 19 to 53 years (31±10), 2 volunteers had transtibial lower limb prostheses and the others were considered healthy. For inclusion, volunteers had to be in good general cardiovascular health. Also, potential participants were excluded in case of extremely dizzy feelings or drug abuse. This study was conducted under the accepted ethical standards for research practice as described in the Helsinki Declaration of 1975, as revised in 2000 and 2008. The study protocol underwent review and approval by the Ethics Committee of the Faculty of Medicine at Universidad de Antioquia. Written informed consent was obtained from all participants prior to enrolment. The test consisted of the following steps. The user administrator added the volunteer's personal information to the database using the website, then, the volunteer was placed on the trike, and it was adjusted adequately until the user felt comfortable. The physiological sensors and the virtual reality headset were attached to the volunteer, and the administrator performed the adjustments of thresholds. The volunteer could now start the video game tutorial, but if the person got dizzy, then the game was set to not use VR and the volunteer could repeat the tutorial or quit the test. Once the tutorial was finished, the videogame test was executed, and the user had to drive along a track and collect different elements for 3 min. During the test, the plethysmographic signal, heart rate, breathing frequency, and weight distribution were recorded and saved in the database for further studies. In the end, the volunteer was given a usability survey based on the system usability scale [14] and a presence and immersion questionnaire [15] which allowed to validate that the system was working correctly and the video game complied with the gameplay and immersion parameters.

## 3   Results

The resulting integrated system shown in Fig. 6 occupies an area of $2 \times 3$ m2. The cables and sensors were firmly attached to the frame of the trike so there were no misalignment problems despite the strong vibration produced during

pedaling. Finally, a protective box was built for the control center, increasing the security of the devices, and avoiding involuntary disconnections. For a better appreciation of the result visit: https://youtu.be/49tSDCCW3Eo

The interface allowed the correct visualization not only of the plethysmographic and the chest volume signals in real-time, but also the pressure distribution color map, the heart rate, and respiratory frequency values. However, these values should not be used in studies that require a level of clinical precision because the system is not considered a medical device for diagnosis yet.

The sampling frequency used to plot the pressure distribution color map was not enough to appreciate the behavior of weight distribution over the seat while pedaling, but it did make it possible to identify the differences in inclination, which were very evident in the 2 volunteers with prostheses. The concept test revealed that there is good acceptance of the system by the participants as an element that they would have at home to exercise and that it is safe to use. Some volunteers reported difficulty sitting down and that it is not very intuitive at first, but they easily assimilated the driving experience using the tutorial. All participants reported feeling the physical exercise and that it can be suitable for all audiences. Regarding the use of virtual reality, all the participants reported enjoying the test, experiencing high immersion, quick response to control the vehicle, and none declared feeling dizzy. Table 1 summarizes the results of the most relevant questions in the survey, where 5 means a lot and 1 very little.
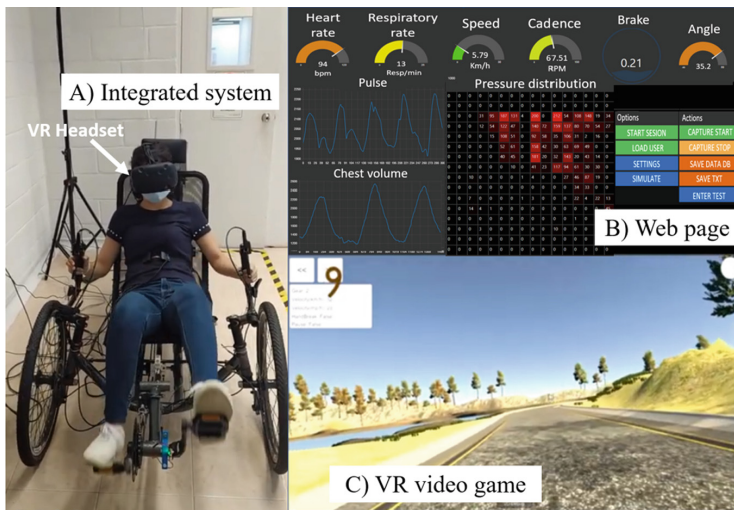


**Fig. 6.** Integrated system with web page and game

**Table 1.** Usability survey results

| Topic | Question | Result |
|---|---|---|
| Usability | Was it easy to use the game together with the trike during the test? | 4.3 |
| | Do you think it was quick to learn | 4.5 |
| | how to use the game with the tricycle? | |
| | Can it be useful for all types of audiences? | 5.0 |
| Immersion | Did you like the virtual environment? | 5.0 |
| | How connected were your senses to the virtual environment? | 4.5 |
| | How natural did the mechanism that controls the movement of | 3.7 |
| | the virtual vehicle feel? | |
| | How confused, dizzy, or disoriented did you feel at the end of the | 1.7 |
| | experimental session? | |
| | How quickly did you adapt to the virtual environment experience? | 4.1 |

## 4    Conclusion

The design and adjustment of the physiological sensors were considered one of the most delicate stages of the project. Therefore, to replicate a prototype like this, it is necessary to adapt these sensing systems to their optimal point before integration. It can be highlighted that the designed system allowed to capture, visualize and store different physiological and mechanical variables, as well as personal user information, adequately and simultaneously, while the participants used the virtual reality game. The fact that no participants reported dizziness could be attributed to the vibration of the trike providing enough haptic feedback to experience a real sensation of driving, however, further studies are needed. In addition, great enthusiasm was generated by the volunteers thanks to the added value of the video game as suggested by the usability survey. However, the participation of a larger number of volunteers is necessary to determine if adjustments must be done so that the system can be adapted to different types of lower limb amputation. A study on the correlation between video game scores and mobility tests is suggested as a next step to determine if the system can also be used as a mobility diagnostic tool for people recovering from trauma with or without prostheses. It is also suggested to perform a longitudinal study with patients to evaluate if the system might fulfill its function as an element for physical training while monitoring physiological responses.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# References

1. Karimé, G., Jorge, F., Luis, M.J.: Physical inactivity, sedentary behavior and chronic diseases. Kor. J. Fam. Med. **38**, 111 (2017)
2. Nicholas, L.: What are the some of the long-term physical effects of using or not using a prosthesis. In: Motion (2010)
3. Israel, D., Gabrielov, R., Itzhak, S., Gabriel, Z., Michael, H.: Evaluation of prosthetic usage in upper limb amputees disability and rehabilitation. **26**, 60–63 (2004)
4. Piazza, L., Ferreira, E.G., Minsky, R.C., Pires, G.K.W., Silva, R.: Assesment of physical activity in amputees: a systematic review of the literature. Sci. Sports **32**, 191–202 (2017)
5. Chamlian Therezinha Rosane: Use of prostheses in lower limb amputee patients due to peripheral arterial disease. Einstein (São Paulo) **12**, 440–446 (2014)
6. Monica, E., Dan, N.: Understanding prosthetic abandonment In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 63, pp. 1644–1648 (2019)
7. Gil, C.V.: Fundamentos de medicina de rehabilitación. Edit, UCR (2007)
8. Bonnechère, B.: Serious Games in Physical Rehabilitation. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-66122-3
9. Tatla S.K., et al.: Evidence for outcomes of motivational rehabilitation interventions for children and adolescents with cerebral palsy: an American Academy for Cerebral Palsy and De- velopmental Medicine systematic review. Dev. Med. Child Neurol. **55**, 593–601 (2013)
10. Peter, L., Julie, B., Gert, K.: Stroke rehabilitation. The Lancet. **377**, 1693–1702 (2011)
11. Kerr, A., Rafferty, D., Moffat, F., Morlan, G.: Specificity of recumbent cycling as a training modality for the functional movements; sit-to-stand and step-up. Clin. Biomech. **22**, 1104–1111 (2007)
12. Diego, V., Andrés Felipe, C., Jonathan, G., Alher Mauricio, H.: Test and fabrication of piezoresistive sensors for contact pressure measurement Revista Facultad de Ingeniería Universidad de Antioquia **82**, 47–52 (2017)
13. Dimitrios, S., Ancret, S., Brandon, B., Hannah, A.D.K, Albert, R., Tobias, L.: Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta- analysis Front. Human Neurosci. **14**, 96 (2020)
14. Brooke, J.: System usability scale usability. gov: improving the user experience: US. Department of Health & Human Services (2018)
15. Thomas, W.: Social presence in multi-user virtual environments: a review and measurement framework for organizational research. North Carolina State Univ. Retrieved Nov. **16** (2008)

# Photobiomodulation in the Treatment of Bovine Subclinical Mastitis

Livia Helena Moreira[1,2(✉)] , Henrique Cunha Carvalho[2,3] ,
Andre Luiz da Silva Mendes[4] , José Carlos Pereira de Souza[5] ,
Leandro Procópio Alves[1,2] , and Renato Amaro Zângaro[1,2]

[1] Universidade Anhembi Morumbi – UAM, São Paulo, SP 04546-001, Brazil
lh.medicinaveterinaria@gmail.com

[2] Centro de Inovação, Tecnologia e Educação – CITÉ, São José dos Campos, SP 12247-016,
Brazil

[3] Universidade Tecnológica Federal do Paraná – UTFPR, Campo Mourão, PR 87301-006, Brazil

[4] Universidade Federal Rural do Rio de Janeiro – UFRRJ, Seropédica, RJ CEP 23890-000,
Brazil

[5] Ministério da Agricultura, Pecuária e Abastecimento – MAPA, Rio de Janeiro,
RJ CEP 20081-250, Brazil

**Abstract.** Mastitis is among the most common, impacting and challenging disease in the dairy industry. Mainly in view of the various disadvantages of conventional therapies, this study aimed to verify the treatment of subclinical mastitis through photobiomodulation. Lactating cows (n = 10) were used, nine (n = 9) with positive subclinical mastitis and one (n = 1) negative used as a control. The photobiomodulation protocol using 5 J/cm$^2$, and occurred on alternate days, totaling seven applications of the Ga-Al-As diode laser (685 nm) inside one cow teat. The milk samples were analyzed before, during and after the treatment to realize the reductase test, colony count and California Mastitis Test (CMT). No changes were found in the reductase test, keeping the milk acidity stable in all samples during and after treatment. In the bacteria isolation and colony count we identified *Staphylococcus aureus*, coagulase-negative *Staphylococcu*s and *Bacillus cereus*, a better performance was observed in animals treated by coagulase-negative *Staphylococcus*, keeping the growth of *Staphylococcus aureu*s stable and increased in *Bacillus cereus*. Regarding the CMT, it was positive for subclinical mastitis before and after treatment, with a decrease in positivity in 60% of the animals 96 h after the last application. The results suggest the development of future research of photobiomodulation protocols for bovine subclinical mastitis with others optical dosimetries, since there were no changes in the milk composition, decreased in the bacteria count, and without generating residues in milk and dairy product losses.

**Keywords:** Photobiomodulation · dairy · mastitis · bovine

# 1   Introduction

Mastitis is a common worldwide inflammatory disease of the mammary gland that causes major impacts on dairy animals, affecting animal welfare and causing great economic losses to the dairy industry through decreased production performance and increased waste milk disposal [1, 2].

The infection transmission occurs mainly between milking due to hygiene failures, and is difficult to control, since the main infection agents are disseminated in the environment where the animals remain [3]. The widespread use of antimicrobials to combat this cause infection has been a serious public health problem, given the risks of the selection of antimicrobial-resistant bacteria, and the presence of these compounds in the dairy products [4], where they must be discarded by the milk-producer [5]. New regulations on veterinary medicine will substantially influence antimicrobial prescribing and usage throughout into the near future. These regulations have been informed by a very large body of work, including the substantial progress towards reduced antimicrobial usage in food animal production [6].

New treatment proposals for subclinical mastitis using phototherapy or photo-biomodulation technique as a new "clean" technology for animal healthcare, the environment and consequently for human beings, and have been recently proposed in the works of Moreira et al. [7, 8], Galstyan & Dobrindt [9], recently in sheep by Silva et al. [10] and in a letter to the editor by Ribeiro and colleagues [11], where they also discuss the clinical challenges and therapeutic advantages of this promising technique, also too with the association of photodynamic therapy.

The objective of this study was to evaluate the effect of photobiomodulation in bovines with subclinical mastitis for the development of preventive protocols for this infection disease.

# 2   Materials and Methods

## 2.1   Animals

This study was approved by the institutional review board (Animal Care and Use Committee) of the University Camilo Castelo Branco (process #1-00034/2012).

Ten lactating Holstein cows (n = 10) were selected, all raised in the confinement management and feeding based on corn silage, protein concentrate, brewing industry residue (barley) and green forage ad libitum, from a commercial farm in Caçapava/SP, Brazil.

Cows were milked using a mechanical herringbone milking parlor ($6 \times 2$) three times a day (4, 12 and 20 h) and per-formed without the calves presence. Routine technical procedures for milking hygiene consisted of asepsis in the mammary gland with Master Iodo (Sani Química Ltda, Valinhos, Brazil), upon entering the milking parlor, then the teats were dried with disposable paper towels and milked, at the end service, the teats were also disinfected with the same product, and the animals were released to the pad-docks.

## 2.2   Analytical and Clinical Exams

All lactating animals underwent visual veterinary inspection of the mammary gland and indirect analysis of the milk quality using the California Mastitis Test (CMT) [12], once a week, according to the manufacturer's instructions, before the second milking of the day (12 h p.m.) for the detection of clinical and/or subclinical mastitis through 120 days. Cows that showed persistent reaction in the CMT exam with three crosses (+++), for a period of three consecutive months without observation of spontaneous cure of subclinical mastitis, were separated for the performance of the photobiomodulation treatment. Under these conditions, the cows (n = 9) were selected using only one mammary quarter of each animal for the experiment. For control group, we used an animal (n = 1) that was in the same period of lactation and that also received the same treatment protocol although it did not have subclinical mastitis.

To the CMT examination the selected cows receives again the mammary gland asepsis routine, and then the first three jets of milk were used for the CMT exam, and after this, 70% ethyl alcohol was passed on the teat to milk microbial collect analysis. The cow teat was cleaned, dried with a paper towel and then collected 20 mL of milk in a sterile bottle for laboratory test. After this procedure, the routine milking was performed and the treatment protocol was applied to the selected teat.

The milk samples collected were immediately stored in isothermal boxes and sent to the laboratory milk analysis. Microbiological identification, colony count and reductase tests were also performed in the laboratory [13]. Finally, ten milk samples were collected from each treated teat, one sample before treatment, seven samples during treatment and two samples after treatment, 48 and 92 h apart from the last treatment.

## 2.3   Photobiomodulation Protocol

Animals were treated with a low-power Gallium-Aluminium-Arsenide (Ga-Al-As) diode laser (Teralaser, MMOptics Ltda, São Carlos, Brazil) with a wavelength in 685 nm, continuous emission mode, power of 20 mW and an energy density of 5 J/cm$^2$ with an exposure time of four minutes (240 s). The light dose was checked before each experiment.

The laser probe was disinfected and inserted internally up to three centimeters into the cow teat canal as showed in Fig. 1, and at the same time the teat was compressed to reduce its length, thus facilitating a better irradiation of the light from laser beam inside the mammary gland. In this sense, after milking the selected cows were treated, consider the time interval of 48 h (alternate days) between each application, consisting of seven total applications per selected teat animal. A negative or decreased CMT within 92 h after the end of treatment was established as a primary objective.

**Fig. 1.** Laser application in the bovine teat canal.

## 3 Results and Discussion

Of the ten treated teats (n = 10) in six teats were isolated Staphylococcus aureus, in two teats coagulase-negative Staphylococcus, and in other two Bacillus cereus as the infection agent of subclinical mastitis.

The average of the CFUs (log CFU) from the animals before treatment (0), during treatment (applications 1 to 7) and after treatment (48 and 92 h) versus the animal control condition are shown in Fig. 2.

The reductase test was performed on all milk samples, and we detected changes in four samples before starting treatment, and no changes were observed during and after treatment in these animals.

Lage and colleagues [14] developed a photobiomodulation LED-based device for the prevention and treatment of teat hyperkeratosis in dairy cows as a preventive proposal for mastitis. Moreira et al. [7] developed a treatment protocol for subclinical mastitis using LEDs associated with a photosensitizer, obtaining good results as an alternative treatment without leaving drug residues in the animal, without milk discarding. The same author and other colleagues [8] recently proposed a device based in infrared LEDs accoupled in an industrial mechanical milking equipment, aiming to prevent subclinical mastitis, but they are necessary more studies to determine the best dosage.

In subclinical mastitis, although it is not possible to diagnose it visually, there is a marked increase in polymorphonuclear leukocytes, which makes it possible to detect this condition by indirect methods, such as CMT, which is the most widespread technique by rural producers because it is easy-to-perform, low-cost and a fast test, for these reasons was the technique used in our study to compare the condition before, during and after treatment, so the results are presented in Table 1.
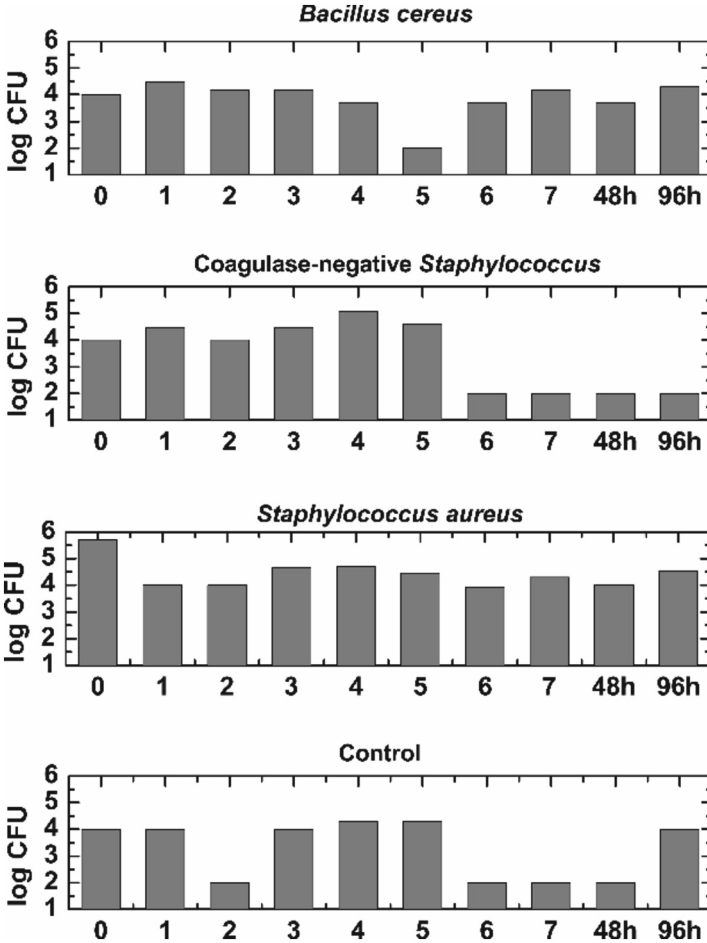
**Fig. 2.** Mean Colony Forming Units (log CFU) of cows photobiomodulation treatment to subclinical mastitis positive microorganisms *vs.* control.

The CMT increases during treatment was expected, corroborated the studies by Albertini et al. [15] and Mansouri et al. [16] that reported that low-level laser therapy induce a reorganization of the inflammatory process, both in cells and in blood vessels; endothelial cells are transformed into capillaries due to the neoforming light effect, as their action on the inflammatory process by modulating neovascularization, which as a consequence reduces the loss of function, increases tissue oxygenation and regional microcirculation, proliferating blood cells such as leukocytes, lymphocytes, polymorphonuclear cells, macrophages and plasma cells, and in this case it was detected in milk by the CMT during and after treatment, due to the photobiomodulation caused by the light absorption. CMT tends to become negative after treatment, as it was observed that 60% of the treated animals obtained a CMT below 3 (+++) at 96 h after treatment.

**Table 1.** CMT before, during and after photobiomodulation treatment.

| Application | Control | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 1st | 0 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 2nd | 1 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 |
| 3rd | 2 | 3 | 2 | 3 | 3 | 2 | 3 | 3 | 2 | 3 |
| 4th | 3 | 3 | 2 | 3 | 2 | 3 | 3 | 2 | 2 | 3 |
| 5th | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 |
| 6th | 1 | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 3 | 2 |
| 7th | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 3 | 3 |
| 48 h | 1 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 2 |
| 96 h | 1 | 2 | 2 | 3 | 2 | 2 | 3 | 3 | 3 | 2 |

Application: 0 = before application; 1–7 = during application; 48 h = 48 h after application; 96 h = 96 h after application; Control = animal control; A–I (n = 9) = treated animals; 0 = animal without subclinical mastitis; 1 = (+) animal weakly positive for subclinical mastitis; 2 = (++) positive animal for subclinical mastitis; 3 = (+++) animal strongly positive for subclinical mastitis.

Moreira et al. [7] used photodynamic therapy in subclinical mastitis confirmed a bactericidal effect with the reduction of the microorganisms Streptococcus dysgalactiae and coagulase-negative Staphylococcus using the laser irradiation at three different spots inside the mammary gland, using a dosage much higher than in this proposed work. The in vitro study by Schultz et al. [17] indicated that Pseudomonas aeruginosa, Staphylococcus aureus, and Escherichia coli were eliminated by laser irradiation when the energy density was greater than only 1.6 J/cm$^2$.

The average CFU of milk samples from treated animals showed the profile of growth and partial inhibition of isolated microorganisms (Staphylococcus aureus, coagulase-negative Staphylococcus and Bacillus cereus) at $10^{-2}$ dilution at energy density of 5 J/cm$^2$, with time of application of four minutes and applying the laser probe (optical fiber) inside the channel of the teat, different from the works of Schultz et al. [17] and Okamoto & Iwave & Morioka [18] that were performed with energy density and application time focusing directly on the microorganism in in vitro assays when obtained a more direct response (partial or total inhibition). Albertini et al. [15] studied the light effect in rat paw edema focusing on the animal's skin, and obtained satisfactory results for the anti-inflammatory effect, as also suggested by Moreira et al. [7] when using photodynamic therapy with 200 J/cm$^2$ to treat subclinical mastitis.

Sharun et al. [2] in a recent review discuss a need for the development of new technologies for the effectiveness in mastitis treatment, since the disease presents different therapeutic responses in antibiotic use, due to microbial resistance to conventional drugs available on the market and the high number of etiological agents cause-disease. New "clean" technologies, without waste generation, are essential for the future of dairy industry in a sustainable way.

## 4 Conclusions

This study suggests the development of future research aiming to prevention of bovine subclinical mastitis applying photobiomodulation technique, based in our results related to bacteria reduction in the treated teat, improvement in the quality of milk by the CMT, and without generating residues in milk and dairy product losses.

**Conflict of Interest.** The authors declare that they have no conflict of interest.

## References

1. Abebe, R., Hatiya, H., Abera, M., et al.: Bovine mastitis: prevalence, risk factors and isolation of *Staphylococcus aureus* in dairy herds at Hawassa milk shed, South Ethiopia. BMC Vet. Res. **12**(1), 270 (2016). https://doi.org/10.1186/s12917-016-0905-3
2. Sharun, K., Dhama, K., Tiwari, R., et al.: Advances in therapeutic and managemental approaches of bovine mastitis: a comprehensive review. Vet. Q. **41**(1), 107–136 (2021). https://doi.org/10.1080/01652176.2021.1882713
3. Langoni, H.: Tendências de modernização do setor lácteo: monitoramento da qualidade do leite pela contagem de células somáticas. Continuous Educ. J. CRMV-SP **3**(3), 57–64 (2000)
4. Aizawa, J., Souza-Filho, A.F., Guimarães, A.S., et al.: Retrospective multicenter study reveals absence of MRSA-associated bovine mastitis in Brazil (1994 to 2016). J. Infect. Dev. Ctries. **13**(6), 581–583 (2019). https://doi.org/10.3855/jidc.11406
5. Santos, M.V.: Suplementação mineral e vitamínica e a relação com mastite e qualidade do leite. Revista Balde Branco **456**, 86–92 (2002)
6. More, S.J.: European perspectives on efforts to reduce antimicrobial usage in food animal production. Ir. Vet. J. **73**, 2 (2020). https://doi.org/10.1186/s13620-019-0154-4
7. Moreira, L.H., de Souza, J.C.P., de Lima, C.J., et al.: Use of photodynamic therapy in the treatment of bovine subclinical mastitis. Photodiagnosis Photodyn. Ther. **21**, 246–251 (2018). https://doi.org/10.1016/j.pdpdt.2017.12.009
8. Moreira, L.H., Lima, C.J., Azevedo, L., et al.: Desenvolvimento do dispositivo de fotobioestimulação para prevenção e tratamento de mastite em vacas lactantes. In: XXVII Congresso Brasileiro de Engenharia Biomédica, pp. 1648–1649, Vitória (2020)
9. Galstyana, A., Dobrindt, U.: Determining and unravelling origins of reduced photoinactivation efficacy of bacteria in milk. J. Photochem. Photobiol. B. **197**, 111554 (2019). https://doi.org/10.1016/j.jphotobiol.2019.111554
10. Silva, L.O., Souza, K.L., Beloti, L., et al.: Use of photodynamic therapy and photobiomodulation as alternatives for microbial control on clinical and subclinical mastitis in sheep. Lasers Med. Sci. **37**, 2305–2310 (2022). https://doi.org/10.1007/s10103-022-03506-2
11. Ribeiro, M.S., Gargano, R.G., Sabino, C.P., et al.: Clinical challenges of antimicrobial photodynamic therapy for bovine mastitis. Photodiagnosis Photodyn. Ther. **21**, 327 (2018). https://doi.org/10.1016/j.pdpdt.2018.01.007

12. Schalm, O.W., Noorlander, D.O.: Experiments and observations leading to development of the California mastitis test. J. Am. Vet. Med. Assoc. **130**(5), 199–204 (1957)
13. Pereira, J.G., Montanhini, M.T.M., Barcellos, V.C.: Testes de redutase para a avaliação da qualidade de leite cru refrigerado. Cient Ciênc Biol Saúde **14**(2), 77–80 (2012)
14. Lage, P.G., Araújo, A.R., Teixeira, A.G., et al.: Photobiomodulation device for prevention and treatment of teat hyperkeratosis in dairy cows. Pesq. Vet. Bras. **34**(6), 515–522 (2014). https://doi.org/10.1590/S0100-736X2014000600004
15. Albertini, R., Villaverde, A.B., Aimbire, F., et al.: Anti-inflammatory effects of low-level laser therapy (LLLT) with two different red wavelengths (660 nm and 684 nm) in carrageenan-induced rat paw edema. J. Photochem. Photobiol. B. **89**(1), 50–55 (2007). https://doi.org/10.1016/j.jphotobiol.2007.08.005
16. Mansouri, V., Arjmand, B., Rezaei-Tavirani, M., et al.: Evaluation of efficacy of low-level laser therapy. J. Lasers Med. Sci. **11**(4), 369–380 (2002). https://doi.org/10.34172/jlms.2020.60
17. Schultz, R.J., Harvey, G.P., Fernandez-Beros, M.E., et al.: Bactericidal effects of the neodymium: YAG laser: in vitro study. Lasers Surg. Med. **6**(5), 445–448 (1986). https://doi.org/10.1002/lsm.1900060505
18. Okamoto, H., Iwase, T., Morioka, T.: Dye-mediated bactericidal effect of He-Ne laser irradiation on oral microorganisms. Lasers Surg. Med. **12**(4), 450–458 (1992). https://doi.org/10.1002/lsm.1900120415

# Author Index