Smart Innovation, Systems and Technologies 104

Suresh Chandra Satapathy
Vikrant Bhateja · Swagatam Das *Editors*

Smart Intelligent Computing and Applications

Proceedings of the Second International Conference on SCI 2018, Volume 1

KES International

Springer

Smart Innovation, Systems and Technologies

Volume 104

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

More information about this series at http://www.springer.com/series/8767

Suresh Chandra Satapathy · Vikrant Bhateja
Swagatam Das
Editors

Smart Intelligent Computing and Applications

Proceedings of the Second International Conference on SCI 2018, Volume 1

Springer

Editors
Suresh Chandra Satapathy
School of Computer Engineering
KIIT Deemed to be University
Bhubaneswar, Odisha, India

Vikrant Bhateja
Department of Electronics and Communication Engineering
Shri Ramswaroop Memorial Group of Professional Colleges
Lucknow, Uttar Pradesh, India

Swagatam Das
Electronics and Communication Sciences Unit
Indian Statistical Institute
Kolkata, West Bengal, India

ISSN 2190-3018 ISSN 2190-3026 (electronic)
Smart Innovation, Systems and Technologies
ISBN 978-981-13-1920-4 ISBN 978-981-13-1921-1 (eBook)
https://doi.org/10.1007/978-981-13-1921-1

Library of Congress Control Number: 2018950782

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Organizing Committee

Organizing Chairs

Dr. Suresh Chandra Satapathy, PVPSIT, India
Dr. B. V. SubbaRao, PVPSIT, India

Program Chairs

Dr. M. V. Rama Krishna, PVPSIT, India
Dr. J. Rajendra Prasad, PVPSIT, India

Program Co-chairs

Dr. S Madhavi, PVPSIT, India
Dr. P. V. S. Lakshmi, PVPSIT, India
Dr. A. Sudhir Babu, PVPSIT, India

Conference Convenor

Dr. B. Janakiramaiah, PVPSIT, India

Conference Co-convenor

Dr. P. E. S. N. Krishna Prasad, PVPSIT, India
Mr. Y. Suresh, PVPSIT, India

Publicity Chairs

Ms. A. Ramana Lakshmi, PVPSIT, India
Mr. A. Vanamala Kumar, PVPSIT, India
Ms. D. Kavitha, PVPSIT, India

Special Session Chairs

Dr. B. Srinivasa Rao, PVPSIT, India
Ms. G. Reshma, PVPSIT, India

Workshop/Tutorial Chairs

Mr. B. N. Swamy, PVPSIT, India
Ms. A. Haritha, PVPSIT, India

Web Master

Mr. K. Syama Sundara Rao, PVPSIT, India
Mr. L. Ravi Kumar, PVPSIT, India

Organizing Committee Members

Computer Science and Engineering

Ms. J. Rama Devi
Ms. G. Lalitha Kumari
Ms. V. Siva Parvathi
Ms. B. Lakshmi Ramani
Mr. I. M. V. Krishna
Ms. D. Swapna
Mr. K. Vijay Kumar
Ms. M. Sailaja
Ms. Y. Surekha
Ms. D. Sree Lakshmi
Ms. T. Sri Lakshmi
Mr. N. Venkata Ramana Gupta
Ms. A. Madhuri
Ms. Ch. Ratna Jyothi
Mr. P. Anil Kumar
Mr. S. Phani Praveen
Mr. B. Vishnu Vardhan
Mr. D. Lokesh Sai Kumar
Ms. A. Divya
Mr. A. Yuva Krishna
Mr. Ch. Chandra Mohan
Mr. K. Venkatesh
Mr. V. Rajesh
Mr. M. Ramgopal

Information Technology

Ms. J. Sirisha
Mr. K. Pavan Kumar
Ms. G. Lakshmi
Mr. D. Ratnam
Mr. M. Sundara Babu

Mr. S. Sai Kumar
Ms. K. Swarupa Rani
Mr. P. Ravi Prakash
Mr. T. D. Ravi Kiran
Ms. Y. Padma
Ms. K. Sri Vijaya
Ms. D. Leela Dharani
Mr. G. Venugopal
Ms. M. Sowjanya
Mr. K. Prudviraju
Mr. R. Vijay Kumar Reddy

Preface

The Second International Conference on Smart Computing and Informatics (SCI 2018) was successfully organized by Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh. The objective of this international conference was to provide a platform for academicians, researchers, scientists, professionals, and students to share their knowledge and expertise in the field of smart computing comprising of soft computing, evolutionary algorithms, swarm intelligence, Internet of things, machine learning, etc. and address various issues to increase an awareness of technological innovations and to identify challenges and opportunities to promote the development of multidisciplinary problem-solving techniques and applications. Research submissions in various advanced technology areas were received, and after a rigorous peer review process with the help of technical program committee members, elite quality papers were accepted. The conference featured special sessions on various cutting-edge technologies which were chaired by eminent professors. Many distinguished researchers like Dr. Lakhmi C. Jain, Australia; Dr. Nabil Khelifi, Springer, Germany; Dr. Roman Senkerik, Tomas Bata University in Zlin, Czech Republic; Dr. B. K. Panigrahi, IIT Delhi; and Dr. S. Das, Indian Statistical Institute, Kolkata, attended the conference and delivered the talks.

Our sincere thanks to all special session chairs, reviewers, authors for their excellent support. We would like to thank Siddhartha Academy for all the support to make this event possible. Special thanks to Sri B. Sreeramulu, Convenor, PVPSIT, and Dr. K. Sivaji Babu, Principal, PVPSIT, for their continuous support. We would like to extend our special thanks to very competitive team members from the Department of CSE and IT, PVPSIT, for successfully organizing the event.

Editorial Boards
Dr. Suresh Chandra Satapathy
Dr. Vikrant Bhateja
Dr. Swagatam Das

Bhubaneswar, India
Lucknow, India
Kolkata, India

Details of Special Sessions

1. Special Session on "Emerging Trends in BigData, IoT, and Social Networks (ETBIS)"

Session Chair

Dr. L. D. Dhinesh Babu, VIT University, Vellore, TN, India

Co-session Chair(s)

Dr. S. Sumathy, VIT University, Vellore, TN, India
Dr. J. Kamalakannan, VIT University, Vellore, TN, India

2. Special Session on "Applications of Computing in Interdisciplinary Domains"

Session Chair

Prof. Dr. V. Suma
Dean, Research and Industry Incubation Centre
Professor, Department of Information Science and Engineering
Dayananda Sagar College of Engineering
Bengaluru, 560078, India

3. "Recent Trends in Data Science and Security Analytics"

Session Chair(s)

Dr. Sireesha Rodda, GITAM University, Visakhapatnam, India
Dr. Hyma Janapana, GITAM University, Visakhapatnam, India

4. Special Session on "Trends in Smart Healthcare Technologies (TSHT)"

Session Chair

Dr. K. Govinda, VIT University, Vellore, TN, India

Co-session Chair(s)

Dr. R. Rajkumar, VIT University, Vellore, TN, India
Dr. N. Sureshkumar, VIT University, Vellore, TN, India

Contents

Bangla Character Recognition by Euclidean Distance Between Center of Gravity and Endpoints 1
Chinmoy Kar and Sreeparna Banerjee

A Dynamic Hierarchical Load Balancing Service Architecture for Cloud Data Centre Virtual Machine Migration 9
Kothapuli Venkata Subba Reddy, Jagirdar Srinivas and Ahmed Abdul Moiz Qyser

Cost Optimization Technique and Resource Utilization of Web and Cloud Services 19
M. Swami Das, A. Govardhan, D. Vijaya Lakshmi and P. S. V. Sucheta

An Empirical Study on Community Detection Algorithms 35
K. Chandusha, S. Rao Chintalapudi and M. H. M. Krishna Prasad

Enhanced Red Wolf Optimization Algorithm for Reduction of Real Power Loss 45
K. Lenin

Estimating Runoff Using Feed-Forward Neural Networks in Scarce Rainfall Region 53
Dillip K. Ghose and Sandeep Samantaray

Non-invasive Heart Rate Measurement on Wrist Using IR LED with IoT Sync to Web Server 65
R. Pawan Sai and M. P. Sunil

Analysis of Constrained Navigation in a Bounded Smart City 75
Arihant Sinha and Jimmy Mathew

Improving Social Safety with Automobile Pilot Adroitness Analyzer 85
Bhupesh Singh, Prashanth Balasubramanian, Jimmy Mathew and Binu John

Abnormal User Pattern Detection Using Semi-structured Server Log File Analysis 97
P. V. Sai Charan

Computational Modeling and Parametric Analysis of an Implantable Patch Antenna Using Finite-Difference Time-Domain Algorithm 107
T. Mary Neebha, M. Nesasudha and Evangelin Chrysolite

A Survey on Acronym–Expansion Mining Approaches from Text and Web 121
R. Menaha and VE. Jayanthi

Deformable Facial Fitting Using Active Appearance Model for Emotion Recognition 135
Lakshmi Sarvani Videla, M. R. Narasinga Rao, D. Anand, Hima Deepthi Vankayalapati and Shaik Razia

Grey Wolf Optimization-Based Improved Closed-Loop Speed Control for a BLDC Motor Drive 145
Devendra Potnuru and Ayyarao S. L. V. Tummala

Energy-Aware Workflow Scheduling Algorithm for the Deployment of Scientific Workflows in Cloud 153
S. Balamurugan and S. Saraswathi

A Framework for Evaluating Performance of MADA-AODV Protocol by Considering Multi-dimensional Parameters on MANET 163
Thati Balamuralikrishna and Mohammed Ali Hussain

A Comparative Analysis of Breast Cancer Data Set Using Different Classification Methods 175
M. Navya Sri, J. S. V. S. Hari Priyanka, D. Sailaja and M. Ramakrishna Murthy

Energy Efficient Forwarding Data and Route Selection (EEFDRS) Algorithm for Wireless Body Area Network 183
G. Smilarubavathy and S. Ayyasamy

Skin Melanoma Assessment Using Kapur's Entropy and Level Set—A Study with Bat Algorithm 193
V. Rajinikanth, Suresh Chandra Satapathy, Nilanjan Dey, Steven Lawrence Fernandes and K. Suresh Manic

Estimation of Distance of a Target Speech Source by Involving Monaural Features and Statistical Properties 203
R. Venkatesan and A. Balaji Ganesh

Termite-Motivated Simulation of Cooperative Behavior 211
Dhruv Chamania, Amit Adate and Parveen Sultana

Combined Feature Extraction for Multi-view Gender Recognition 219
A. Annie Micheal and P. Geetha

Resource Allocation Using Modified Banker's Algorithm for Next-Generation Wireless Networks 229
Patil Harshali and Purohit Seema

Synthesis Methods of Baugh-Wooley Multiplier and Non-restoring Divider to Enhance Primitive's Results of QCA Circuits 237
Bandan Kumar Bhoi, Neeraj Kumar Misra, Manoranjan Pradhan and Rashmishree Rout

Elegant Energy Competent Lighting in Green Buildings Based on Energetic Power Control Using IoT Design 247
B. Tarakeswara Rao, B. V. V. S. Prasad and Subba Rao Peram

Energy-Efficient Routing Protocols for WSNs: A Comparative Analysis 259
B. A. Mohan and H. Sarojadevi

Analysis of Cancer Data Set with Statistical and Unsupervised Machine Learning Methods 267
T. Panduranga Vital, K. Dileep Kumar, H. V. Bhagya Sri and M. Murali Krishna

Performance Analysis of Array Multiplier Using Low-Power 10T Full Adder 277
D. S. Shylu Sam and A. Christina Roseline

A Secure and Lightweight Authentication Protocol for Multiple Layers in Wireless Body Area Network 287
N. V. Abiramy and S. V. Sudha

Energy Characterization of Bluetooth in Opportunistic Mobile Crowdsensing Platform 297
Kalyani Sahoo, Ramesh K. Sahoo and Srinivas Sethi

Emphysema Medical Image Classification Using Fuzzy Decision Tree with Fuzzy Particle Swarm Optimization Clustering 305
Swathi Jamjala Narayanan, Rajkumar Soundrapandiyan, Boominathan Perumal and Cyril Joe Baby

Design of LSTM-Based RNN for Prognosis Prediction of High-Risk Diseases from Patient Diagnostic Histories 315
K. Sathyabama, K. Saruladha and M. Hemalatha

Websocket-Evented Real-Time Online Coding Collaborator 325
Varun Bhatia, Satej Joshi and Radhika Chapaneri

Analytical Survey on Parameters for Designing an Efficient 5G Antenna System 335
M. Benisha, R. Thandaiah Prabu and V. Thulasi Bai

A Comprehensive Survey of Machine Learning-Based Network Intrusion Detection 345
Radhika Chapaneri and Seema Shah

Modified Ant Colony Optimization Algorithm for Task Scheduling in Cloud Computing Systems 357
G. Narendrababu Reddy and S. Phani Kumar

Design of 5G mm-Wave Antenna Using Line Feed and Corporate Feed Techniques 367
R. Thandaiah Prabu, M. Benisha, V. Thulasi Bai and R. Ranjeetha

Multiple Output Off-Line Flyback Converter with a Single Switch 381
R. Rashmi and M. D. Uplane

Predictive Analytic as a Service on Tax Evasion Using Feature Engineering Strategies 393
S. Kishore Babu and S. Vasavi

An Efficient Integrated ERP System Using Multilayer Perceptron 403
Ch. Rupa, J. Ranga Rao and P. Raveendra Babu

A Comprehensive Survey of Services Provided by Prevalent Cloud Computing Environments 413
N. Joshi and S. Shah

Conditional Generative Recurrent Adversarial Networks 425
Siddharth Seth and Mukesh A. Zaveri

A Multiphase Pre-copy Strategy for the Virtual Machine Migration in Cloud 437
Ruchi Shukla, Rajeev Kumar Gupta and Ramgopal Kashyap

Fuzzy C-Means-Based JPEG Algorithm for Still Image Compression 447
Vanitha Kakollu, G. Narsimha and P. Chandrasekhar Reddy

Supportive Communication System for the Elderly Disabled People 459
Pushpa Kotipalli, E. R. Praveen Kumar, M. A. S. Mohan Raju and D. Murali Krishna

Real-Time Digital Control of Synchronous Buck Converter for Low-Power Application 469
R. Rashmi and M. D. Uplane

Integrated Parallel K-Nearest Neighbor Algorithm 479
Rashmi Agrawal

Predicting the Performance of Disability Students Using Assistive Tools with the Role of ICT in Mining Approach 487
P. Saraswathi and N. Nagadeepa

A New Approach for Line Loadability Enhancement in Restructured Power System 495
D. Ragaleela and S. Sivanagaraju

A Robust Gene Data Classification Model Using Modified Manhattan Distance-Based Weighted Gene Expression Graph Classifier 505
N. Sevugapandi and C. P. Chandran

Identity-Based Security Scheme in Internet of Vehicles 515
Rakshanda Agarwal, Sai Satya Pranay, K. Rachana and H. Parveen Sultana

IoT-Based Smartbots for Smart City Using MCC and Big Data 525
Praveen Kumar Singh, Rajesh Kumar Verma and P. E. S. N. Krishna Prasad

Generation of Random Fields for Image Segmentation Based on MRF Energy Level Set Method 535
Rambabu Pemula and C. Nagaraju

Recognition of Islanding Data for Multiple Distributed Generation Systems with ROCOF Shore Up Analysis 547
Ch. Rami Reddy, K. Harinadha Reddy and K. Venkata Siva Reddy

PSO Algorithm Support Switching Pulse Sequence ISVM for Six-Phase Matrix Converter-Fed Drives 559
Ch. Amarendra and K. Harinadha Reddy

Spam Mail Detection Using Data Mining: A Comparative Analysis 571
Soumyabrata Saha, Suparna DasGupta and Suman Kumar Das

Selection of Battery Size by Using Power Flow Decision Program for Microgrids 581
Ch. Padmanabha Raju and D. Dhanalakshmi

An Effective Multi-faceted Cost Model for Auto-scaling of Servers in Cloud 591
S. Phani Praveen and K. Thirupathi Rao

Sign Language Translator Using LabVIEW Enabled with Internet of Things 603
M. Pradeep Kumar, M. Thilagaraj, S. Sakthivel, C. Maduraiveeran, M. Pallikonda Rajasekaran and S. Rama

Study of Galaxy Evolution Through GA-SVM 613
M. Raja Sekar and N. Sandhya

Energy Distribution and Coherence-Based Changes in Normal and Epileptic Electroencephalogram 625
Revati Shriram, V. Vijaya Baskar, Betty Martin, M. Sundhararajan and Nivedita Daimiwal

Welch's Power Spectral Density of Cranial PPG Signal Using AVR ATmega 8535 Microcontroller 637
Nivedita Daimiwal, S. Poornapushpakala, Betty Martin, M. Sundararajan and Revati Shriram

Multi-mode Routing Mechanism with Cryptographic Techniques and Reduction of Packet Drop Using 2ACK Scheme MANETs 649
V. Lakshman Narayana and C. R. Bharathi

Automatic Early Stage Glaucoma Detection Using Cascade Correlation Neural Network 659
Thresiamma Devasia, K. Poulose Jacob and Tessamma Thomas

Performance Evaluation of MRI Pancreas Image Classification Using Artificial Neural Network (ANN) 671
B. Aruna Devi and M. Pallikonda Rajasekaran

User Behavioral Characteristics Identification from Mobile Call Logs 683
Sravya Paritala, Rishitha Bobba, Haritha Akkineni, V. S. Lakshmi Papineni and Lakshmi Gogulamudi

Local Contrast Regularized Contrast Limited Adaptive Histogram Equalization Using Tree Seed Algorithm—An Aid for Mammogram Images Enhancement 693
V. Muneeswaran and M. Pallikonda Rajasekaran

Design of Multi-layer Perceptron for the Diagnosis of Diabetes Mellitus Using Keras in Deep Learning 703
N. Muni Kumar and R. Manjula

A QoS-Enhanced Smart Packet Scheduler for Multi-core Processors in Intelligent Routers Using Machine Learning 713
Suman Paul and Malay Kumar Pandit

Achieving Robustness of Mesh Watermarking Techniques Toward Mesh Geometry and Topology-Invariant Operations 721
Sagarika Borah and Bhogeswar Borah

Author Index 729

About the Editors

Dr. Suresh Chandra Satapathy is currently working as Professor, School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, India. He obtained his Ph.D. in computer science and engineering from JNTU, Hyderabad, and M.Tech. in CSE from NIT Rourkela, Odisha, India. He has 27 years of teaching experience. His research interests are data mining, machine intelligence, and swarm intelligence. He has acted as program chair of many international conferences and edited six volumes of proceedings from Springer LNCS and AISC series. He is currently guiding eight scholars for Ph.D. He is also Senior Member of IEEE.

Dr. Vikrant Bhateja is Professor, Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC), Lucknow, and also Head (Academics and Quality Control) in the same college. His areas of research include digital image and video processing, computer vision, medical imaging, machine learning, pattern analysis and recognition, neural networks, soft computing, and bio-inspired computing techniques. He has more than 90 quality publications in various international journals and conference proceedings. He has been on TPC and chaired various sessions from the above domain in international conferences of IEEE and Springer. He has been Track Chair and served in the core-technical/editorial teams for international conferences: FICTA 2014, CSI 2014, and INDIA 2015 under Springer ASIC series; INDIACom-2015 and ICACCI 2015 under IEEE. He is Associate Editor in *International Journal of Convergence Computing* (IJConvC) and also serving in the editorial board of *International Journal of Image Mining* (IJIM) under Inderscience Publishers. At present, he is Guest Editor for two special issues floated in *International Journal of Rough Sets and Data Analysis* (IJRSDA) and *International Journal of System Dynamics Applications* (IJSDA) under IGI Global publications.

Dr. Swagatam Das received B.E. Tel.E., M.E. Tel.E. (Control Engineering), and Ph.D., all from Jadavpur University, India, in 2003, 2005, and 2009, respectively. He is currently serving as Assistant Professor in the Department of Electronics and Communication Sciences at the Indian Statistical Institute, Kolkata, India. His research interests include evolutionary computing, pattern recognition, multiagent systems, and wireless communication. He has published one research monograph, one edited volume, and more than 200 research articles in peer-reviewed journals and international conferences.

Bangla Character Recognition by Euclidean Distance Between Center of Gravity and Endpoints

Chinmoy Kar and Sreeparna Banerjee

Abstract Character recognition is one of the classic examples of pattern recognition and image processing. Many works has been carried out by researchers in this field. Bangla character recognition becomes an interesting field of research in last few decades. However, researches on Bangla character recognition still in progress. This paper contributes an approach toward character recognition by calculating Euclidean distance between center of gravity (CoG) of a character image and its endpoints. The CoG is calculated based on a research work by Onodera (Singapore ICCS/ISITA, 2:548–552, 1992, [1]). Endpoints of a character are generated from skeletonized image followed by removal of isolated pixels and CoG. Parameters like number of endpoints, mean of normalized Euclidian distance and mean CoG are used for recognition. The above-mentioned method is applied on handwritten characters of different sizes and width with satisfied recognition accuracy.

1 Introduction

Bangla character recognition is one of most attractive fields of research among researchers. Bangla is a popular language in India. It is widely used in West Bengal, Tripura, Assam and Andaman and Nicobar Islands. Bangla script used in other Indian languages like Assamese and Manipuri. Bangla is also official language of People's Republic of Bangladesh. The use and popularity makes Bangla character recognition as subject of research.

Significant work started few decades earlier on Bangla optical character recognition by pattern matching algorithms [2]. Since then various methods are applied

C. Kar (✉)
Sikkim Manipal University, Sikkim, India
e-mail: info.chinmoy@gmail.com

S. Banerjee
Maulana Abul Kalam Azad University of Technology, Kolkata, India
e-mail: sreeparnab@hotmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_1

for recognition purpose, like features extraction methods [3], multilayer perceptrons trained by backpropagation algorithm [4] and deep learning of artificial neural network [5, 6]. Research work on Bengali consonant recognition is also found in [7]. Several researches are carried out on online handwriting recognition of a few Indian scripts [8, 9, 10] including online Bangla handwritten texts [11]. Research work is also carried out on various Indian languages [12], for example Oriya [13], Devanagari [14], Tamil [15], and Telugu [16].

The present paper is motivated by STRICR-FB, which is an example of feature vector-based method for character recognition used on Japanese character [17] and a simple approach is proposed to identify a character based on the Euclidean distance (ED) between CoG and endpoints. Since the method attempts to identify characters of different sizes, the distances must be normalized. The mean normalized ED is calculated as the ratio of CoG to endpoints and maximum ED.

2 Method of Recognition

2.1 Preprocessing

The following steps are used for preprocessing. Steps further divided into three steps:

1. Scanning

   Most of the images are collected from various sources. Some handwritten scanned images are used further.

2. Cropping

   Images of handwritten characters are cropped to maintain a standard between size of various handwritten images.

3. Noise removal using Gabor filter.

   Gabor filter used here to remove blur and noise from image. Most of the images in our experiment are collected from standard sources which are already processed.

2.2 Image Conversion

An input image of handwritten character is converted into binary image before any further process. The standard MATLAB functions are used for this purpose.

2.3 CoG Calculation

CoG is calculated based on following methods (Eqs. 1 and 2):

$$Xcog = \sum_{x=1}^{m} \sum_{y=1}^{n} x.Cxy / \sum_{x=1}^{m} \sum_{y=1}^{n} Cxy \cdot \quad (1)$$

$$Ycog = \sum_{x=1}^{m} \sum_{y=1}^{n} y.Cxy / \sum_{x=1}^{m} \sum_{y=1}^{n} Cxy \cdot \quad (2)$$

where C is a character matrix of size x × y. Xcog is a xth coordinate and Ycog is the yth coordinate of center of gravity (CoG) of the character.

2.4 Skeletonized Representation

Skeletonized representation of a character image is generated by following operations:

1. Removes isolated pixels form a character image.
2. Performs morphological closing.
3. Finds skeleton of the closed image.

All the operations are stated above are done by using standard MATLAB functions. Figure 1 describes how an image is converted into skeletonized image.

2.5 Endpoint Detection of a Character Image

Finding of endpoints of a character image is done by the following steps:

The following two steps are used to achieve the endpoints of a character. Here, 4X4 matrix is used to represent an image and standard methods are applied to achieve the endpoints of that image. In the step 1, irregularities in the boundary regions are removed (shown in Fig. 2). Step 2 is dedicated for endpoint detection (Fig. 3). Figure 4 shows an example of test image with five endpoints.

1. Removes spur pixels from the skeleton.
2. Finds endpoints of Skeletonized image.

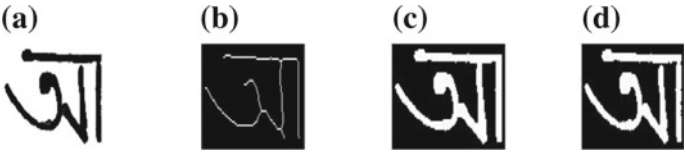


Fig. 1 **a** Original image; **b** skeletonized image; **c** closed image; **d** after removing isolated pixels

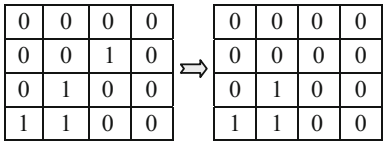


Fig. 2 Removal of spur pixels

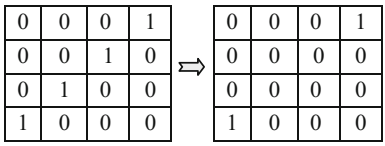


Fig. 3 Endpoints detection

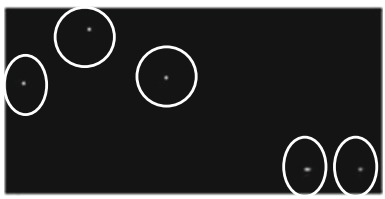


Fig. 4 Example of test image with five endpoints (shown here by using circles)

2.6 *Finding of Euclidean Distance (ED) Between CoG and Endpoints*

The following Eq. 3 is used in the process of calculating ED between every endpoint and CoG. Pxy is an endpoint of C, and then ED (E_{xy}) between Pxy and CoG is given as:

$$\mathrm{ed}_1 = \mathrm{Exy} = \sqrt{(\mathrm{x} - \mathrm{Xcog})^2 + (\mathrm{y} - \mathrm{Ycog})^2}\,. \tag{3}$$

Since many endpoints are recognized by our method, the Euclidean distances must be normalized in order to perform a meaningful comparison. The normalized Euclidean distance is calculated as the ratio of each Euclidean distance over the maximum Euclidean distance (maxE). Then the mean and variance of the normalized Euclidean (NE) distances is calculated by Eq. 4.

$$\mathrm{NE} = \frac{1}{n}\sum_{k=1}^{n}\frac{ed_k}{maxE} \tag{4}$$

2.7 *Mean NE of All Test Images of Same Character*

$$\mathrm{MNE} = \frac{1}{m}\sum_{x=1}^{m}\mathrm{NE_x}\,. \tag{5}$$

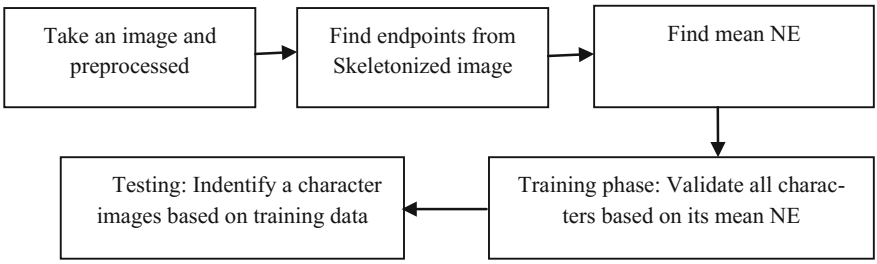


Fig. 5 Block diagram of proposed system

In Eq. 5, m is the number of test image of same character. MNE is mean normalized Euclidean distances. This value will be used for identification purpose.

2.8 Block Diagram of the Proposed System

Block diagram of proposed system is mentioned in Fig. 5.

3 Experimental Result

We have taken a list of Bangla vowel characters for our experiment as shown in Table 1. English alphabets are used here (in suffix) against each Bangla character for better understanding. The set of 20 of each character is taken for our experiment. The whole datasets are divided randomly into two parts. The 75% data are taken for training purpose, and 25% data are preserved for testing purpose.

The MNE of above characters is calculated in MATLAB simulation tool. Table 2 shows the results which we received from training set.

From the above table, we can see the values are different for each character. In case of any similarities between features of two characters, we can distinguish it by seeing other features. No two characters from above are matched exactly with all its features.

Table 1 The set of Bengali vowel characters

অ (A)	ঋ (II)	ৠ (R)	ও (O)
আ (AA)	ঊ (U)	এ (E)	ঔ (AU)
ই (I)	ঊ (UU)	ঐ EI)	

Table 2 Feature values of each character

Characters	Features		
	MNE	Mean endpoints	Mean CoG (x, y)
A	3.304292	5	(32, 48)
AA	3.855232	6	(31, 55)
I	3.86287	5	(58, 41)
II	4.840439	7	(65, 57)
U	2.850235	4	(43, 38)
UU	4.091816	7	(52, 42)
R	4.787708	7	(31, 43)
E	2.241953	3	(32, 46)
AI	3.003382	4	(54, 48)
O	1.838577	3	(38, 42)
AU	2.842854	5	(62, 51)

Table 3 Categories of test results

Case: I (Accurately identified)	Case: II (Identified with conclusion)	Case: III (Unable to identify)
Approx 40% of characters are accurately identified by seeing its feature value	The features values of 30% (approx) test images fall between two characters listed in Table 2. In this case, assumption is made based on difference between feature value of test image and values of Table 2	Approx 30% of test images are unable to identify because of wrong interpretation of weight value

The above table categorizes the result of test images into three cases. The testing is done on 25% of test images.

4 Conclusion

This paper proposed a method for character recognition which is based on mean Euclidian distance, number of endpoints and density. The number of endpoints is calculated here by a simple method which may change with respect to width and size of a character. This paper is motivated by the Dann Barnes and Milos Manic paper and used the modified concept on completely different character set. The proposed approach is never used earlier on Bangla character. The accuracy of the proposed method is definitely less than Soumik Bhattacharya et al., but the approach proposed here is completely different. The major difficulties arises with case II images of Table 3, which can identify accurately in future.

Acknowledgements I would like to acknowledge Dr. Ujjawal Bhattacharya Associate Professor, Computer Vision and Pattern Recognition Unit, ISI Kolkata, for giving free access to handwritten characters dataset.

References

1. Onodera, Y., Watanabe, H., Taguchi, A., Iijima, N., Sone, M., Mitsui, H.: Translation and rotation-invariant pattern recognition method using neural network with back-propagation. Singapore ICCS/ISITA **2**, 548–552 (1992)
2. Chaudhuri, B.B., Pal, U.: A Complete Printed Bangla OCR System. Pattern Recognition, vol. 31, 531–549 (1998)
3. Bhattacharya, S., Sen Mitra, D., Bhattacharya, U., Parui, S.K.: An end to end system for Bangla online handwritten recognition. In: 15th International Conference on Frontiers in Handwriting Recognition (2016)
4. Bhattacharya, U., Sridhar, M., Parui, S.K.: On Recognition of Handwritten Bangla Character. ICVGIP, pp. 817–828 (2006)
5. Iamasa-at, S., Punyaphol, H.: Handwritten Character Recognition Using Histogram of Oriented Gradient Features in deep Learning of Artificial Neural Network (2013)
6. Acharya S., Pant A.K., Gyawali P.K.: Deep learning based Large Scale Handwritten Devanagari Character Recognition (2015)
7. Das, P., Dasgupta, T., Bhattacharya, S.: A handwritten Bengali consonant recognition scheme based on the detection of pattern primitives. In: International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 72–77 (2016)
8. Bharath, A., Madhvanath, S.: Allograph modeling for online handwritten characters in Devanagari using constrained stroke clustering, ACM Trans. Asian Language Info. Proc. **13**(3), 1–12 (2014)
9. Parui, S.K., Guin, K., Bhattacharya, U., Chaudhuri, B.B.: Online handwritten Bangla character recognition using HMM. In: Proceedings of 19th lnternational Conference on Pattern Recognition, pp. 1–4 (2008)
10. Sundaram, S., Ramakrishnan, A.: Two dimensional principal component analysis for online tamil character recognition. In: Proceedings of 11th lnternational Conference on Frontiers in Handwriting Recognition, pp. 88–94 (2008)
11. Bhattacharya, U., Chaudhuri, B.B.: Databases for research on recognition of handwritten characters of Indian scripts. In: Proceedings of the 8th International Conference on Document Analysis and Recognition (ICDAR-2005), Seoul, Korea, vol. II, pp. 789–793 (2005)
12. Pal, U., Chaudhuri, B.B.: Indian Script Character Recognition: A Survey. Pattern Recognition, vol. 37, 1887–1899 (2004)
13. Mohanti, S.: Pattern recognition in alphabets of Oriya Language using Kohonen Neural Network. IJPRAI **12**, 1007–1015 (1998)
14. Ramakrishnan, K.R., Srinivasan, S.H., Bhagavathy, S.: The independent components of characters are 'Strokes'. In: Proceedings of the 5th ICDAR, pp. 414–417 (1999)
15. Suresh, R.M., Ganesan, L.: Recognition of Printed and Handwritten Tamil Characters Using Fuzzy Approach. In: Proceedings of Sixth ICCIMA'05 286–291 (2005)
16. Sukhaswami, M.B., Seetharamulu, P., Pujari, A.K.: Recognition of Telugu characters using neural networks. Int. J. Neural Syst. **6**, 317–357 (1995)
17. Barnes, D., Manic,m M.: STRICR-FB, a novel size-translation- rotation-invariant character recognition method. In: 3rd International Conference on Human System Interaction (2010)

A Dynamic Hierarchical Load Balancing Service Architecture for Cloud Data Centre Virtual Machine Migration

Kothapuli Venkata Subba Reddy, Jagirdar Srinivas and Ahmed Abdul Moiz Qyser

Abstract Growth and focus are towards the high-performance computing, and the paradigm shift for computational research is moving towards the data-centre-centric management of the application. During the last decade, the demand for high performance and timely responsive nature of the applications has increased to the peak and to match up with the demand every service provider has moved applications to the cloud data centre. Nevertheless, cloud data centres struggled to provide the desired response. The parallel research outcomes have demonstrated the uplifting of load balancing mechanism on cloud data centres. However, the architectural dependencies have been the bottlenecks to improve the performance of the load balancing algorithms. The optimal settings of the architectural components in data centres are yet to be proposed to standardize the framework with the proper justification of the hardware and maintainability cost. This paper proposes optimal dynamic and hierarchical hybrid data centre architecture to enhance the load balancing performance from the virtual components of the cloud data centres. It also proposes a performance comparison matrix to evaluate the performance of the existing data centre architectures and the proposed hybrid architecture. The final outcome of this paper is the simulated results from the optimal dynamic and hierarchical hybrid data centre architecture for virtual machine migration policy to be optimal to support the load balancing mechanisms.

K. Venkata Subba Reddy
Department of CSE, Kallam Haranadhareddy Institute of Technology, Guntur, India
e-mail: kvsreddy2012@gmail.com

J. Srinivas (✉) · A. Abdul Moiz Qyser
Department of CSE, Muffakham Jah College of Engineering and Technology, Hyderabad, India
e-mail: jagirdar.srinivas@gmail.com

A. Abdul Moiz Qyser
e-mail: aamoiz@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_2

1 Introduction

The extensive upfront of the computing performance in order to maximize the performance of the applications related all domains like education, research, business and social media is the core demand for high-performance computing. The upfront of the application performance demand can be catered with the suitable configuration of the high-performance hardware and related software protocols. The high-performance hardware resources cause the constraint of cost, which also can be reduced with the proper deployment and management on the cloud data centres. The cost constraints can be controlled with the benefit of cloud service model pay per use strategies. Henceforth, the majority of the service providers have decided to move their application into the cloud, thus the data centres on cloud.

In conjunction with the hardware and load balancing algorithms, the high responsive nature of the application also demands the support from the data centre internal architecture in order to maintain the high-level availability. The parallel research outcomes have demonstrated the bottleneck for further performance improvement by only enhancing the load balancing mechanism [1, 2]. The high availability of the application points to optimal redundancy control of the data centre components. Thus, the parallel research outcomes also have demonstrated the use of optimal setting in many means [3, 4]. The data centres on cloud are the collective framework of all major computing and communication hardware components; thus, the in-depth analysis and understanding of all hardware components are the major goals for the research in order to achieve the optimal policy.

The performance of the cloud data centre for achieving the highest level of load balancing needs to be validated over a fixed format of performance parameters rather than the traditional ways where it is measured based on the availability or high computations speed or higher level of storage or higher level of redundancy. The recent research outcomes have failed to propose a standard performance evaluation matrix, which must propose a set of parameters focusing on all aspects of the data centres for achieving high level of load balancing performance. The demand for the computing growth in terms of the performance of data centres on the cloud, and the formulation of the performance parameters are the demands of the current research trends.

This paper proposes a dynamic hierarchical load balancing service architecture for cloud data centres that is most suitable for load balancing mechanisms and also proposes a generic performance evaluation matrix to evaluate the performance of the cloud data centres.

2 Traditional Data Centre Architecture and Bottlenecks

In order to understand the bottlenecks and scope for the improvements for common data centres architectures [Fig. 1], the traditional data centre architecture and the limitations for performance enhancements are identified.

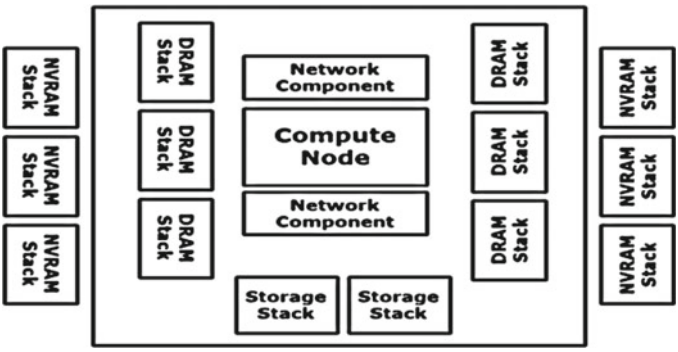


Fig. 1 Traditional data centre architecture

Table 1 Data centre sustainable architecture

Node count	Teraflops	Memory bandwidth (In TBPS)	Network bandwidth (In TBPS)	Power consumption (In Watt)
10	0.001	0.4	0.1	0.02
100	0.1	40	10	2
1000	0.5	400	100	20
10000	5	4000	1000	200
100000	10	4000	1000	200

The major differentiating factor between the computational devices and the traditional data centres is the network of connected components and devices [1, 5]. The demands for network bandwidth in non-cloud data centres are limited as the distribution of the application is also limited geodetically [2, 3, 4].

Also, the demand for throughput management is also the core component of the traditional data centres. In traditional data centres [1], the generic architecture for 1 million sub-nodes should be considered as constant node performance with 10 teraflops, memory bandwidth should match at least 4 TBPS, intra-node network bandwidth should match at least 1 TBPS, and the power consumption is nearly 100 W per min [Table 1].

With the detailed understanding of the traditional non-cloud data centres, the bottlenecks for effective performance enhancement of load balancing mechanisms are listed below:

A. **Power Loss**

Running the load balancing algorithms on traditional data centres demands a high amount of power. Also, the initial power requirements for data centres are very high. The actual power requirements for the load balancing algorithms are half of the total power consumed by the data centres. In future, scope of this paper demonstrates the power-aware data centre framework to address this problem.

B. **Flexibility for Expansion**
The traditional data centres follow the analogy of one application and one server deployment models. Hence, the scope for elasticity is very much limited. Conversely, in the proposed method the virtualization of the components is suitably addressed.

C. **High Carbon Emission**
Recent researches and surveys have demonstrated that the traditional data centres emit nearly 35 million tons of carbon into the environment due to the overuse of standard hardware resources. In the cloud data centre architecture with the help of the optimal distribution of the components and replacement of hardware resources with software modules to a great extent address the problem.

D. **Modular Distribution**
Traditional data centres are not supportive to the modular management of the hardware components. In the proposed framework, the true modular approach for data centre architecture is produced.

E. **Energy Efficiency**
In the connection with the modular distribution limitations [6] of the traditional data centre, it cannot be optimized for the energy efficiency [7, 8]. However, the proposed framework demonstrates the use of smart power grid association for the data centre power supply in the future part of the research.

3 Generic Architecture for Cloud Data Centre Virtualization

The benefits understood in the previous section of the paper define the reason for massive migration of data centres into cloud in order to reach maximized availability and minimized manageability. Henceforth, in this part of the paper, the limitations of the existing data centre architecture are projected in order to propose a new architecture most suitable for load balancing algorithms.

The generic architecture of the cloud traditional data centres [Fig. 2] is the composition of hardware and software service components such as computing, storage, networking [9], cooling devices and software service components such as replication control, power management and security.

The data centres are generally configured for hosting business application for profitability, and decrease in the performance may lead to loss in business. Hence, the data centre performances are to be measured based on the business components. The information technology specific data centre performance measures are majorly classified into four categories as capacity parameter, resource optimization parameters, service parameters and finally the asset management parameter. Here, the proposed matrix for parameterized impacts is presented in Table 2.

In order to achieve a higher rate of flexibility in terms of virtual machine migration and hardware cost optimization for cloud data centres, here formulization

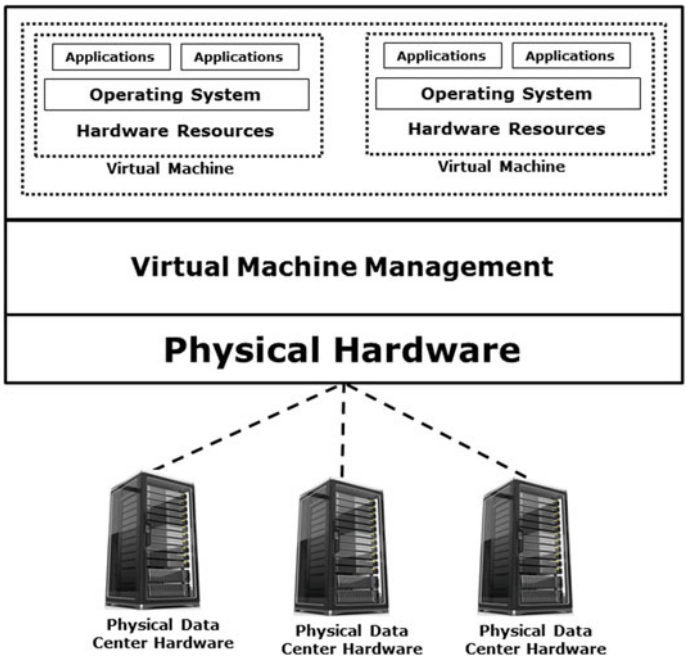


Fig. 2 Generic architecture for cloud data centre

Table 2 Business-oriented application-based data centre performance impact parameters

Data centre impact	Reasons of impact	Impact on business
Capacity parameter	Delay in resource provisioning	Low revenue generation
	Lack of demand forecasting	Low growth in profit
Service parameter	Low system availability	Low customer satisfaction
	Low availability of problem solutions	Low customer trust
Resource optimization parameter	Poor cooling systems	Low return on investment ratio
	Low energy efficiency	Low environment control
Asset management parameter	Low inventory management	Low provisioning
	Poor resource management	Unexpected delay in provisioning

of the resource distribution parameters in the state space "S" considering the total number of resources allocated for the data centre as R is presented, and thus we derive Eq. 1:

$$R_{res} + R_{un} = R, \tag{1}$$

where

R_{run} signifies the quantity of VMs on running state,
R_{res} signifies the quantity of VMs on reserved state,
R signifies the total numbers of VM hardware that can be delivered by the data centre and
Q_v signifies the size of the ready queue for the data centre

The VMs in the queue are noted by Q_V, and thus Eq. 2:

$$RVirtual_{queue} \leq Q_V. \tag{2}$$

$RVirtual_{queue}$ signifies the quantity of VMS on ready queue.
The VMs on the reserved queue are more than the VMs on the ready queue

$$R_{\mathrm{r}\,es} > 0 \Rightarrow RVirtual_{queue} = 0. \tag{3}$$

Hence, we realize that the number of VMs on the ready queue is more than the VMs in the running state. Equation 4 can be written as:

$$RVirtual_{queue} > 0 \Rightarrow R_{run} = R. \tag{4}$$

Equation 5 estimates the peak load of the data centre as:

$$D_{LOAD} = R + Q_v + 1_{\mathrm{E}}. \tag{5}$$

From Eq. 6, the cardinality of the same state space can be derived as:

$$|S| = 2.(D_{LOAD} + 1).(R + Q_V + 1). \tag{6}$$

Equation 7 represents the complexity of any cloud data centre as:

$$O(D_{LOAD}.(R + Q_V)). \tag{7}$$

4 Performance Evaluation Matrix for Data Centre

Along with the proposed new architecture for cloud data centres, this paper also contributes towards the performance evaluation matrix on which the proposed architecture needs to be tested. The work presented in the paper has been carried out using a cloudlet which is cloud data centre on the Internet using a mobility-enhanced small-scale implementation. The cloudlets allow the data centres to provision resources and carry out the resource extensive tasks to be simulated. The parameters to be used for further optimization of the resources in the proposed data centres are listed in Table 3.

Table 3 Proposed performance evaluation matrix

Parameter name	Meaning	Possible values
Cloudlet ID	Sequence number of the request for resources	Integer sequence number
STATUS	Status of the request fulfilment	Success/Failure
DATA CENTRE ID	Simulated data centre ID	Integer sequence number
VM ID	Virtual infrastructure ID	Integer sequence number
TIME	The throughput time	Integer number
START TIME	Start time of the job	Integer number
FINISH TIME	End time of the job	Integer number

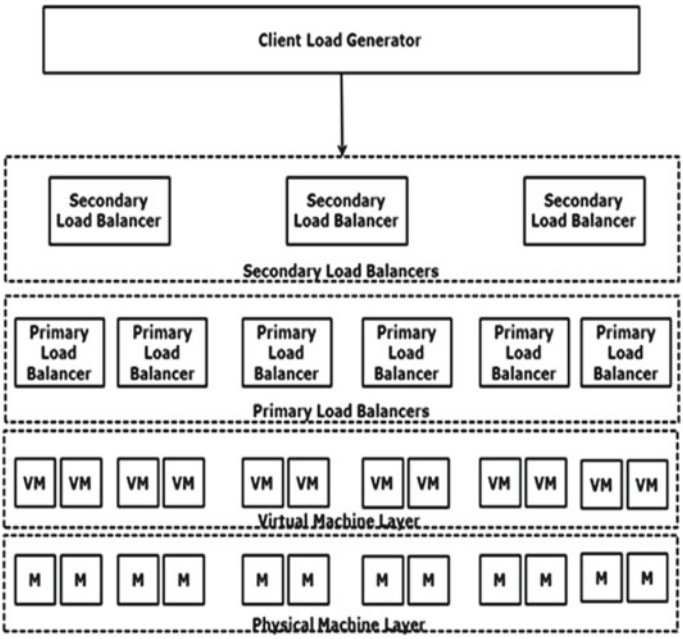


Fig. 3 Proposed cloud data centre architecture to support load balancing strategies

5 Proposed Architecture

The proposed dynamic hierarchical load balancing service architecture for cloud data centre virtual machine migration is the major outcome of this paper. An array of primary load balancers and secondary load balancers is introduced into the traditional data centre architecture.[Fig. 3]. Table 4 represents the simulation results achieved after testing the proposed dynamic hierarchical architecture for load balancing using CloudSim simulator for availability. The simulation was carried out

Table 4 Availability of the proposed data centre architecture

Cloudlet ID	STATUS	Data centre ID	VM ID	Time	Start time	Finish time
0	SUCCESS	3	0	320	0.1	320.1
5	SUCCESS	3	0	320	0.1	320.1
1	SUCCESS	3	1	320	0.1	320.1
6	SUCCESS	3	1	320	0.1	320.1
2	SUCCESS	3	2	320	0.1	320.1
7	SUCCESS	3	2	320	0.1	320.1
4	SUCCESS	3	4	320	0.1	320.1
9	SUCCESS	3	4	320	0.1	320.1
3	SUCCESS	3	3	320	0.1	320.1
8	SUCCESS	3	3	320	0.1	320.1
101	SUCCESS	3	101	320	200.1	520.1
106	SUCCESS	3	101	320	200.1	520.1
103	SUCCESS	3	103	320	200.1	520.1
108	SUCCESS	3	103	320	200.1	520.1
100	SUCCESS	3	100	320	200.1	520.1
105	SUCCESS	3	100	320	200.1	520.1
102	SUCCESS	3	102	320	200.1	520.1
107	SUCCESS	3	102	320	200.1	520.1
104	SUCCESS	3	104	320	200.1	520.1
109	SUCCESS	3	104	320	200.1	520.1

for 50 virtual machines, each running 5 cloudlets. The random load was generated to check the availability and random allocation readiness of the virtual machines. The achieved results are satisfactory. In this part of the paper, enhancement of the load balancing using the virtual machine migration is processed. There are many policies available in the literature for virtual machine migration. The notable policies are maximum correlation (MC), minimum utilization (MU), random selection (RS) and minimum migration time (MMT).The selection of the most efficient strategy continues to be a challenge.

6 Conclusion

This paper demonstrated the traditional data centre architecture and discussed the bottlenecks. Further, the paper demonstrated the cloud computing core feature applicabilities for enhancing the data centre performance. Also, this paper highlights the limitations of the generic cloud data centre architecture and highlighted the limitations for applying the virtual machine migrations. This paper framed the

performance evaluation matrix to evaluate the performance of any cloud data centres in order to validate the availability of physical hosts and virtual machine. The paper listed the load balancing algorithms and introduced the future enhancement

References

1. Zhang, Z., Cheng, X., Su, S., Wang, Y., Shuang, K., Luo, Y.: A unified enhanced particle swarm optimization-based virtual network embedding algorithm. Int. J. Commun. Syst. **26**(8), 1054–1073 (2013)
2. Cheng, X., Su, S., Zhang, Z., Shuang, K., Yang, F., Luo, Y., Wang, J.: Virtual network embedding through topology awareness and optimization. Comput. Netw. **56**(6), 1797–1813 (2012)
3. Sun, G., Anand, V., Yu, H., Liao, D., Cai, Y., Li, L.: Adaptive provisioning for evolving virtual network request in cloud-based data centres. In: Proceedings of IEEE GLOBECOM, pp. 1617–1622 (2012)
4. Beloglazov, A., Abawajy, J., Buyya, R.: Energy-aware resource allocation heuristics for efficient management of data centres for cloud computing. Future Gen. Comput. Syst. **28**(5), 755–768 (2012)
5. Cheng, X., Su, S., Zhang, Z., Wang, H., Yang, F., Luo, Y., Wang, J.: Virtual network embedding through topology-aware node ranking. Comput. Commun. Rev. **41**(2), 38–47 (2011)
6. Botero, J.F., Hesselbach, X., Duelli, M., Schlosser, D., Fischer, A., de Meer, H.: Energy efficient virtual network embedding. IEEE Commun. Lett. **16**(5), 756–759 (2012)
7. A. Berl, E. Gelenbe, M. Girolamo and and G. Giuliani "Energy-efficient cloud computing", Comput. J., vol. 53, no. 7, pp. 1045 –1051 2010
8. Kaplan, J., Forrest, W., Kindler, N.: Revolutionizing data center energy efficiency (2008)
9. An Efficient Cloud Framework for Health Care Monitoring System, Dr. B. Eswara Reddy, Dr. T.V. Suresh Kumar, Gandikota Ramu, 2012 International Symposium on Cloud and Services Computing

Cost Optimization Technique and Resource Utilization of Web and Cloud Services

M. Swami Das, A. Govardhan, D. Vijaya Lakshmi and P. S. V. Sucheta

Abstract Use of web services is increasing rapidly and extensively in banking, e-commerce and other applications. Web-based applications are dynamically associated with high interaction in cloud-based web applications. Cloud computing is an internet computing that provides services such as Platform as Service (PaaS), Software as Service (SaaS), Infrastructure as Service (IaaS) that include computation, memory, networks, etc., by pay as you use service. Web services are rapidly expanded to cloud computing resources which are scalable, efficient resource utilization, improved performance, quality of services and optimized cost. We have discussed the related issues, proposed model, service layer agreement with the operational cost to web and cloud applications. In the future, the paper can be extended to implement various deployment models for minimizing cost and effective utilization of resources.

M. Swami Das (✉)
Department of CSE, MREC, Hyderabad, Telangana, India
e-mail: msdas.520@gmail.com

A. Govardhan
Jawaharlal Nehru Technological University Hyderabad, Kukatpally, Hyderabad, Telangana, India
e-mail: govardhan_cse@yahoo.co.in

D. Vijaya Lakshmi
Department of IT, MGIT, Hyderabad, Telangana, India
e-mail: vijayadoddapaneni@yahoo.com

P. S. V. Sucheta
MGIT, Hyderabad, Telangana, India
e-mail: suchetapallempati6@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_3

1 Introduction

Web services are functional services provided by internet service provider (ISP) over the internet to invoke/access web pages to clients. Web services use programming languages like scripting, HTML, XML, PHP, SOAP, REST and other technologies like J2EE, Servlets and JSPs. Cloud computing is an Internet-based computing that provides computer information processing and data to the computation of other devices on high demand. Cloud uses a shared configurable resource (for example, computer networks, servers, storage, applications and services) to achieve high economic and high availability with affordable prices. In 1960–1970, client–server computing models are used for services. In 1970, telecommunications used virtual private network (VPN) services. VPN services are available at lower cost with features like switching, traffic and load-balancing servers. In 1970, time sharing become popularized by remote job entry using systems.

Grid computing (1997–1999): Grid computing has resource networked communication from multiple locations to invoke services. It is a distributed system with non-interactive workloads, high-performance computing, and a cluster of computing geographically located. In 1989, Tim Berns Lee was the inventor of internet and web applications.

Distributed computing (1970–1980): In distributed systems where computers are located at various locations to communicate and coordinate to invoke/access services. It uses service-oriented architecture (SOA)-based systems. Fog computing provides services such as storage, computation and networking using internet cloud data centers. In 1990, telecommunications offers web services.

Internet (1960–2016): Global system, interconnected with computer networks using TCP/IP protocol, interconnects to billions of devices in the world. In the year 2000, the cloud computing has come into existence. In 2005, world population was 6.5 billion in which 16% of the total population uses the internet. World population in 2010 is 6.9 billion in which 30% of people use the internet. In the year 2014, out of 7.2 billion population, 40% of the population is using the internet. In 2016, world population is 7.4 billion and internet usage is 46.1%. This usage includes applications such as browsing, email, VoIP, file sharing through P2P networks [1, 2].

The following are the benefits of cloud-based services:

i. Flexibility in adding technical infrastructure for the end users.
ii. Minimum cost.
iii. Enable the users to access various devices.
iv. Minimum maintenance.
v. Enable shared resources across the pool of users and efficient utilization of resources.
vi. Improved performance of service provider by using web services and system interfaces.

vii. Increase multiple users to work on same data simultaneously.
viii. Improve scalability, reliability, elasticity and security [1].

The features of cloud computing all things are distributed, scalability, cloud computing resources, architecture, performance tools, characteristics, deployment models, service level agreement (SLA), and other issues [2, 3].

In the development of cloud computing applications, various models can be used by service providers, designers and service organizations. The cost is the one of the most important aspects of design and development of cloud applications; the development of organizations must minimize the development cost and affordable cost to consumers. Different cost-saving approaches were used by physical data centers to meet the user needs and offer service toward as services to users/consumers.

The rest of the paper is organized as follows: Section 2. discusses the related works; Sect. 3 presents the proposed model. Section 4 discusses service layer agreement and operational cost. Section 5 discusses web services; Sect. 6 discusses cloud services and followed by Sect. 7 conclusion.

2 Related Works

2.1 Problem Definition

Organizations need to provide web- and cloud-based services with minimum cost, high interactivity, auto-scaling with the increased workload, fault tolerance, high dynamic traffic, efficient implementation and effective utilization of web services. These web services are connected to cloud services for accessing various applications. Cloud development has main issues that can be related to business, technical, IT resources, security, management, integration, governance, cost, performance and others.

2.2 Literature Survey

M. U Banu et al. [4] proposed that IaaS does have a set of virtual machines (VMs). Each VM instance is based on service provider request. It has several components such as monitoring engines which calculate the demand of workload and usage charge. Work load analyzer measures time intervals request and utilization of resources. Prediction model predicts the demand for workload and finally resource broker (RB) performs planning and delivers subscription to the IaaS provider.

N. Latha et al. [5] proposed a bender decomposition algorithm that uses optimal cloud resource provisioning (OCRP) method providing virtual machines, cloud providers, cost of the service provider, and the decision of virtual machine offered

cloud service providers reserving the cloud service. The Benders algorithm decomposes and optimizes bigger problems into multiple smaller problems.

Rahul Prasad Kanu et al. [6] proposed dynamic cluster configuration algorithm for selecting cloud service. The jobs in the queue scheduling algorithm were based on this policy. The algorithm traverses all the switches linearly to ensure the network latency to be minimum in the locality and is based on cloud provider features; the cost is optimized by optimum cluster dynamic mapreduce jobs.

Today, data centers consume enormous energy, as the performance of cloud computing provides cloud devices that run on various scenarios. Yuan Tian et al. [7] proposed a model consisting of stochastic service decision rules; the performance analysis of the model is MDP (which is a continuous-time Markov chain) with policies and cost functions and metric formulations.

Yu Gu et al. [8] proposed a disaster recovery service model, which aims to provide high data reliability, low backup cost and short recovery time to the cloud customers. The DR model cloud architecture provides QoS satisfaction, constraints, formulation and evaluation performance, backup and recovery. The disaster recovery (DR) model cloud with multiple cloud service providers, data disaster recovery service, and intelligent data strategies is discussed.

The growth of big data demands for processing computation and communication in data centers, factors task assignment, placement of data, movement of data and expenditure operational data centers; Lin Gu et al. [9] proposed a network model, constraints for data placement, data loading, which are discussed; the data processing in 2D Markov chain is expected to completion of time, joint optimization of MINLP problems.

The big data process requires computation, storage and communication to DCs; the cost optimization data are available in three factors, such as task assignment, data placement and data movement. In operational expenditure of data centers, Lin G et al. [10] proposed a method of 2D Markov chain, with mixed integer nonlinear programming with network model, task model and data loading functionalities. In cloud computing multi-objective optimization method, Liyun Zuo et al. [11] proposed a resource cost framework model that uses multi-objective Ant colony optimization to improve the performance, span, cost, dead line violation rate and resource utilization.

Internet and cloud computing provides information, shared resources on demand pay as you manner. To meet the service level agreement (SLA) by cloud providers, W. Zheng et al. [12] proposed an analytical approach based on percentile of performance for unreliable infrastructure as a service clouds. This improves response load, fault frequency, multiplexing and instant processing time. The model has IaaS cloud control flow requests for arrival process, which are in queue map to management of VMs to find error failure detections and message resubmissions.

Cloud computing technology provides the users to pay as you like for web and cloud service models. In these services, the cost is influenced by service invocation time. Lianyong Qi et al. [13] proposed a cost optimization method (CS-COM), by considering multiple impact factors. There are some factors such as job sizes,

execution duration, service invocation time, service quality, and using CloudSim simulated the experiments.

Web services and cloud computing use SOA that requires quality; the QoS management is to identify the role, model, customer perceptions, collection of QoS information, and to aggregate and evaluate the results. Amir Vahid Dastjerdi et al. [14] discussed load balancing, QoS management and cost optimization models; web service selection approaches in selection, optimization and decision-making process.

Cloud service providers need to maintain QoS of cloud providers; John P. Sahlin et al. [15] suggested to improve the performance of map/reduce, virtualization, using TOC optimization methods. Cloud computing uses efficient allocation of resources (disks, servers, networks, etc.), in heterogeneous applications. Abdul Hameed et al. [16] proposed a MapReduce, web applications, resource allocation policies, power management techniques. Cloud computing has grown rapidly, which gives economic benefit to the cloud providers. Danilo Ardagna et al. [17] discussed various system models, applications in user and quality control activities.

3 Proposed Model

Web services are the application services provided by service providers. The Web services also involve higher interactivity with cloud-based application services. The problem with these applications is scalable, fault tolerance, high dynamic traffic and efficient implementation of web services. Architecture has limitations; Amazon European data centers. Amazon EC2 (Elastic Compute Cloud) server. The proposed approach is shown in Fig. 1.

1. Design a web service framework with required QoS parameters
2. The web services also communicate and coordinate the information from cloud-based applications.

The cloud applications such as IaaS, PaaS and SaaS can be utilized in public, private and hybrid deployments. One of the importance of quality features for web and cloud applications are response time, availability, throughput, scalability, caching and other parameters [18].

The end users: The users may be mobile users, devices, or smart devices, who can access the web- and cloud-based applications.

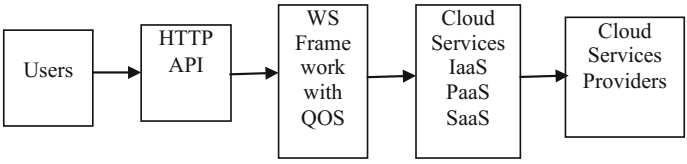


Fig. 1 Web services and cloud services

HTTP API: The end user can communicate and coordinate with the messages using network protocols; for example, HTTP protocols. Web service framework is the design architecture of the modules, interfaces and components.

Cloud services: Cloud computing is a model for enabling on-demand network access to configurable computing resources such as servers, storage, application and services with the minimum management of service provider interactions. Infrastructure, platform and software services, the quality of service, need to be improved, result in cost-saving, improved performance, create on-demand infrastructure and provide high-security applications. Cloud computing services such as public, private and hybrid clouds are offered by multiple service providers. It requires management of allocation, optimization of virtual machines, data centers, networking architecture, and cloud security which are the most essential for various SLAs, for example cloud on Google, IBM cloud, Amazon web services.

Cloud center is a server which is an infrastructure with power devices consisting of UPS and generators. These provide the services to consumers with affordable cost. The designer will consider the network communication among the data centers. The deployments of cloud can be as follows:

1. **Public cloud**: Provides support infrastructure facilities to customers.
2. **Private cloud**: Private cloud is a cloud model that which users can access environment in secured applications, to one organization.
3. **Hybrid cloud**: A company uses both private and public clouds.

Data centers are the computers with services, storage, communication to be provided on demand to the user to deliver cloud services, and the architecture of clouds of various types is shown in Table 1. As Tier1, Tier2, Tier3 and Tier4, the various tiers expected to invoke services as follows:

Data Preprocessing: In general, real-world dataset is incomplete, noisy, and inconsistent. By using preprocessing techniques, the data was converted into standard, consistent and complete format. The web services data set (such as WSDL documents, UDDI services) during preprocessing requires the information to be presented in a standard format. For example, representing the huge data attributes using dimensionality reduction process after the preprocessing stage.

Cloud service providers, for example, Windows Azure, have management and power command tools; the management portal is easy to use, which helps the cloud users to manage applications in web- and cloud-based applications. The application Windows Azure has the following:

Table 1 Availability of cloud services by various architectures

Tier	Percentage of availability of services (%)	Time is taken to access the information
Tier1	99.671	allows 1729.22 min or 28.8 h
Tier2	99.741	allows 1361.3 min or 22.6 h
Tier3	99.982	allows 94.6 min or 1.57 h
Tier4	99.995	status allows 26.2 min or 0.438 h

(1) **Hyper visitors**: Windows Azure is a specialized design to optimize homogeneous data center environment by Microsoft.
(2) **Reliability:** Windows Azure guarantees 99.95% of computing reliability and storage reliability. It provides multiple ways to ensure reliability. Azure has service bus queue to increase the reliability.
(3) **Operating support**: Windows Azure supports both Windows and Linux systems for virtual systems.
(4) **Cost**: Cost is calculated based on computing the usage cost including the sum of the following cost.

   (i) Web site cost
   (ii) Virtual machines cost
   (iii) Cloud service cost
   (iv) Mobile service cost
   (v) Data management and component cost based on configuration.

The consumer may use the cloud services to pay as you approach.

Cloud service resource management: Cloud resources are data and computing centers that give a significant investment in capital outlay and ongoing cost. Optimized cloud resource management is not only to improve resource utilization and performance, but also to reduce the cost of the budget, manpower, for both cloud providers and users. Resource utilization computing resources geographically dispersed data centers and optimized networks, when saving electrical power. Cloud centers are used to reduce the operational cost by providing quality of service (QoS), management of virtual machines [19].

3.1 Cost and Resource Optimization Approach and Algorithm

Figure 2 shows a performance model that has three components: load analyzer, load monitor and predict and cost utilization components. Table 2 describes the various QoS parameters related to cloud and web services also described in Algorithm 1, and Table 3 shows the evaluation parameters of cloud DC.

Algorithm 1: Optimum cost and maximum utilization of web and cloud services.

(1) Read the input
(2) If input classified into domains
(3) The input is processed to Web service
(4) Find load of the web server (load analyzer)
(5) Process load monitor
(6) Predicting the cost of the web and cloud service with required QoS parameters
(7) If cost and effective utilization are satisfied process, the operations go to step 2 until cost and resources are satisfied

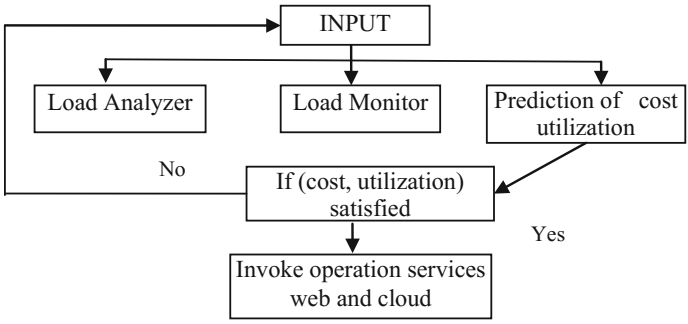


Fig. 2 Cost optimization and resource utilization for web- and cloud-based applications

Table 2 Performance of measure parameters

Component	Performance	Scalability	Storage	Logic
Delay minimized	Optimum concurrency	Auto-scalability	Max. storage	Logical relationship between modules
Maximum request flow	Minimum cost	Management components	Fault tolerance	Dynamic traffic

Table 3 Evolution parameters of cloud data centers

Elements parameters	Measurement	Insight with elements
Topology	{Ring, Star etc.}	Topology depends on network traffic
Data traffic	10 to 100 Gbps	The bandwidth used for data transfer
Control traffic	20 to 40%	Control the traffic monitoring
Data planes latency	Unidirection 4 ms	Data directions uni- or bidirectional
Control planes	40 ms	Control planes used to control the DC network
No. of DC locations	1 to 8 data	The cost affordable data center locations are required
CPU cores per unit	1 Gbps distance in cm For example 20 cm	The CPU configuration based on the data traffic, network, and cloud application with load
No. of cores per server	8 or 16	The servers used to locate and access the web and cloud applications

(8) Web service communicated to optimum cloud server to invoke operations
(9) Optimize the cost and maximum utilization of web and cloud resources.

Network function and cloud center data function are to optimize the data center deployments network load and data center cost.

Total Cost = $\sum$(Website Cost + Virtual Machine Cost + Cloud Service Cost + Cloud Management Cost)

The development cost includes the software cost, cloud service cost for various services such as PaaS, IaaS, and SaaS; the deployment models such as public, private and hybrid cloud services offer VM, CPU allocation, storage and bandwidth, and network topologies. The load analyzer determines the number of requests processed at a unit amount of time and availability of DC with services, network resources.

For example, LoadControl function $F_L(N_{DC,}\ F_d,\ d)$, where F_L denotes load control function, N_{DC} denotes Network data center, F_d represents function group of demands, and demand function to choose the cost of access cloud service including network cost, utilization of data center, and cloud services.

Minimizing the network cost by utilization of DC and data function can be achieved by control servers with optimal workload distribution by using Eqs. 1–5. Demand is a binary variable $d = \{0, 1\}$, minimizes data center cost by server control function, requires data planes, server-requested control planes, and servers control and data planes.

$$Minimum\ Cost = W_{net}X\ C_{net} + W_{dc}X \tag{1}$$

where W_{net} is weighing of the network, W_{dc} is the weight for data center, C_{net} is cost network, and C_{dc} is the cost of the data center.

$$Network\ Load\ Min = \sum\{Data\ Plane + Control\ Plane + Dema1\} \tag{2}$$

$$\begin{aligned} &Minimum\ Data\ Center\ \text{Cost} \\ &\quad = \sum \{No.\ of\ Servers\ Required\ Data\ Plane \\ &\quad + No.\ of\ Servers\ Required\ Control\ Plane \\ &\quad + No\ .of\ Servers\ Required\ for\ Data\ and\ Control\ Plane\} \end{aligned} \tag{3}$$

$$Weight\ W_{net} = \left[\frac{(\lambda_i)}{(Out\ C_{net} - \text{Min}\ C_{net})}\right] \tag{4}$$

where Wnet is the weight of network, Out C_{net} is the output cost of network, $MinC_{net}$ is the minimum cost of network.
Where is the value from 0 to 1, incrementing {0.1, 0.2, … 0.9, 1.0} if value of λ is 1 means optimum value for data centers and networks.

$$Weight\ W_{dc}\left[\frac{(1-\lambda i)}{(Out\ D_C - Min\ D_C)}\right] \tag{5}$$

where W_{dc} is the weight of DC, λ_i is the value of D_C from 0 to 1, $OutD_C$ is the output of D_C, $MinD_C$ is the minimum of D_C.

Evaluation parameters can be network topologies, data traffic, control traffic, data plane latency, control plane latency, number of data centers used, CPU cores used based on the demand of cloud services based on user requests shown in

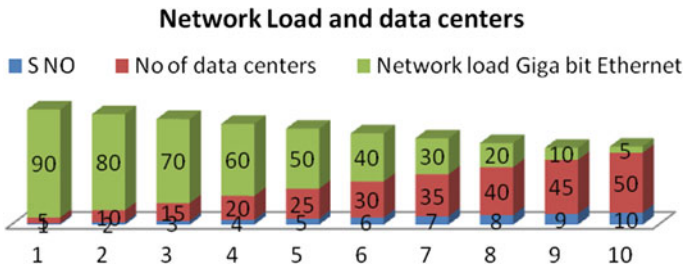


Fig. 3 Network load and data centers, optimize network load by adding DC based on demand: in X-axis (Data center), Y-axis (Network load). The figure shows that the demand to adding data centers at instance with optimized load balance and cost

Table 3 [20]. Case study result features described in the following are shown in Fig. 3, which represents based on network load, add DC increase cost but optimize the model depending on demand of services.

4 Service Layer Agreement (SLA) and Operational Cost

Automatic management of virtual machines will reduce the manpower cost and time. Primary machines were controlled by intelligent automatic resource managers. One needs to manage systems to optimize global utility by considering service layer agreement (SLA) and operational cost.

(i) Data backup and synchronization: Cloud computing service provider should maintain a good privacy for backup and synchronization of information of various terminals on daily basis, and information may include dataset, images, files, emails and other documents. For example, Amazon cloud drive, Sky Drive, Cloud use data increase the scalability with best utilization services.
(ii) Interoperability is the two or more systems which are capable of communicating each other to exchange data, of various domains to access various resources and application services (For example. Educational, business, Medical, E-Government, social and others); these can be interrelated into five ways:

(1) **Product testing**: For product development we need to follow the standard model and testing methodologies.
(2) **Product engineering**: To achieve high interoperability by software implementations.
(3) **Industry/community partnerships**: Common standards are used to allow software systems to communicate each other organizations
(4) **Technology and IP**: IP speed up and reduce complexity and effective use of library resources in the development of communication systems, these use open-source platform technologies.

(5) **Standards Implementation**: Follow the international standards to communicate Web applications. For example, Flight Reservation Application development with standard communication protocols, programming languages, web services, and architectures and others [21].

5 Web Services

Web services are modular functional-based application services that are located, published and invoked across the web, to perform various functions, for example, B2B applications (web-based train booking, hotel booking, flight reservation, hospital management, etc.). Most essential elements for developing web applications are SOAP (XML message, a simple mechanism for exchanging structured typed information between systems), WSDL and UDDI. WSDL is a element to describe the services, request message, port, and binding with SOAP. There are three types of documents that are static, dynamic and active; this will provide web services published in service descriptions (directory services); the advantage of these applications can be used to create a single, required functions in a flexible platform. It also gives options to the users to choose a particular hardware and software with optimum cost. The application will use the content management systems used to store information repositories where the easy to access and reuse. UDDI is the web service directory services that are used to find and locate the service by two ways looking at the well-known location and broadcasting a request. For example, Microsoft, IBM has UDDI Business Registry. Directory has three-part service provider, web services offered and binding [1].

6 Cloud Services

The cloud service models can be classified into three types as IaaS, PaaS, and SaaS, and these have different levels of integration (1) cloud computing and internet of things (IoT) integration, (2) embedded devices components. The device components are categorized into three, namely management, seamless integration, monitoring of cloud platforms and networking environment. The user can receive data from IoT and connected devices, processing and accessing large amounts of heterogeneity data with the use of middleware [22]. Cloud services, are Infrastructure as a Service (IaaS) for large cloud applications, if the service fails due to SLA on the availability of cloud service. Then, the provider will provide high availability of the information in the cloud for large-scale service applications which is necessary to migrate the physical machines among the pools to running service in cloud, turn on the service in cloud, turn of cloud service and interaction with the cloud systems to invoke services Markov chain approach gives reduction in complexity of solution time.

Cloud services use virtual machines. These machines are with some specific characteristics, such as frequency of CPU cores, Memory, and storage. For backup and recovery data centers organized into pools. For example, the requirement of developing web-based cloud application has Datacenter1, Data center2,…. Data center: i. Data center: n. Primary Machines are organized into pools, the States of the systems are hot running (ready), warm turned on (not ready) and cold turned off (default), the delay because of cold turned off state of systems.

To develop a model for IaaS with scalability and availability, scheduling is carried out to complete the task [23]. The work may be virtual computation using threads, process, and data flow and is scheduled for system hardware, processors, networking links and expansion cards. Schedulers are implemented to keep all resources busy (i.e., load balancing) with multiple users to share the systems efficiently to achieve the quality of service. Scheduler goal is to maximize the throughput. The rule-based scheduling algorithm is widely used for many cloud computing systems [24].

6.1 Infrastructure as a Service (IaaS)

Cloud service providers offer services and jobs that run on various virtual machines (VMs); performance depends on the hardware of heterogeneity and virtual instances. Infrastructure as a Service (IaaS) provides on-demand computational services that will predict the performance of the system by heterogeneity within the same virtual machine. This machine performance is measured by considering network, I/O operations, bandwidth, and other parameters. Table 4 shows the cloud service provider system configurations.

Hardware (H_1, H_2, … H_n), Application (a_1 at host1, a_2 at host-2, … a_n at host n) performance h_1 compared with h_2 and $h_3 \ldots h_n$, h_2 compared with h_1, h_{n-1} with h_n and H_n is compared with h_1, $h_2, \ldots h_{n-1}$ Map ReducePerformance of Map Reduce applications.

6.2 Software as Service (SaaS)

Web customers are to share the services with licensed software as a service to authorized users; the server softwares are accessed with minimum cost using VMs through networked systems.

Table 4 Cloud service provider system configurations

Cloud provider	CPU type	Frequency	Cache
Amazon	IntelXeon,72 vCPUs twice	2.2 GHz	25 Mb
IBM	Intel Xeon E5-2660 v2 Cache10-Core Processor	2.2 GHz	25 Mb

6.3 Platform as Service (PaaS)

This is used to develop, manage and run the cloud applications. The end users can access services using VMs. The virtual machines have ranges of small, medium and large, extremely large instances. Cloud services offer tenants at different computing systems regarding disk, CPU, memory. The operating system examples such as Ubuntu server, Linux kernel, public clouds Amazon EC2, Google and others are also provided. Virtual machines are used to access remote desktop applications; these can use machines, and benchmarks Hadoop mainly focus on Virtual Machine Resources, Central Processing Units and I/O bandwidth of communication channel.

6.4 Performance

Performance is one of the essential various quality parameters, which can be measured to analyze the cloud infrastructure and services. It is one of the most essential and important parameters for cloud-based web applications across calculated virtual machines in the cloud. The system has hardware's placed differently based on types of applications, system configurations (CPU, Cache, Memory, and Disk) and other I/O utilities. Comparing Intel processor and AMD processor's performance, Intel is equipped with larger cache than AMD processors. The memory bandwidth, a performance comparison of bench mark application [25].

Input dataset job: I, Size of data block distributed file system: B, Number of mapreduce jobs: n_m, $n_{r,}$ virtual machines mapreduce slots $m_{S,}$ timely execution of mapreduce job: $m_{e.}$ The time for data transfer depends on non-overlapping wave, data processing rate and data processing performance of virtual machine. Availability of CPU and I/O resources is hosted on a primary machine of VM_i. The aggregated I/O and CPU utilize reduced job function. Table 5 gives cloud service applications with some co-located VMs. The execution time of each job is varied upon processor in which load of a system was included (files or no. of jobs executed at a unit time) and located in different virtual machines.

Software and infrastructure are consumed from the cloud. Infrastructure and deployment cost of the smaller organizations, risk, performance, reliability, and security consider these parameters such as cloud cost, clients, workforce.

Today, corporate environment uses various cloud computing technologies that have best practices and applications, software, data center cloud providers, cloud users, IT services in the cloud. Cloud computing model is on demand of access and

Table 5 Cloud service applications with some co-located VMs

Application	Hardware	No. of co-located virtual machines
1	Intel I7 Processor	6
2	Inter I5 Processor	5

a shared pool of configurable computing resources (ex. networks, servers, storage applications) with minimum management and service provider interaction [26].

6.5 Cloud Deployment Models

Cloud deployment model services public, private and hybrid or combination of public and private. To design a cloud service that provides minimum price, Infrastructure as a Service (IaaS) has elastic architecture, designed for failure. High scalable architecture leads to the less robust system. In cloud environment, the cloud resources are shared for load balancing. For example: In Amazon web services, partitioning the clusters into regions and data centers. Performance parameters are available for resources, network, bandwidth, I/O operations, architecture, hardware, latency issues, etc. Cloud resource management and deployment are 1. Automatic scalability and 2. robust architecture.

Cloud vendors provide services with optimum cost and location-based services. These services vendor design-required quality parameters like cost, location, availability. To select and implement cloud services with quality parameters, the organizations must provide data privacy, policies, in cloud data centers. For example (India), SLA service level (more than 99.55%) accountability. The Cloud providers, for example, Amazon, support cost-efficient, integrated monitoring tool which will assess performance; Microsoft Azure has IaaS that offers computing services networking and security constraints; Rackspace is an open-source cloud provider (IaaS); many more companies use data centers, the optional managed and easy to use control panel is Google providing a Google App Engine cloud computing PaaS model to IaaS model; this service is integrated with Google services. For example, HP has IaaS cloud products and services that offer services public, private and hybrid clouds. IBM resources, size, and data centers are best for a large enterprise with heavy data processing, security concerns, and it has cloud provider with good management, software, security and administrators. Finally, factors influence cloud cost (1) complexity of price due to items, providers, user do not have visit selection of lower price, (2) change of cloud provider, cloud cost is instance, discount, and storage and (3) management and optimization such as policies, allocate cost to services, forecast future and analyzing trends [27].

7 Conclusion

Web services are the functional components of SOAP, UDDI and WSDL. Cloud computing is an internet-based computing. Organizations may improve the cloud computing with user flexibility of various services and various deployments. The applications need to have minimum cost and efficient utilization of resources with secured solutions. Most of the organization will use the web- and cloud-based

applications auto-scaling, increase workload, fault tolerance, cost and effective utilization of resources. Cloud service data center models are of Tier1, Tier2, Tier3, and Tier4. The proposed model and algorithm are to improve the performance of cost and effective utilization of resources which is described in Algorithm 1 and Fig. 2. In Fig. 3 which represents based on network load, add DCs to increase cost but optimize the model depending on demand of services. The SLA automatically manages the virtual machines with policies for data backup, cloud storage, interoperability and other communication standards. Performance needs to be considered by the quality of service parameters of web and cloud services response time, availability and others. In the future, the paper can be extended to manage deployment models with automatic scalability, cost and efficient utilization of services.

References

1. Kalil, M., Al-Dweik, A., Abu Sharkh, M.F., Shami, A., Refaey, A.: A framework for joint wireless network virtualization and cloud radio access networks for next generation wireless networks. In: IEEE Access, vol. 5, pp. 20814–20827 (2017)
2. Oujani, A., Jain, R.: A Survey of Cloud Computing Simulations and Cloud Testing, pp. 1–8. https://www.cse.wustl.edu/~jain/cse567-13/ftp/cloud.pdf
3. Shawish, A., Salama, M.: Cloud Computing: paradigms and technologies. In: Inter-cooperative Collective Intelligence: Techniques and Applications, pp. 39–674 (2013)
4. Banu, M.U., Saravanan, K.: Optimizing the cost for resource subscription policy in IaaS cloud. Int. J. Eng. Trends Technol. (IJETT) **6**(5), 296–301 (2013)
5. Latha, N., Deepa, S.T.: Cost optimization in cloud services. Int. J. Comput. Appl. **106**(5), 14–17(2014)
6. Prasad Kanu, R., Shabeera, T.P., Madhu Kumar, S.D.: Dynamic cluster configuration algorithm in mapreduce cloud. Int. J. Comput. Sci. Inf. Technol. **5**(3), 4028–4033 (2014)
7. Tian, Y., Lin, C., CHen, Z., Wan, J., Peng, X.: Performance evaluation and dynamic optimization of speed scaling on web services in cloud computing. In: Tsinghua Science and Technology, IEEE, pp. 298–307 (2013)
8. Gu, Y., Wang, D., Liu, C.: DR-cloud multi-cloud based disaster recovery service. Singhua Sci. Technol. **19**(1), 13–23 (2014)
9. Lin, G., Zeng, D., Li, P., Guo, S.: Cost Minimization for big data processing in Geo distributed data centers. IEEE Trans. Emerg. Top. Comput. **2**(3), 314–323 (2014)
10. Lin, G., Zeng, D., Li, P., Guo, S.: Cost minimization for big data processing in geo-distributed data centers. IEEE Trans. Emerg. Top. Comput. **2**(3), 314–323 (2014)
11. Zuo, L., Dong, S., Zhu, C., Hara, T.: A multi-objective optimization scheduling method based on the ant colony algorithm in cloud computing. IEEE **3**, 2687–2699 (2015)
12. Zheng, W., Zhou, M., Wu, L., Xia, Y., Luo, X., Pang, S., Zhu, Q., Wu, Y.: Percentile performance estimation of unreliable IaaS clouds and their cost-optimal capacity decision. IEEE Access **5**, 2808–2818 (2016)
13. Qi, L., Yu, J., Zhou, Z.: An invocation cost optimization method for web services in cloud environment. Hindawi Sci. Program. **2017**, 1–10 (2017)
14. Vahid Dastjerdi, A., Buyya, R.: A taxonomy of QoS management and service selection methodologies for cloud computing, pp. 109–139 (2011)
15. Sahlin, J.P., Mazzuchi, T.: Optimizing QoSin distributed systems/cloud computing architectures. Int. J. Comput. Appl. **42**(18), 14–20 (2012)

16. Hameed, A., Prakash Jayaraman, P., Zeadally, S., Vishnu, A.: A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems. Comput. pp. 751–774 (2016)
17. Ardagna, D., Casale, G., Ciavotta1, M., Perez, J.F., Wang, W.: Quality-of-service in cloud computing:modeling techniques and their applications. J. Internet Serv. Appl. pp. 1–17 (2014)
18. Fankhauser, T., Wang, Q., Gerlicher, A., Grecos, C., Wang, X.: Web Scaling frameworks for web services in the cloud. IEEE Trans. Serv. Comput. **9**(5), 728–741 (2016)
19. Singh, A., Juneja, D., Malhotra, M.: A novel agent-based autonomous and service composition framework for cost optimization of resource provisioning in cloud computing. http://www.sciencedirect.com
20. Basta, A., Blenk, A., Hoffmann, K.: Towards a cost optimal design for a 5G mobile core network based on SDN and NFV. IEEE Trans. Netw. Serv. Manag. pp. 1–14 (2017)
21. Yang, C., Huang, Q.: Spatial Cloud Computing: A Practical Approach. CRC Press, pp. 191–298(2014)
22. Diaz, M., Martin, C., Rubio, B.: State-of-the-art, challenges, and open issuesin the integration of Internet of things and cloud computing. J. Netw. Comput. Appl. **67**(C), 99–117(2016)
23. Ghosh, R., Longo, F., Frattini, F., Russo, S., Trived, K.S.: Scalable analytics for IaaS cloud availability. IEEE Trans. Cloud Comput. **2**(1), 50–70 (2014)
24. Tsai, C.-W., Huang, W.-C., Chiang, M.-H., Chiang, M.-C., Yang, C.: A hyper-heuristic scheduling algorithm for cloud. IEEE Trans. Cloud Comput. **2**(2), 236–250 (2014)
25. Xu, F., Liu, F., Jin, H.: Heterogeneity and interference-aware virtual machine provisioning for predictable performance in the cloud. IEEE Trans. Comput. **65**(8), 2470–2483 (2016)
26. Serrano, N., Gallardo, G., Hernantes, J.: Infrastructure as a service and cloud technologies. IEEE Softw. **32**(2), 30–36 (2015)
27. https://www.rightscale.com/blog/cloud-cost-analysis/where-10b-waste-public-cloud-costs

An Empirical Study on Community Detection Algorithms

K. Chandusha, S. Rao Chintalapudi and M. H. M. Krishna Prasad

Abstract Social networks are simply networks of social interactions and personal relationships. They have several properties, and community is one among them. These communities can be arranged by individuals in such a way that within the group they can connect more frequently compared to the outside of the group. Community detection can discover groups within a network where individuals' group memberships are not explicitly given. These networks are represented in the form of graph. When graph size is increased then the number of communities will also be increased. Because of this complexity and dynamic nature of the graph, community detection in social network becomes a challenging task. Hence, more research is going on community detection, resulting in plenty of algorithms that come into picture to find effective way of detecting communities in a graph. In this paper, authors have presented different community detection algorithms and also discussed their pros and cons. Finally, authors stated some of the research challenges in this area.

1 Introduction

Social media is an interactive and Internet-based application. Nowadays social media is used everywhere on smart phones, tablets and computers to create highly interactive platform for users. In the physical world, it is very difficult to find similar interests, but it is much easier to find friends in social networks with same kind of interests. As a result, we can connect with the world using social networks. Social

K. Chandusha (✉) · M. H. M. Krishna Prasad
Jawaharlal Nehru Technological University, Kakinada, Andhra Pradesh, India
e-mail: chand.luckystar@gmail.com

M. H. M. Krishna Prasad
e-mail: krishnaprasad.mhm@gmail.com

S. R. Chintalapudi
Pragati Engineering College (Autonomous), Surampalem, Andhra Pradesh, India
e-mail: srao.chintalapudi@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_4

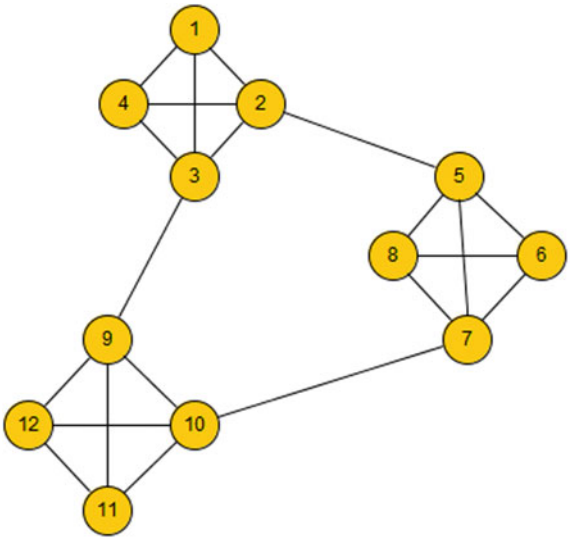


Fig. 1 The community structure. Here three groups of nodes are having strong inner connections and less connection among the groups

networks can be represented as graphs. It contains nodes and edges. Nodes indicate the individuals, and edges indicate the relationship between individuals. The study of graph leads to understand the data and form different communities. Social network plays a major part in different areas like marketing, social relations, anthropology, biology, economics, geography, social psychology and major challenging issues like privacy, perception, immunity, safeguarding, authentic connection, content quantity over quality in the area of social network research. It contains various fields. Social network analysis is one of the fields which is constrained for access problems and providing the solutions to solve this. It can provide the relationship between group of people or objects. This field main aim is visualizing and understanding the relationship within the nodes

Figure 1 shows three communities in which nodes within the community are strongly connected to each other compared to the nodes outside of the community.

Community detection is discovering communities within a network in which individual group members are not definitely given. Detecting community structure in complex network is not an easy task. The practical applications of community detection are to simplify visualization of large graphs, for detecting fraud activities in telecommunication network, for link prediction and for detecting terrorist in online social network.

2 Community Detection Algorithms

Community detection approaches are basically split into four groups. They are node-centric-, group-centric-, network-centric- and hierarchy-centric-based approaches. Every category has its own properties.

2.1 Node Centric Community Detection Algorithms

In this algorithm, every node satisfies the properties of mutuality, reachability of members, nodal degree and relative frequency. These are commonly used in traditional social network analysis. The limitations involved in these community detection algorithms are too strict, not scalable. These are usually used in network analysis with networks of small size. Louvain method and Ramadan F. Mohamed et al. are examples of this category.

Louvain method [1, 2] It is a heuristic algorithm for performing community detection based on modularity optimization [3]. Implementation of this is done in Pajek software. Here modularity is a quality function used for graph division in communities. The modularity function can be calculated by using these formulae.

$$Q = \frac{1}{2m} \sum_{\substack{i,j=0 \\ i \neq j}}^{n} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \tag{1}$$

n represents number of nodes in community. m represents edges. C_i, C_j indicates communities to which node x_i and x_j has been assigned. K_i and K_j indicates the degree of node x_i and $x_{j..}$ Here modularity function can be rewritten as

$$Q = \sum_{c=1}^{n_c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right] \tag{2}$$

n_c indicates number of communities. l_c depicts total number of edges joining nodes from community c. d_c represents sum of the degree of the node from c. The modularity function has drawback that dense communities are weakly connected. In that case, dense communities are taken as separate communities. In order to solve that problem, we are introducing a parameter r called as resolution parameter. By using this r, we can find new Modularity function [4] as in Eq. (3)

$$Q_r = \sum_{c=1}^{n_c} \left[\frac{l_c}{m} - r \left(\frac{d_c}{2m} \right)^2 \right] \tag{3}$$

If the resolution parameter r value > 1, then large number of small communities will be desired. If r value < 1, then small number of large communities will be produced. If r value $= 1$, then it gives new modularity function.

Ramadan F. Mohamed et al. [5, 6] is a combination of Girvan–Newman and DBScan algorithm. DBScan is referred as density-based clustering algorithm. In this algorithm, the clusters with random shapes in spatial database along with the noise will be discovered. In this approach, DBScan algorithm is used to remove the noise. It has two input parameters. Eps and minpts. The density completely depends on the radius of the clusters (eps) and minpts [7]. Then, Girvan–Newman clustering

algorithm is applied [8] to form the communities. In Girvan–Newman, the betweenness will be calculated for all the edges in a network. The edge which contains the highest betweenness will be removed. Then, betweenness is recalculated for all edges affected by the removal. This process is repeated until no edges remain.

2.2 Group-Centric Community Detection Algorithms

In this algorithm, entire group satisfies the properties like group density with some conditions such as group density >= given threshold. It supports recursive pruning. LFK and Tibely algorithms are examples of this category.

Lancichinetti Fortunato kertesz (LFK) [9] algorithm is based on multi-resolution method. It is used to set the community size. This LFK algorithm's main goal is to examine the natural communities of each node of a network. Here, community is a subgraph. It is used to identify maximization of fitness of its nodes.

$$f_G = \frac{K_{in}^G}{\left(K_{in}^G + K_{out}^G\right)^{\alpha}} \tag{4}$$

Here G is community. K_{in} represents the double value of internal connections in the community of G and K_{out} depicts the number of connecting edges of G with the remaining of the graph. α is a positive parameter. It controls the community size as in Eq. (6)

$$f_G^A = f_{G+\{A\}} - f_{G-\{A\}} \tag{5}$$

Here G + {A}(G-{A}) represents the subgraph obtained from module G with the node A inside (outside). Then add the new node or delete the node from subgraph. The fitness of node A is defined as variations according to the fitness in the subgraph G with and without node A. Natural community of node A can be defined iteratively. Initially G = A. Then each node fitness function [10] connected with G is computed. The node with greatest fitness is joined to G, forming a grater community. The node with negative fitness is deleted from community. This algorithm stops when either all nodes have negative fitness or no nodes are left. This procedure is repeated for every node, which has not yet been assigned to natural community.

Tibely algorithm [11] is proposed by Tibely. It is an improvement of LFK approach. In this, fitness function consists of two components. They are separation fitness term and cohesion fitness term. If both are combined, it reflects the Euclidian distance from (0, 0). Separation fitness term ranges from 0 to 1. Cohesion fitness needs to be transformed in order to match first term.

$$f_G = \sqrt{\left(\left(f_G^s\right)^2 + \left(f_G^c\right)^2\right)} \tag{6}$$

Separation fitness term f_G^S is local fitness function, which has been introduced in LFK algorithm with α parameter set to 1. Cohesion fitness term uses second lowest eigen value of Laplacian matrix of community λ_2. Graph Laplacian matrix can be denoted as L = A − D. Here, A represents adjacency matrix and D represents diagonal matrix. The second eigen value of L represents the algebraic connectivity of a graph. Cohesion fitness value can be computed using the Eq. (7)

$$f_G^c = \frac{1}{2} + \frac{1}{2}\frac{\log|\lambda_2|}{\log|G|} \tag{7}$$

2.3 Network-Centric Community Detection Algorithms

In this algorithm, the entire network can be divided into some disjoint sets of nodes. In these algorithms, network connections are global and the clustering depends upon the vertex similarity, multi-dimensional scaling, block model similarity, spectral clustering and maximizing the modularity. The goal of these algorithms is partition of entire network into many disjoint sets. These algorithms need the user to indicate the number of communities beforehand. Clauset, Newman and Moore and Radicchi et al. are examples of this category.

Clauset, Newman and Moore [12] suggested the fast greedy [13] modularity optimization algorithm. Here, modularity [14] is one of the properties of network defined by Newman. If many edges are existing within the community, then the division is good. A_{vw} represents the network of neighborhood matrix elements. It is 1, if both vertices v and w are joined else it is zero. It includes storing the neighborhood matrix of the graph in form of integers and then merges the rows and columns until communities are merged. The straightforward implementation involves storing the graph neighborhood matrix like an array of integers and continuously combining the rows and columns pairs as equivalent communities are combined. Subject to sparse graphs, these are main relevant things in the field, and on the other hand, by using this way time and memory are wasted for storing and merging zero values, which is the extensive majority neighborhood matrix.

$$\mathrm{Q} = \sum_i e_{ij} - a_i^2 \tag{8}$$

Here e_{ij} indicates the division of edges that combine vertices in community i to j, a_i indicates the division of ends of edges that are connected to i.

Find the change in Q for pair of communities and select the maximum change among them and continue the process for all the pair of communities. In multi-graphs,

communities are represented by vertices and edges are used to connect one vertex with another vertex. The edges which are internal called self edges ΔQ_{ij} become time-consuming. The adjacency matrix elements $A'_{ij} = 2me_{ij}$; here two communities i, j are replaced with ith and jth rows and columns by their aggregation. Compute change in modularity and find the pair i, j for which change in modularity is maximum.

Radicchi et al. [15] designed an algorithm to speed up the community identification. It is same as Girvan–Newman algorithm, depending on edge removal. But in this authors used another measure rather than modularity [16]. The measure is count of shortest loop of edges in network. So, this algorithm runs faster. Radicchi et al. finds the shortest loops of edges [17] in a network to calculate edge clustering coefficient. It can be computed as in Eq. (2).

$$C_{ij} = \frac{Z_{ij} + 1}{\min(K_{i-1}, K_{j-1})} \tag{9}$$

Here Z_{ij} represents the number of triangles the edge (i, j) belongs to and K_i, K_j represents the node degrees of i and j, respectively. In this algorithm edges with small values of C_{ij} will be removed and C_{ij} can be calculated for all the remaining edges iteratively. The C_{ij} value will be low for edges between communities. This algorithm is efficient for finding the well-known community structure in the network. The time taken for calculating edge clustering coefficient of an edge is the compound of K_i and K_j.

2.4 Hierarchy-Centric Community Detection Algorithms

These algorithms completely depend on network topology to construct a hierarchal design of communities. It can be represented with two approaches. One is divisive hierarchical clustering. In this approach, nodes are partitioned into definite sets and then again every set is again divided into smaller sets. It is highly stable, but practically it is highly expensive. Second one is agglomerative clustering; in this approach, each node is initialized like a community. Then two communities are selected and combined into larger communities. It considers ultimate modularity increase and largest node similarity. It is very sensitive to the nodes transformed order and adopts combining criteria. Girvan and Newman and Tyler et al. are examples of this category.

Girvan and Newman [18, 19] suggested a hierarchal divisive algorithm based on edge betweenness. Edge betweenness is nothing but total count of shortened paths that run through the edge. In this approach, edge betweenness [20] can be calculated iteratively for all the edges. In the next step, highest betweenness edge is removed and edge betweenness for the remaining networks are recalculated. This step is repeated until all the edges are removed or modularity functions is optimized. This algorithm's time complexity is $O(m^2n)$; here m represents number of

edges and n represents number of nodes. Edge betweenness can be calculated by using Eq. (10)

$$B(e) = \sum_{U \in V, V \in U} \frac{\sigma_e(U, V)}{\sigma(U, V)} \tag{10}$$

Here σ(u, v) depicts the count of shortened paths between u and v, and σ_e(u, v) depicts the count of shortened paths between u and v that include e.

Tyler et al. [21] suggested a variation of Girvan–Newman algorithm to increase the speed of edge betweenness computation. In Girvan–Newman algorithm, one has to compute edge betweenness for all the edges in every iteration, whereas Tyler et al. calculates partial edge betweenness score of all edges using a subset of vertices. If arbitrary sample is selected, this will produce a probabilistic measurement of betweenness. It leads to correct betweenness when sample size is large. Since it is using probabilistic approach, result will contain statistical errors. But Tyler et al. shows that efficient result can be obtained when sample size is small and it offers substantial speed improvement compared to original Girvan–Newman algorithm.

3 Comparative Analysis of Algorithms with Advantage and Disadvantages

Table 1.

Table 1 The advantages, disadvantages and time complexity of different community detection algorithms

Algorithm	Methodology	Time complexity	Advantages	Disadvantages
Girvan–Newman	Edge betweenness	O(mn)	Gives good division of network	Tie with edge betweenness score. High computational complexity
Tyler et al.	Partial betweenness by using subset of vertices	$0(n^3 \log n)$	Speed is improved	Entire graph will be taken It takes more time
Radicchi et al.	Edge clustering coefficient	$O(m^4/n^2)$	Speeds up the identification of communities	Sparse graph expansion takes 0(m) time and speed depends on convergence

(continued)

Table 1 (continued)

Algorithm	Methodology	Time complexity	Advantages	Disadvantages
Clauset, Newman and Moore	Modularity optimization	0(N $\log^2$ N)	It is efficient than modularity of Newman and Girvan	It wastes time and memory for storing and merging 0 matrix elements
LFK	Multi-resolution method	0(n^2 log (n)) in worst case	It is efficient in setting the community size	It is iterative method
Tibley et al.	Fitness function	Range 0–1 in separation fitness	Eigen values are computed	Less cohesive group again divided into communities
Louvain	Modularity quality function	0(n log n)	Allows to change resolution parameter	Dense community
Ramadan F. Mohamed et al.	Eliminating noise points	O(m^2)in worst case	Noise resistant. and less clusters in arbitrary shape and size	It takes more time. It is expensive

4 Research Challenges in Community Detection

After reviewing the above community detection algorithms, authors identified some research challenges in this area. They are:

Scalability and Quality

Majority of the community detection algorithms work effectively in compact networks only, but in real scenario social networks are vast in size. Hence, present algorithms are less effective in detecting community and even few algorithms are unable to find communities also. Hence, there is a scope for developing scalable algorithms and for testing the algorithm with different benchmarks to reveal the quality of communities in real-world networks.

Direct Versus Undirect Network:

Majority of present algorithms work well for undirected networks, and few algorithms even work for directed network as well, but these algorithms do not produce efficient community detection. Hence, there is a huge scope to find the algorithm which works well in both direct and undirected networks.

5 Conclusion and Future Scope

In this survey, authors presented different community detection algorithms on undirected graphs. Majority of present algorithms work well on small-scale networks but are not suitable for wide networks. These community detection techniques are based on edge betweenness, edge clustering coefficient, modularity and fitness function. The specialty of this work is that it reveals the literature review of different community detection algorithms and provides a large amount of information under a single paper. After reviewing all the existing algorithms, this survey concludes that scalability and efficiency are the major factors affecting community detection.

References

1. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities In large networks. J. Stat. Mech. Theory Exp. 1742–5468 (2008)
2. Rotta, R., Noack, A.: Multilevel local search algorithms for modularity clustering. J. Exp. Algorithms, vol. 16, Article no 2.3 (2011). DOI=http://doi.acm.org/10.1145/1963190.1970376
3. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**(2), 026113 (2004)
4. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010). https://doi.org/10.1016/j.phyrep.2009.11.002(2010)
5. ElBarawy, Y.M., Mohamedt, R.F., Ghali, N.I.: Improving social network community detection using DBSCAN algorithm. In: Computer Applications and Research (WSCAR), 2014 World Symposium, (2014). IEEE
6. Mehjabin, et al.: Community detection methods in social networs. In: I.J. Education and managementenigineering, 2015,1,8–18 (2015) (http://www.mecs-press.net), https://doi.org/10.5815/ijeme
7. Chakraborty, S., Nagwani, N., Dey, L.: Performance comparison of incremental K-means and incremental DBSCAN algorithms. Int. J. Comput. Appl. **27**, 14–18 (2011)
8. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. U.S.A. **99**, 7821–7826 (2002)
9. Lancichinetti, A., Fortunato, S., Kertesz, J.: Detecting the overlapping and hierarchical community structure in complex networks. In New J. Phys. **11**, 033015 (2009)
10. Gergely, T.: Criterions for locally dense subgraphs. In: arXiv:1103.3397 [physics.soc-ph] (2011)
11. Mohamed Nasr, M. et al.: A proposed algorithm to detect the largest community based on depth level. Int. J. Adv. Netw. Appl. **09**(02), 3362–3375 (2017). ISSN:0975-0290
12. Clauset, A., Newman, M., Moore, C.: Finding community structure in very large networks. Phys. Rev. E **70**, 066111 (2004). [Online]. Available: http://www.citebase.org/cgibin/citations?id=oai:arXiv.org:cond-mat/0408187
13. Clauset, et al.: A comparative analysis of community detection algorithms on artificial networks. Scientific reports|6:30750| (2016) https://doi.org/10.1038/srep30750
14. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**, 026113 (2004)
15. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D.: Defining and identifying communities in networks. Preprint cond-mat/0309488 (2003)

16. Rao Chintalapudi, S., Prasad, M.H.M.K.: A survey on community detection algorithms in large scale real world networks. In: 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, pp. 1323–1327 (2015)
17. Fortunato, S.: Community detection in networks: a user guide (2016) arXiv:1608.00163v2 [physics.soc-ph]
18. Girvan, M., New man, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA **99**, 7821–7826 (2002)
19. Rao Chintalapudi, S., Krishna Prasad, M.H.M.: Community Detection in Large-Scale Social networks: A Survey (2018). https://doi.org/10.4018/978-1-5225-2814-2.ch012
20. Bojewar, S., Naik, A.P.: A survey paper on Techniques used for Community detection in social networks. In: International Conference on Emanations in Modern Technology and Engineering (ICEMTE-2017), vol. 5(3) (2017)
21. Tyler, J.R., Wilkinson, D.M., Huberman, B.A.: Email as spectroscopy: automated discovery of community structure within organizations. In: M. Huysman, E. Wenger, V. Wulf (eds.) Proceedings of the First International Conference on Communities and Technologies, Kluwer, Dordrecht (2003)

Enhanced Red Wolf Optimization Algorithm for Reduction of Real Power Loss

K. Lenin

Abstract This paper projects enhanced red wolf optimization (ERWO) algorithm for solving optimal reactive power problem. Projected ERWO algorithm hybridizes the wolf optimization (WO) algorithm with particle swarm optimization (PSO) algorithm. Each red wolf has a flag vector, in the algorithm, and length is equivalent to the whole sum of numbers which features in the dataset of the wolf optimization (WO). Due to the hybridization of both WO with PSO exploration, the ability of the proposed red wolf optimization algorithm has been enhanced. Efficiency of the projected enhanced red wolf optimization (ERWO) algorithm has been evaluated in standard IEEE 118 bus test system. Results indicate that enhanced red wolf optimization (ERWO) algorithm performs well in solving the problem. Actual power losses are reduced, and control variables are well within the limits.

1 Introduction

Various techniques are employed for solving optimal reactive power optimization problem, and previously many divergent kind of mathematical methods [1–7] have been used. Also numerous types of evolutionary algorithms [8–15] are utilized to solve the problem. For last twenty years, various types of programming and probabilistic-based approach [16–20] have been used to solve the problem. Enhanced red wolf optimization (ERWO) algorithm is projected to solve the problem. Both exploration and exploitation in the search has been improved in order to reach the global optimal solution. In the basic wolf optimization algorithm (WO) [21], exploration spaces are missing the diversity and the high-quality diversity is needed to upgrade the performance of the algorithm to find an optimal solution. Particle swarm optimization (PSO) [22] has good feature of exploration ability, and it has been hybridized with wolf optimization algorithm (WO) to produce an enhanced

K. Lenin (✉)
Department of EEE, Prasad V. Potluri Siddhartha Institute of Technology, Kanuru, Vijayawada 520007, Andhra Pradesh, India
e-mail: gklenin@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_5

version called as enhanced red wolf optimization (ERWO). PSO will aid to form better preliminary population to WO. Efficiency of the proposed enhanced red wolf optimization (ERWO) algorithm is appraised in standard IEEE 118 bus test system. Projected approach minimizes real power loss considerably when compared with other standard reported algorithms and voltage profiles within the limits.

2 Problem Formulation

Minimizing the real power loss in the transmission network is the main objective of the problem, and it is described in Eq. (1) as follows:

$$F = PL = \sum_{k \in Nbr} g_k \left(V_i^2 + V_j^2 - 2V_i V_j cos\theta_{ij} \right) \tag{1}$$

where F—objective function, P_L—power loss, g_k—conductance of branch, V_i and V_j—voltages at buses i, j, Nbr—total number of transmission lines in power systems.

Objective function is given as Eq. (2), to minimize the voltage deviation in PQ buses,

$$F = PL + \omega_v \times VD \tag{2}$$

VD is voltage deviation and mathematically given as in Eq. (3):

$$VD = \sum_{i=1}^{Npq} |V_i - 1| \tag{3}$$

Equality Constraint

The equality constraint is signified by the power balance Eq. (4) as follows,

$$P_G = P_D + P_L \tag{4}$$

Inequality Constraints

The inequality constraints are given in Eqs. (5) and (6) as below,

$$P_{gslack}^{min} \le P_{gslack} \le P_{gslack}^{max} \tag{5}$$

$$Q_{gi}^{min} \le Q_{gi} \le Q_{gi}^{max}, i \in N_g \tag{6}$$

Bus voltage magnitudes upper and lower bounds are given in Eq. (7) as follows,

$$V_i^{min} \le V_i \le V_i^{max}, i \in N \tag{7}$$

Transformers tap ratios upper and lower bounds are given in Eq. (8) as below,

$$T_i^{min} \le T_i \le T_i^{max}, i \in N_T \tag{8}$$

Compensators reactive powers upper and lower bounds are given in Eq. (9) as follows

$$Q_c^{min} \le Q_c \le Q_C^{max}, i \in N_C \tag{9}$$

where N—number of buses, N_T—number of transformers, and Nc—number of shunt reactive compensators.

3 Red Wolf Optimization

Red wolf optimization mimics the communal management and hunt deeds of red wolves in nature. There are three fittest candidate solutions assumed as α, β and γ to lead the population toward promising regions of the exploration space in each iteration of red wolf optimization. φ is named for the rest of red wolves, and it will assist α, β, and γ to encircle, hunt, and attack prey, that is, to find enhanced solutions. In order to scientifically replicate the encompassing behavior of red wolves are given in Eq. (10).

$$\begin{aligned} \vec{G} &= \left| \vec{F} \cdot \overrightarrow{Y_P}(t) - \vec{Y}(t) \right|, \\ \vec{Y}(t+1) &= \overrightarrow{Y_P}(t) - \vec{H} \cdot \vec{G} \end{aligned} \tag{10}$$

where t indicates the current iteration, $\vec{H} = 2\overrightarrow{b} \cdot \overrightarrow{r_1} - \overrightarrow{b}$, $\vec{F} = 2 \cdot \overrightarrow{r_2}$, $\widehat{Y_P}$ the position vector of the prey, $\overrightarrow{Y}$ is the position vector of a red wolf, $\overrightarrow{b}$ is linearly decreased from 2.0 to 0, and $\overrightarrow{r_1}$ and $\overrightarrow{r_2}$ are arbitrary vectors in [0, 1].

In order to mathematically simulate the hunting behavior of red wolves, Eqs. (11)–(13) are proposed,

$$\begin{aligned} \overrightarrow{G_\alpha} &= \left| \overrightarrow{F_1}, \overrightarrow{Y_\alpha} - \vec{Y} \right| \\ \overrightarrow{G_\beta} &= \left| \overrightarrow{F_2}, \overrightarrow{Y_\beta} - \vec{Y} \right| \\ \overrightarrow{G_\gamma} &= \left| \overrightarrow{F_3}, \overrightarrow{Y_\gamma} - \vec{Y} \right| \end{aligned} \tag{11}$$

$$\begin{aligned} \overrightarrow{Y_1} &= \overrightarrow{Y_\alpha} - \overrightarrow{H_1} \cdot \overrightarrow{G_\alpha} \\ \overrightarrow{Y_2} &= \overrightarrow{Y_\beta} - \overrightarrow{H_2} \cdot \overrightarrow{G_\beta} \\ \overrightarrow{Y_3} &= \overrightarrow{Y_\gamma} - \overrightarrow{H_3} \cdot \overrightarrow{G_\gamma} \end{aligned} \tag{12}$$

$$\vec{Y}(t+1) = \frac{\overrightarrow{Y_1} + \overrightarrow{Y_2} + \overrightarrow{Y_3}}{3} \tag{13}$$

In this work, a new enhanced red wolf optimization (ERWO) algorithm is proposed for solving reactive power problem and the position of a red wolf was updated by Eqs. (13) and (14) is used to discrete the position.

$$flag_{i,j} = \begin{cases} 1 & Y_{i,j} > 0.50 \\ 0 & otherwise \end{cases} \tag{14}$$

where i indicates the *j*th position of the *i*th red wolf and $flag_{i,j}$ is features of the wolf.

4 Particle Swarm Optimization (PSO)

In particle swarm optimization (PSO) [22], the positions and velocities of the particles are modernized as in Eqs. (15) and (16) as follows:

$$v_{t+1}^{i} = \omega_t \cdot v_t^i + cg_1 \cdot Rm_1 \cdot \left(m_t^i - y_t^i\right) + cg_2 \cdot Rm_2 \cdot \left(m_t^g - y_t^i\right) \tag{15}$$

$$y_{t+1}^{i} = y_t^i + v_{t+1}^i \tag{16}$$

The current position of particle is y_t^i, and search velocity is v_t^i. Global best-found position is m_t^g. In uniformly distributed interval (0, 1), Rm_1 and Rm_2 are arbitrary numbers. Here cg_1 and cg_2 are scaling parameters.ω_t is the particle inertia. The variable ω_t is modernized through Eq. (17) as below,

$$\omega_t = (\omega_{max} - \omega_{min}) \cdot \frac{(t_{max} - t)}{t_{max}} + \omega_{min}. \tag{17}$$

Maximum and minimum of ω_t are represented by ω_{max} and ω_{min}; maximum number of iterations is given by t_{max}. Until termination conditions are met, this process will be repeated.

5 Enhanced Red Wolf Optimization (ERWO) Algorithm for Solving Reactive Power Problem

In the simulation red wolves, α, β and γ determine the position of the prey. $\vec{H} = 2\overrightarrow{b} \cdot \overrightarrow{r_1} - \overrightarrow{b}$ directs the exploration and exploitation process by reducing the value from 2 to 0. When $\left|\vec{H}\right| < 1$, it converged toward the prey, and if $\left|\vec{H}\right| > 1$ it

diverged away. The first best minimum loss and variables are accumulated as "α" position, score and as like second best, third best accumulated as "β" and "γ" position and score.

```
Commence
Initialize the parameters
Initialize b, H and F; beginning positions of Red wolves has been
stimulated.
i = 1: population size
j = 1: n
When (i, j) > 0.500
(i) = 1;
 Else
(j) = 0;
End if
End for
Work out the maximum fitness of Red wolves as follows,
Primary maximum fitness of the Red wolf is designated as "α"
Second maximum fitness of the Red wolf is designated as "β"
Third maximum fitness of the Red wolf is designated as "γ"
While k < maximum iteration
For i = 1: population size
Exact Location of the existing Red wolf has been revised periodically
End for
For i = 1: population size
For i=1:n
If (i, j) > 0.500
(j) = 1;
Else
(j) = 0;
End if
End for
Sporadically revise the values of b, H and F;
At this stage Fitness of Red wolves has been calculated
The assessment of red wolves "α","β" and "γ" has to be revised
k=k+1;
End while
Re-examine the value of" α "as the optimal characteristic division;
End
```

6 Simulated Outcomes

Enhanced red wolf optimization (ERWO) algorithm is tested in standard IEEE 118-bus test system [23].

Total number of generator buses—54, load buses—64, number of transformers —9. Table 1 gives the limitation of the values.

Table 2 gives statistical comparison of real power has been listed, and the performance of proposed enhanced red wolf optimization (ERWO) algorithm has

Table 1 Limitation values

Bus number	5	34	37	44	45	46	48
QCmaximum	0	13	0	9	9	9	14
QCminimum	−39	0	−24	0	0	0	0
BUS number	74	79	82	83	105	107	110
QCmaximum	11	19	19	19	19	5	5
QCminimum	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 2 Statistical comparison of real power loss

Actual power loss (p.u)	Biogeography-based optimization [24]	Improved local search biogeography-based optimization/ strategy1 [24]	Improved local search biogeography-based optimization / strategy1 [24]	Projected ERWO
Minimum	128.770	126.980	124.780	116.440
Maximum	132.640	137.340	132.390	118.935
Moderate	130.210	130.370	129.220	117.021

been improved with real power loss which has been considerably reduced when compared to other standard reported algorithms.

7 Conclusion

Optimal reactive power problem has been successfully solved by enhanced red wolf optimization (ERWO) algorithm. Both the exploration and exploitation in the search have been improved in order to reach the global optimal solution. And the validity of the proposed ERWO algorithm has been tested in standard IEEE 118 bus test system. Performance comparison with other standard algorithms gives superior results. Real power loss has been reduced, and voltage profile index is within the specified limits.

References

1. Alsac, O., Scott, B.: Optimal load flow with steady state security. IEEE Trans. PAS, 745–751 (1973)
2. Lee, K.Y., Paru, Y.M., Oritz, J.L.: A united approach to optimal real and reactive power dispatch. IEEE Trans. Power Apparatus Syst.: PAS-104, 1147–1153 (1985)

3. Monticelli, A., Pereira, M.V.F., Granville, S.: Security constrained optimal power flow with post contingency corrective rescheduling. IEEE Trans. Power Syst:PWRS-2(1), 175–182 (1987)
4. Deeb, N., Shahidehpur, S.M.: Linear reactive power optimization in a large power network using the decomposition approach. IEEE Trans Power Syst. **5**(2), 428–435 (1990)
5. Hobson, E.: Network constrained reactive power control using linear programming. IEEE Trans. Power Syst. PAS-99(4), 868–877 (1980)
6. Lee, K.Y., Park, Y.M., Oritz, J.L.: Fuel-cost optimization for both real and reactive power dispatches. IEE Proc 131C(3), 85–93 (1984)
7. Mangoli, M.K., Lee, K.Y.: Optimal real and reactive power control using linear programming. Electr. Power Syst. Res. **26**, 1–10 (1993)
8. Canizares, C.A., de Souza, A.C.Z., Quintana, V.H.: Comparison of performance indices for detection of proximity to voltage collapse. **11**(3), 1441–1450 (1996)
9. Anburaja, K.: Optimal power flow using refined genetic algorithm. Electr. Power Compon. Syst **30**, 1055–1063 (2002)
10. Devaraj, D., Yeganarayana, B.: Genetic algorithm based optimal power flow for security enhancement. IEE Proc.-Gen. Transm. Distrib. **152** (2005)
11. Berizzi, A., Bovo, C., Merlo, M., Delfanti, M.: A ga approach to compare orpf objective functions including secondary voltage regulation. Electr. Power Syst. Res. **84**(1), 187–194 (2012)
12. Yang, C.-F., Lai, G.G., Lee, C.-H., Su, C.-T., Chang, G.W.: Optimal setting of reactive compensation devices with an improved voltage stability index for voltage stability enhancement. Int. J. Electr. Power Energy Syst. **37**(1), 50–57 (2012)
13. Roy, P., Ghoshal, S., Thakur, S.: Optimal var control for improvements in voltage profiles and for real power loss minimization using biogeography based optimization. Int. J. Electr. Power Energy Syst. **43**(1), 830–838 (2012)
14. Venkatesh, B., Sadasivam, G., Khan, M.: A new optimal reactive power scheduling method for loss minimization and voltage stability margin maximization using successive multi-objective fuzzy lp technique. IEEE Trans. Power Syst. **15**(2), 844–851 (2000)
15. Yan, W., Lu, S., Yu, D.: A novel optimal reactive power dispatch method based on an improved hybrid evolutionary programming technique. IEEE Trans. Power Syst. **19**(2), 913–918 (2004)
16. Yan, W., Liu, F., Chung, C., Wong, K.: A hybrid genetic algorithm interior point method for optimal reactive power flow. IEEE Trans. Power Syst. **21**(3), 1163–1169 (2006)
17. Yu, J., Yan, W., Li, W., Chung, C., Wong, K.: An unfixed piecewise optimal reactive power-flow model and its algorithm for ac-dc systems. IEEE Trans. Power Syst. **23**(1), 170–176 (2008)
18. Capitanescu, F.: Assessing reactive power reserves with respect to operating constraints and voltage stability. IEEE Trans. Power Syst. **26**(4), 2224–2234 (2011)
19. Hu, Z., Wang, X., Taylor, G.: Stochastic optimal reactive power dispatch: Formulation and solution method. Int. J. Electr. Power Energy Syst. **32**(6), 615–621 (2010)
20. Kargarian, A., Raoofat, M., Mohammadi, M.: Probabilistic reactive power procurement in hybrid electricity markets with uncertain loads. Electr. Power Syst. Res. **82**(1), 68–80 (2012)
21. Kaveh, A., Shokohi, F.: Application of Grey Wolf Optimizer in design of castellated beams. Asian J. Civil Eng. **17**(5), 683–700 (2016)
22. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proceeding of IEEE International Conference on Neural Networks, Perth, Australia, pp. 1942–1948 (1995)
23. IEEE: The IEEE 30-bus test system and the IEEE 118-test system (1993). http://www.ee.washington.edu/trsearch/pstca/
24. Cao, Jiangtao, Wang, Fuli, Li, Ping: An improved biogeography-based optimization algorithm for optimal reactive power flow. Int. J. Control Autom. **7**(3), 161–176 (2014)

Estimating Runoff Using Feed-Forward Neural Networks in Scarce Rainfall Region

Dillip K. Ghose and Sandeep Samantaray

Abstract This work depicts the use of two neural network techniques: (i) feed-forward back propagation network (FFBPN) and (ii) radial basis function network (RBFN), to forecast runoff as a function of rainfall, temperature and loss due to evapotranspiration. For model architecture, the criteria for evaluation are convergence of mean square error and coefficient of determination. For Tan-sig function in FFBPN, 4-2-1, 4-3-1, 4-5-1, and 4-9-1 architectures are taken into consideration for computation of performance. For Tan-sig function, the best model architecture is found to be 4-2-1 which possess MSE training value 0.000808, MSE testing value 0.004119, RMSE training value 0.028392, RMSE testing value 0.064165 and coefficient of determination for training 0.9914 and testing 0.9255. FFBPN performs the best among four networks with model architecture 4-5-1 using Log-sig transfer function. Similarly for the case of LRN, with Tan-sig function the best model architecture is found to be 4-5-1 which possess MSE training value 0.000483, MSE testing value 0.001025, RMSE training value 0.02316, RMSE testing value 0.03085 and coefficient of determination in training and testing as 0.9925, 0.9611, respectively. Overall results found that LRN performs best as compared to FFBPN for predicting runoff in the watershed. This result will help for planning, design and management of hydraulic structures in the vicinity of the watershed.

1 Introduction

The reliability of water availability has a key function in scheduling water resources project. Runoff calculation is made ensuing precipitation and its abstracts along with weather parameters on the watershed are considered as the initial step. In general, measurement of runoff is either short or long. In such cases, the second step may be the development of precipitation–evapotranspiration–runoff co-relations.

D. K. Ghose (✉) · S. Samantaray
Department of Civil Engineering, National Institute of Technology, Silchar, Silchar, Assam, India
e-mail: dillipghose2002@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_6

These co-relations can be used to predict the runoff from the observed precipitation, temperature and evapotranspiration. Precipitation is the release of water from the atmosphere to reach the surface of the earth. It is the major contribution made by water to a catchment area and needs careful assessment in hydrological study. For this type of framework, estimating and monitoring the rainfall–evapotranspiration–runoff in watershed are needed for eastern India. Runoff from land surface is the flow of water that comes from excess water from rain, over the Earth's surface. It is a major component in regional and global hydrological cycles. It is crucial to understand complex relationships between rainfall and runoff processes and then to estimate surface runoff for efficient design, planning and management of watershed. This can be achieved using hydrological modeling. Physical, statistical, computational and combined methods are adopted to develop model. Physically model occurs directly. Smith and Eli [4] have reported that the runoff response of a watershed focused to rainfall input with variation of weather climatic characteristics. Chen and Adams [1] have found conceptual rainfall–runoff models which are widely employed in hydrological modeling. Zhao [7] has developed the Xinanjiang (XAJ) model for rainfall–runoff process to use across China. Xiong et al. [5] have developed rainfall–runoff model to perform the multilayer perceptron neural network (MLPNN). The development of hybrid neural networks is used with a considerable attention [3]. For rainfall–runoff model, the radial basis function neural networks have similar modeling results to that of the multilayer perceptron neural network (Dawson and Willby, Zakermoshfegh [2, 6].

The objective of this study is to estimate runoff in watersheds. Precipitation and its abstracts are used to develop models, and analysis is needed for describing the results. It has been observed from the literature review that most works have computed, using rainfall–runoff relationship. Present work motivates to study the following objectives: study on correlation of rainfall–runoff and predication of runoff using precipitation maximum temperature, minimum temperature and evapotranspiration.

2 Study Area

Saintala watershed of Balangir district, Odisha, India, is taken into consideration for the proposed study area. The study is made for predicating runoff in three watersheds to assess the drainage capacity of watershed during monsoon period ranging from 1993 to 2012. The watersheds are located in the upper part of Hirakud reservoir. The latitude 21° 88′ 36″ N and longitude 84° 90′ 75″ E are the geocoordinates of the watersheds shown in Fig. 1.

The average monthly precipitation, maximum monthly average temperature and minimum monthly average temperature and evapotranspiration data for monsoon month (May to October) from the period 1993 to 2016 spanning over 24 years are collected from IMD, Bhubaneswar. The monthly runoff and data are collected from soil conservation office, Balangir.

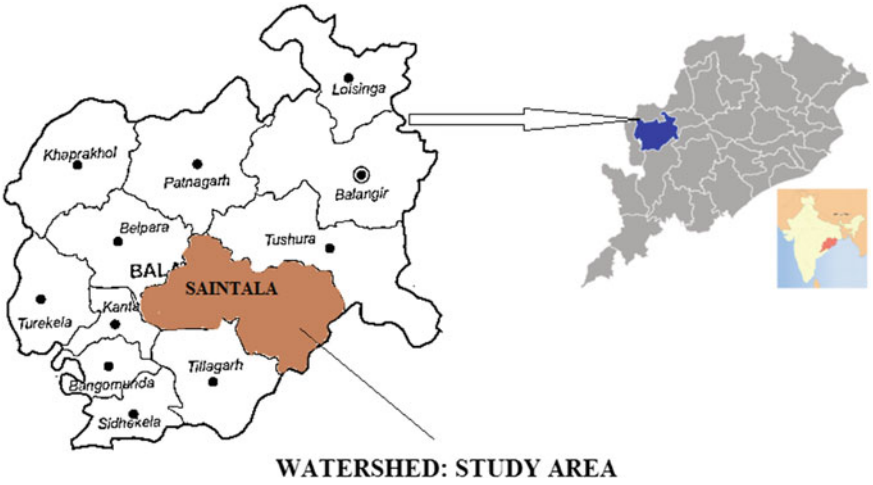


Fig. 1 Study Area: Saintala watershed

3 Methodology

3.1 Artificial Neural Network

ANN consists of three context units: input to output via hidden layer. Nodes of input unit represent the number of input features responsible for output, and number of output nodes represents the output features to be predicted. The network receives the data from input layer transferred the information to entire units for processing. Middle unit encompasses the processing and forwarding the information to the output for executing the result.

3.2 Feed-Forward Back Propagation Network

FFBPN is an organized technique for functioning the process of ANN. Descent-based delta-learning algorithm is used through efficient method for varying weights in FFBPN. The gradient is then used in a simple stochastic manner for adjusting weights to compromise the error. Local minimization of gradient is needed for quick convergence of finding errors. The architecture of FFBPN is represented in Fig. 2.

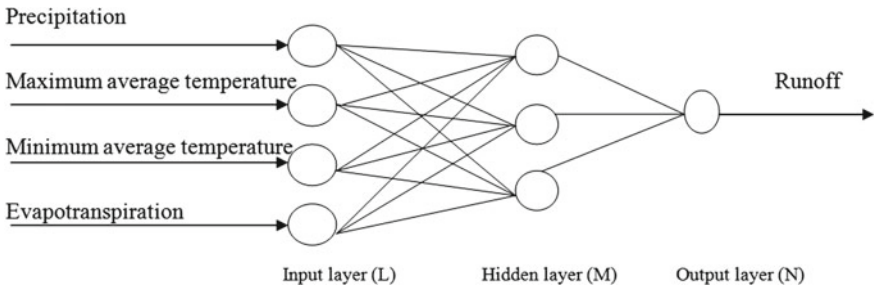


Fig. 2 Architecture of FFBPN model

3.3 Radial Basis Function Network

Radial basis function network is used for approximating function and recognizing patterns. Gaussian potential is used in networks for regulating networks. The interpolation function is a linear combination of basis functions.

$$S(x) = \sum_{i=1}^{n} w_i v_i(x) \tag{1}$$

As basis functions b_i, radial basis function of the form is

$$v_{i(x)} = \varnothing(\|x - x_i\|) \tag{2}$$

where $\varnothing$ is mapping, and $\|x - x_i\|$ is Euclidean distance, $v_{i(x)}$ is output (Figs. 3, 4 and 5).

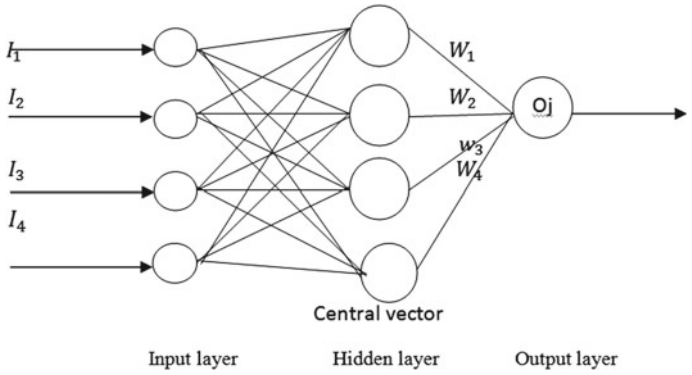


Fig. 3 Architecture of radial basis function network

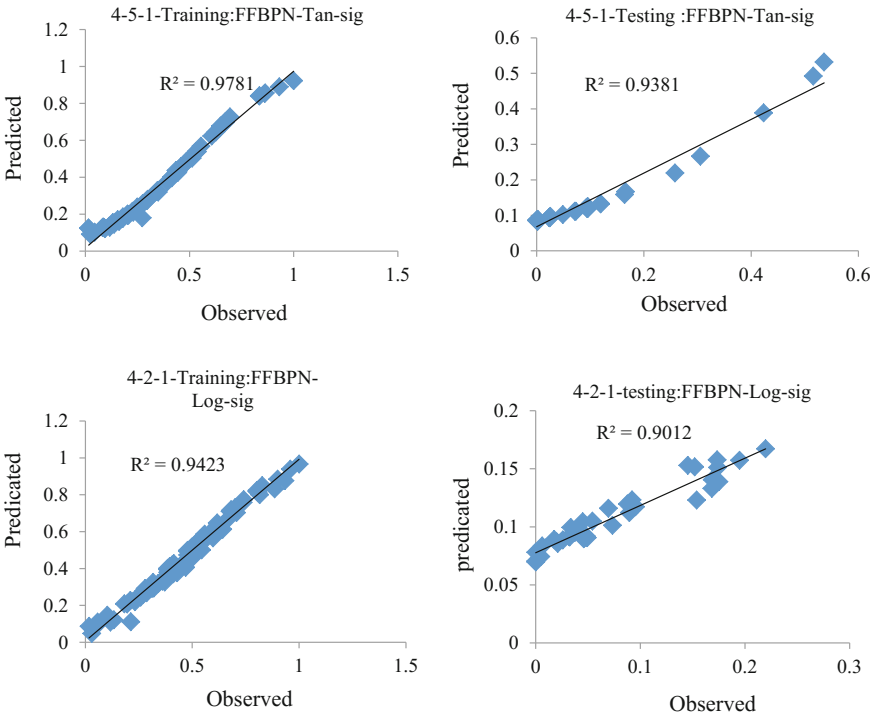


Fig. 4 Observed versus predicted runoff using FFBPN

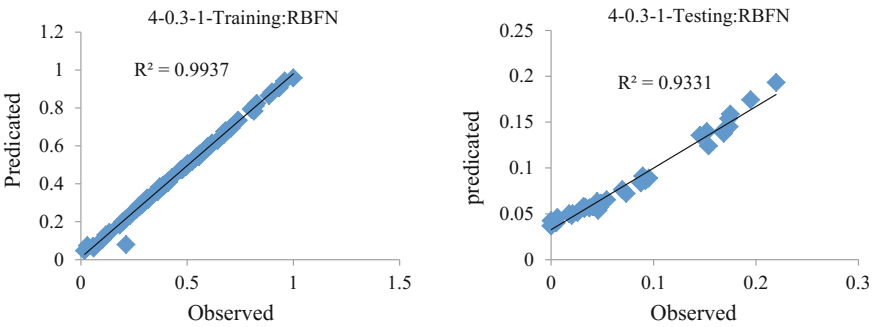


Fig. 5 Observed versus predicted runoff using RBFN

Usage of activation function for computing the output is maintained through the training algorithm of RBFN. Gaussian activation function is used for computing RBFN. The Gaussian function is defined as

$$f(x) = e^{-x^2} \tag{3}$$

And its derivative which is given by

$$f'(x) = -2xf(x) \tag{4}$$

The output of i_m unit $v_i(x_i)$ in the hidden layer.

$$v_i(x_i) = e\left(- \sum_{j=1}^{r} \left[x_{ji} - \widehat{x_{ji}}\right]^2 / \sigma_1^2 \right) \tag{5}$$

where

x_{ji} Centre of the RBF unit for input variables
σ_i Width of the ith RBF unit
$x_{ji} = j_{th}$ Variable of input pattern

Output of the neural network is given as

$$Y_{net} = \sum_{i=1}^{H} w_{im} v_i(x_i) + w_0 \tag{6}$$

where,

H Number of hidden layer nodes (RBF function)
Y_{net} Output value of m_{th} node in output layer for the n_{th} incoming pattern
w_{im} Weight between i_{th} RBF unit and m_{th} output node
w_0 Biasing term at n_{th} output node

Then error is calculated after stopping condition of the network. Stopping condition includes the weight change and numbers of epochs.

3.4 Processing and Preparation of Data

The monthly rainfall, maximum monthly temperature, minimum monthly temperature are collected from India meteorological department, and daily runoff data are collected from soil conservation department Sambalpur, India, for the period of the monsoon months (May to October), from 1993 to 2016. The data from 1993 to 2010 are used for training and data from 2011 to 2016 are used for testing the network. Daily data are converted into monthly data, which is finally used for

training and the testing the model. The input and output data are scaled in such a way that each data plunge within a specified range before training. The process involved is called normalization so that the normalized values are bounded within the range of 0 to 1. The normalization equation used for scaling the data is

$$X_t = 0.1 + \frac{X - X_{min}}{X_{max} - X_{min}} \tag{7}$$

where $X_t =$ transformed data series, X = original input data series, $X_{min} =$ minimum of original input data series, $X_{max} =$ maximum of original input data series.

3.5 Evaluating Criteria

The evaluating criteria to ascertain the best model are coefficient of determination, mean square error and root-mean-square error. To choose the ideal model for this study area, the condition is MSE, RMSE should be least and coefficient of determination should be highest.

$$\text{Co} - \text{efficient of determination } \left(\text{R}^2\right) = \left[\frac{\text{n}\sum \text{xy} - (\sum \text{x})(\sum \text{y})}{\sqrt{\left[\text{n}\sum \text{x}^2 - (\sum \text{x})^2\right]\left[\text{n}\sum \text{y}^2 - (\sum \text{y})^2\right]}}\right]^2 \tag{8}$$

The value of coefficient of determination indicates the percent of variation in one variable explained by the other variable.

$$\text{Mean squared error (MSE)} = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y_i} - y_i\right)^2 \tag{9}$$

where,

$\widehat{y_i}$ Predicted value of runoff
y_i Actual value of runoff

$$\text{Root mean squared error (RMSE)} = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{x_i} - x_i\right)^2\right]^{1/2} \tag{10}$$

RMSE is the root-mean-squared error. RMSE has also nonnegative value and closer to zero is the best performance.

4 Results and Discussion

4.1 Results at Saintala

Here to evaluate the model efficiency of various architectures, different transfer functions like tangential sigmoidal and logarithmic sigmoidal are used to establish the model performance. For every architecture, the criteria for evaluation are mean square error training, testing, root-mean-square error training, testing and coefficient of determination. In Table 1 for Tan-sig function in FFBPN, 4-2-1, 4-4-1, 4-5-1, and 4-7-1 architectures are taken into consideration for computation of performance. For Tan-sig function, the best model architecture is found to be 4-5-1 which possesses MSE training value 0.000496, MSE testing value 0.002495, RMSE training value 0.03282, RMSE testing value 0.04174 and coefficient of determination for training 0.9781 and testing 0.9381. For Log-sig 4-2-1, 4-4-1, 4-5-1, and 4-7-1, architectures are taken into consideration for computation of performance. For Log-sig function, the best model architecture is found to be 4-2-1 which possesses MSE training value 0.000785, MSE testing value 0.006057, RMSE training value 0.04012, RMSE testing value 0.069052 and coefficient of determination value training 0.9423, testing value 0.9012. The comprehensive results are presented in Table 1.

Similarly for the radial basis function network the results are discussed below for Saintala station. For Tan-sig 4-0.2-1, 4-0.3-1, 4-0.5-1, 4-0.7-1 and 4-0.9-1 architectures are taken into consideration for computation of performance. For Tan-sig function, the best model architecture is found to be 4-0.3-1 which possesses MSE training value 0.000598, MSE testing value 0.005014, RMSE training value 0.02441, RMSE testing value 0.07053 and coefficient of determination training value 0.9937, testing value 0.9331 are taken into consideration for computation of performance presented in Table 2.

5 Simulation

The graphs with best values for runoff from precipitation, maximum temperature, minimum temperature, evapotranspiration using feed-forward back propagation (FFBPN) and radial basis function network (RBFN) are presented with Tan-sig and Log-sig transfer function. The best value for each of the evaluating criteria is represented. The graphs below show how these best values result in the variation between the observed runoff and predicted runoff.

Table 1 Results of FFBPN at Saintala

Model input	Sigmoidal function	Architecture (L-M-N)	MSE		RMSE		R^2	
			Training	Testing	Training	Testing	Training	Testing
Precipitation Maximum temperature Minimum temperature Evapotranspiration	Tan-sig	4-2-1	0.000609	0.005226	0.037363	0.054134	0.9613	0.9143
		4-4-1	0.000827	0.005736	0.032374	0.093876	0.9232	0.8871
		4-5-1	**0.000496**	**0.002495**	**0.03282**	**0.04174**	**0.9781**	**0.9381**
		4-7-1	0.000581	0.00409	0.02705	0.05675	0.9671	0.9013
	Log-sig	**4-2-1**	**0.000785**	**0.006057**	**0.04012**	**0.069052**	**0.9423**	**0.9012**
		4-4-1	0.001195	0.004088	0.03455	0.06394	0.9793	0.8114
		4-5-1	0.000991	0.008029	0.06122	0.083450	0.9115	0.8717
		4-7-1	0.000725	0.007336	0.07536	0.091133	0.9233	0.8934

Table 2 Results of RBFN at Saintala

Model input	Architecture with spread value (L-M-N)	MSE		RMSE		R^2	
		Training	Testing	Training	Testing	Training	Testing
Precipitation Maximum temperature Minimum temperature Evapotranspiration	4-0.2-1	0.000639	0.011779	0.02523	0.1089	0.9886	0.6203
	4-0.3-1	**0.000598**	**0.005014**	**0.02441**	**0.07053**	**0.9937**	**0.9331**
	4-0.5-1	0.000826	0.008557	0.02871	0.09252	0.9852	0.790
	4-0.7-1	0.000541	0.003128	0.02321	0.05593	0.9905	0.9622
	4-0.9-1	0.000970	0.003557	0.03112	0.05964	0.8345	0.6980

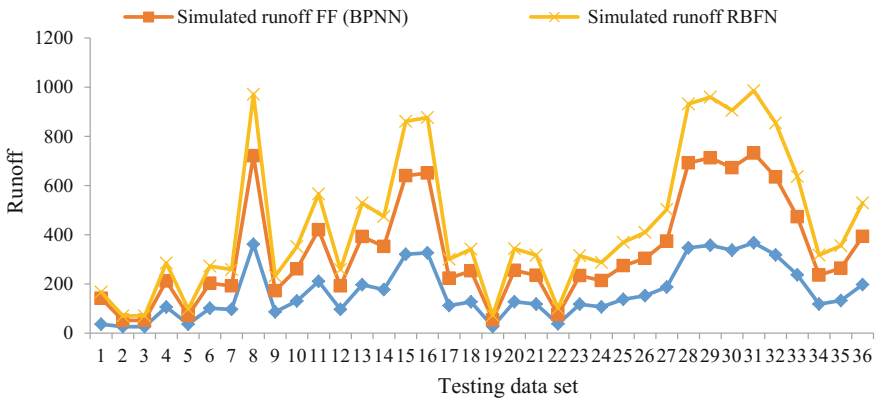


Fig. 6 Actual versus simulated runoff using FFBPN, and RBFN at Saintala in testing phase

5.1 Assessment of Actual Runoff Versus Simulated Runoff at Saintala During Testing Phase

The variation of actual runoff vs. simulated or predicted runoff is exposed in Fig. 6. Results show that the estimated peak runoff is 318.286 and 314.806 mm for LRN and FFBPN against the actual peak 328.3 mm for the watershed Saintala.

6 Conclusions

Study with different parameters like precipitation, evapotranspiration and temperature have been considered for predicting runoff. At Saintala watershed, among two neural networks with the evaluation criteria MSE, RMSE, and coefficient of determination, RBFN performs best with architecture 4-5-1 following Tan-sig transfer function by considering both training and testing performance criteria. Similarly, FFBPN performs best at architecture 4-5-1 with all criteria considered for model performance. This work will help for estimating the parameters related to watershed, for planning, designing and management of the watershed. The output of the research indicates the suitability of method for estimation of runoff in arid region. However, the combination technique needs to be investigated for improving integrated model techniques for future research.

References

1. Chen, J.Y., Adams, B.J.: Integration of artificial neural networks with conceptual models in rainfall-runoff modeling. J. Hydrol. **318**(1–4), 232–249 (2006). https://doi.org/10.1061/(asce)he.1943-5584.0000970
2. Dawson, C.W., Wilby, R.L.: Hydrological modelling using artificial neural networks. Progr. Phys. Geogr. **25**(1), 80–108 (2001). 10.1.1.906.3917
3. Lee, D.S., Jeon, C.O., Park, J.M., Chang, K.S.: Hybrid neural network modeling of a full-scale industrial wastewater treatment process. Biotechnol. Bioeng. **78**(6), 670–682 (2002). https://doi.org/10.1002/bit.10247
4. Smith, J., Eli, R.B.: Neural-network models of rainfall runoff process. J. Water Resour. Plann. Manag. ASCE **121**(6), 499–508 (1995)
5. Xiong, L., Shamseldin, A.Y., O'Connor, K.M.: A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system. J. Hydrol. **245**(1–4), 196–217 (2001). https://doi.org/10.1016/j.jhydrol.2003.08.011
6. Zakermoshfegh, M., Ghodsian, M., SalehiNeishabouri, S.A.A., Shakiba, M.: River flow forecasting using neural networks and auto-calibrated NAM model with shuffled complex evolution. J. Appl. Sci. **8**, 1487–1494 (2008). jas.2008.1487.1494
7. Zhao, R.J.: The Xinanjiang model applied in China. J. Hydrol. **135**(1–4), 371–381 (1992)

Non-invasive Heart Rate Measurement on Wrist Using IR LED with IoT Sync to Web Server

R. Pawan Sai and M. P. Sunil

Abstract This work focuses on monitoring the heart rate of a person in a non-invasive manner from the user's wrist. The principle concept used here is Photoplethysmography (PPG) which detects the change in blood volume as the heart pulsates. This setup includes an Infrared (IR)—detector pair that senses the pulses and a self-designed circuit that processes the collected signal. This data is understood by graphical visualizations on MATLAB and an intelligent algorithm is devised to estimate the dynamic heart rate of a person. For the simplicity of user interface, the calculated data is displayed on the TFT LCD screen and also recorded in the SD card. This recorded data is synchronized to the cloud (web server) via a Wi-Fi module. A front end, easy to understand graph is displayed for the user in the web to analyze the behavior of his/her heart, upon which medical remedies can be advised.

1 Introduction

Heart is the only organ in the human body that keeps functioning at all times from the time of birth till the inevitable death occurs. All other organs in the body take rest and rejuvenate while one sleeps, but the heart does not. It is essential for a person to ensure that their heart is functioning normally and to maintain a balance in their health. Even today, the cardiologists all over the globe do not have a satisfactory explanation on what really triggers the very first "heart pulse"—The sign of Life.

Now the following questions arise—How are we going to monitor the functioning of our Heart? What does the pulse rate indicate? What should be one's ideal pulse rate? What is the inference provided on our health system based on the

R. Pawan Sai (✉) · M. P. Sunil
Department of Electronics and Communication Engineering,
School of Engineering and Technology, Jain University,
Bengaluru, Karnataka, India
e-mail: pawansair@gmail.com

M. P. Sunil
e-mail: mp.sunil@jainuniversity.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_7

readings? What measures are needed to be taken if the observations are unfavorable as compared to our ideal pulse rate?

The following sections provide the answers to the above questions.

1.1 Heart Rate

The heart constantly contracts and expands, which gives the pulsating motion. The expansion of the heart indicates the movement of deoxygenated blood into the heart and the contraction indicates supply of oxygenated blood to the main blood stream to all other parts of body. Heart Rate is the rate at which the heart beats for every minute. The heart beat in our body is usually known as pulse rate. Whenever the heart pumps, we can feel the pulsing in some of the blood vessels in the body close to the skin surface such as neck, wrist, behind the ear, temple region and armpits. By counting these pulses against time, we can find our heart rate.

However, would it not be effective if a smart device can measure this for us with much more efficiency than measuring it ourselves?

1.2 Photoplethysmography (PPG)

PPG [1, 2] is an optical technique that detects the blood volume changes in the dermal layer of our skin where the arteries are present. In this work, we are using a reflective mode of PPG, i.e., the reflected beam is captured by the photo detector. Infrared LED is used as the optical source, since it's wavelength of 940 nm has a deeper penetration into the skin surface irrespective of texture and color of the skin. This IR LED transmits the IR signal through the surface of the skin on the wrist as a result of which the signal penetrates into the blood vessels and some part of signal which is reflected back as shown in Fig. 1. The reflected signal is detected by the photo detector [2]. When the heart pulsates, there is a change in blood volume, which results in a train of pulses at the photo detector output.

2 Literature Survey

There are many existing devices available in the market that focuses on pulse oximetry such as Dr.Trust, Newnik and Choicemmed MD300, but these devices have to be clipped onto the finger for the measurement. These devices also use the concept of PPG [1] using the IR LED and photo detector, but how long can a person have his fingers [3] clipped to measure his heart rate in a day. This provides a setback to measure one's heart rate when one is undergoing different emotional states. So, for continuous measurement, there are other devices such as Fitbit, Apple Watch,

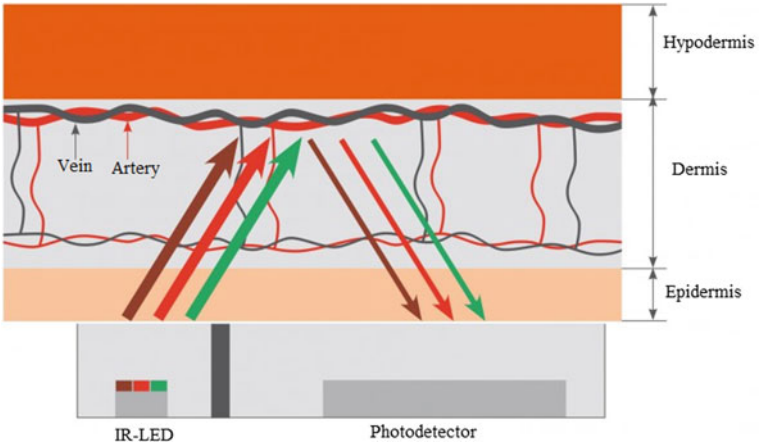


Fig. 1 Action of PPG, photo taken from Medical Design Technology [11]

Microsoft health band and many more that are wrist bands and track heart pulse on the wrist. However, by seeking out user reviews, these devices seem to perform inaccurately due to movements of the wrist. Also, these devices rely on Bluetooth to synchronize the recorded data to the computer to view logged heart rate.

Thus, this work eliminates the noise occurred due to wrist movements by applying the algorithm discussed in the following section. Also, this is the first device that embeds a Wi-Fi device and synchronizes the data to a web server as the user enters his/her regular Wi-Fi zone.

3 Interfacing Block Diagram

The interfacing block diagram of the device is as shown in Fig. 2. The sensor [4] circuit contains the SFH7050 [5] which is an IC that has inbuilt IR-photodetector pair [5]. The IR LED is isolated from the detector with a barrier between them to prevent crossover problem. This SFH7050 is placed directly on top of the wrist and the reflected IR beam is detected by the photodetector. However, this detected beam is very weak and cannot be processed. Therefore, this signal is amplified in two stages with gain of 100 at each stage. The IC MCP602 [6] has two internally in-built Op-Amps which are used here. Also, the maximum rate at which the heart can beat is known to be around 180 bpm (beats per minute), thus a low pass filter of 2.5 Hz is incorporated before every amplification stage. At the end of the amplification circuit, we will obtain a strong pulse that can be considered as a valid input to the Atmega2560 controller (8-bit) [7].

The code is programmed using Embedded C using the Arduino [8] IDE. The general functionality of the code involves, prompting the user to check his/her heart

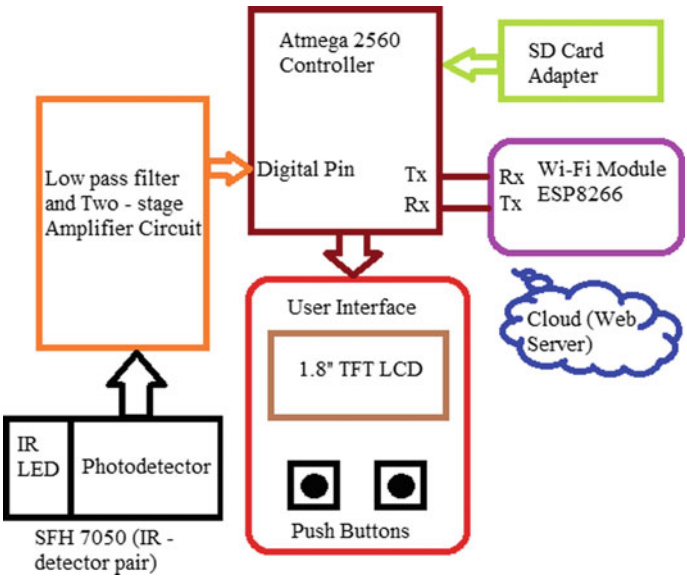


Fig. 2 Interfacing block diagram

rate. The user enters into the measurement mode by selecting the corresponding push button. The algorithm as discussed in the following section is devised into the program code, which calculates the dynamic heart rate and displays the result on the LCD. This is a low-power, 1.8" TFT LCD that is active at 3.3 V. Also, the measured data is recorded in the SD card. Upon, the input received from the user through the corresponding push button, this data is synced to the Cloud (Web Server) via the Wi-Fi module. The ESP8266 is the Wi-Fi module used and follows UART Interface. The detailed working of the ESP8266 and its interface to the microcontroller is discussed under Sect. 3.3.

3.1 Signal Conditioning Circuit

The detected IR beam by the photodiode is very weak and needs to be amplified and processed in order to understand and analyze the pulses. Hence, the output of the detector is given to the Signal Conditioning circuit as shown in Fig. 3.

In simple terms, the signal conditioning circuit removes any optical noise present and boosts the strength of the signal.

Since the reflected signal strength magnitude is very weak. So, it is difficult to read the output across the microcontroller. In order to increase the strength of this signal amplification and filtering circuit is used. Amplification and filtering circuit are used to increase the signal strength and to remove the noise in the signal; a low

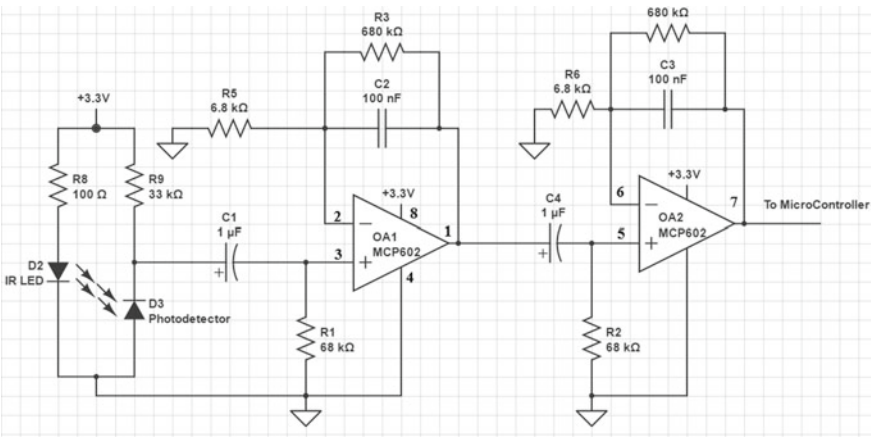


Fig. 3 Signal conditioning circuit

pass filter is used accordingly. The circuit diagram consists of two stage of signal conditioning circuit with an active low pass filters with a cut off frequency of 2.5 Hz. So, the maximum measurable heart rate is around 180 bpm.

For the purpose of amplification, an operational amplifier IC called MCP602 is used. It is a dual Op-Amp IC from microchip. The main advantage of using this IC is, it is single powered supply and it also provides rail to rail swing output. The filtering part removes the higher frequency noises present in the signal. The gain of each stage of amplification is 100, thus it will give the total amplification gain of 10000.

To avoid the DC components in the signal a capacitor of 1uF is connected at the input of each stage. The expression used to calculate the gain and cut off frequency of low pass filter is shown below.

The Gain of each stage is given as,

$$G = 1 + \frac{R_f}{R_i} \tag{1}$$

The Cut-off frequency of the low pass filter is,

$$f = \frac{1}{2\pi R_f C_f} \tag{2}$$

Thus, the gain of each stage is 101 as computed by Eq. 1 and the cut-off frequency is 2.34 Hz as computed by Eq. 2. The two stage amplifiers and filters circuit provide a required gain, to boost the weak signal detected by the photo detector sensor and convert it into a pulse. The detected output is given to the digital Pin of ATMEGA2560 microcontroller which will read the pulse and the following algorithm is designed to calculate the heart rate.

3.2 *The Pulse Algorithm*

This algorithm as shown in Fig. 4, has an efficiency of ±1 bpm. However, it is recommended for the wrist to be at rest while taking the measurement for accurate readings otherwise, we will get incorrect readings due to optical noise while moving. The duration between successive pulses is measured using Serial Port of Arduino IDE, as shown in Fig. 5 and the pulses are graphically viewed in MATLAB waveform window as shown in Fig. 6.

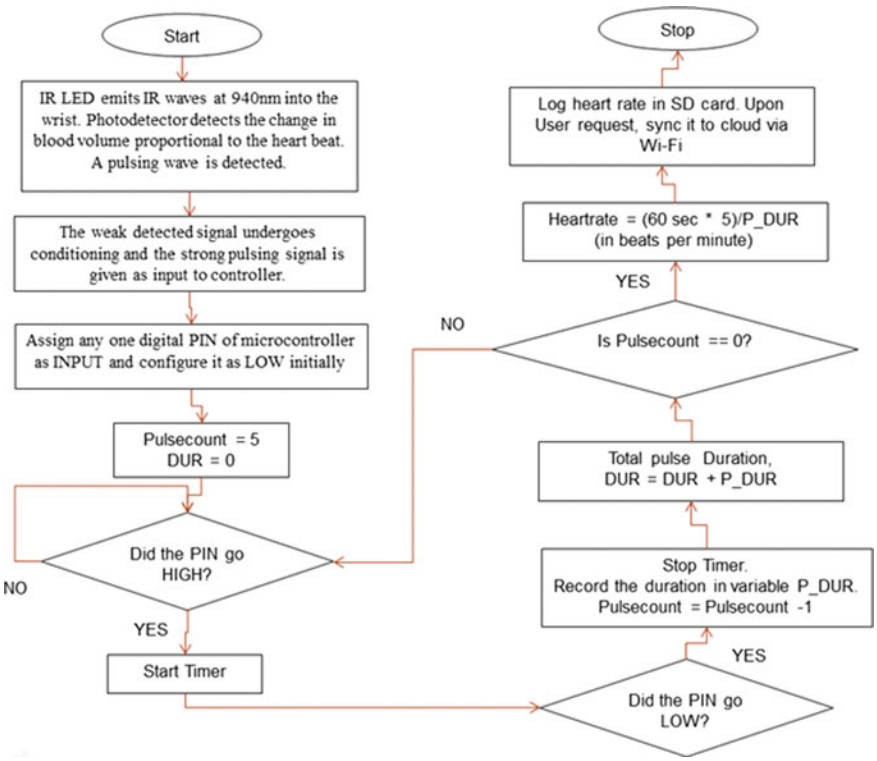


Fig. 4 Flowchart of the pulse algorithm

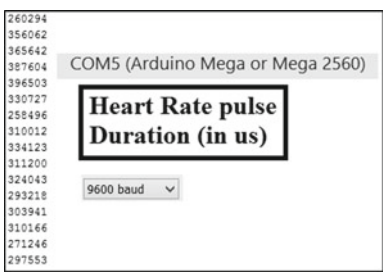


Fig. 5 Duration between successive heartbeats

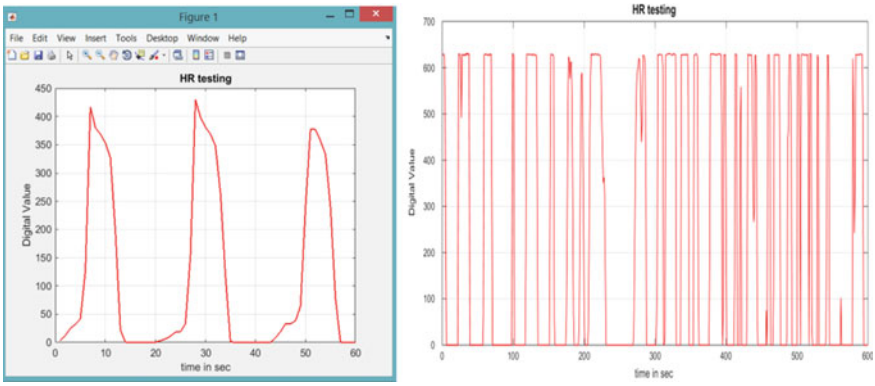


Fig. 6 MATLAB [1] output of Heartbeats

3.3 Wireless Synchronization Using ESP8266 Wi-Fi Module

Wi-Fi is one of the most prominent networks and has a wide future. It is highly user-friendly for the user to sync his data to the cloud upon entering a Wi-Fi zone. Also, the portability of data is very easy as a small link on the Web Server can be shared to anyone especially, our personal family doctors.

The ESP8266 [9] is a Wi-Fi module that can be powered up between 1.8 and 3.3 V, hence a low power device. Wi-Fi is an asynchronous communication model and the ESP8266 communicates at 115200 [9] bauds per second. This is also known as the baud rate. The UART interface to the microcontroller is simple, where the Tx and Rx are crossed between the controller and the module. ESP8266 listens and responds to the Attention (AT) commands. Hence, the program code has the following list of AT commands that sets up the connection by using the TCP via channel 0.

Firstly, the Wi-Fi module acts as a station and connects to a nearby Wi-Fi Access Point (AP), whose log in credentials are specified in the code. The measured heart rate data are logged in the SD card in a .txt file. This data is accessed and the total size of data to be transmitted is calculated in bytes and upon receiving the SEND command, the data is transmitted to the Web Server or the IP address specified in the program code.

Once the data is received at the web server, an acknowledgment is received by the module and the microcontroller indicates to the user that the synchronization has been completed via the LCD.

The user-friendly graphical representation of the data can be done using Web development tools and it is not in the scope of this paper.

The following are the list of AT commands specified in the program code. The ESP8266 module acts as a TCP client while sending data.

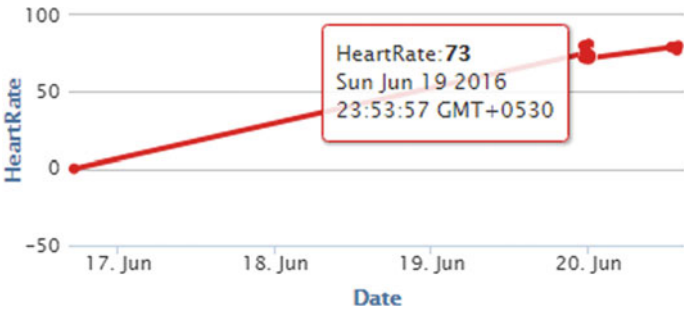


Fig. 7 Graphical representation of the Heart rate data over many readings

$$\mathrm{AT} + \mathrm{CIPMUX} = 1$$

This enables multiple connections to be established.

$$\mathrm{AT} + \mathrm{CIPSTART} = 0, \mathrm{TCP}, \ 54.86.175.203, 80$$

The communication is started via channel 0 using the TCP protocol through Port 80 to the specified IP address.

$$\mathrm{AT} + \mathrm{CIPSEND} = 0, 48$$

48 bytes of data are sent via the channel 0. Then we will get a > prompt. Then we need to enter the following command.

$$> \mathrm{GET}/\mathrm{update?key} = 3\mathrm{G5QIEOLKJTDV8Y6\&field1} = 74\,\mathrm{HTTP}/1.1$$

A GET request is sent to the mentioned API key and the measured heart rate 74 is sent to the web server.

Then we wait for the acknowledgement of SEND OK and bytes received. After which, we terminate the connection using the following command–

$$\mathrm{AT} + \mathrm{CIPCLOSE} = 0$$

The graphical data obtained on the web is shown in Fig. 7.

4 Results

The device is strapped to the wrist of the user with the help of Velcro. The user then navigates to the Heart rate calculator icon on the LCD via push buttons. Once the user triggers the start button, the device starts reading the reflection of IR beam from the wrist which is detected by the photo detector and starts recording the

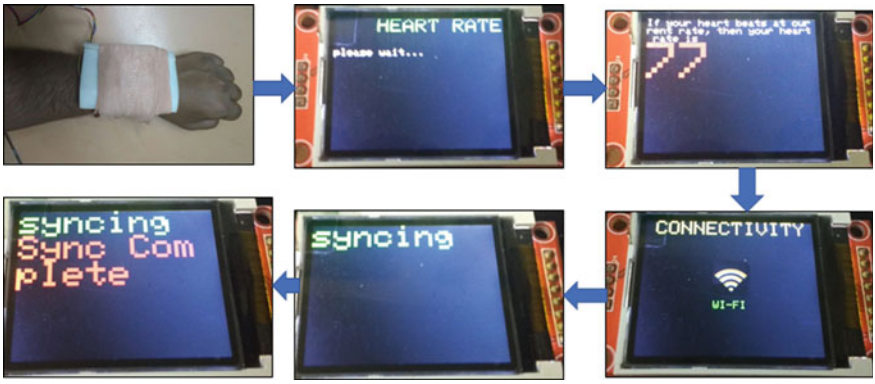


Fig. 8 User operation to measure heart rate and sync it to web server

Table 1 Comparison of results between device and manual measurements

Trials	Heart rate measured using device (bpm)	Heart rate measured manually (bpm)	Accuracy (in %)
At rest	74	74	100
At rest	79	76	96.2
After walking	80	79	98.75
After lunch	73	75	97.3
After jogging	81	81	100

pulses. Then it uses The Pulse Algorithm as explained in previous sections to calculate the Heart rate and displays the result on the LCD and stores the result in the SD card.

Further, the user navigates to the Wi-Fi icon and turns on the Wi-Fi, which connects the device to the Access Point. The calculated Heart rate are retrieved from the SD card and synced to the Web Server and the user is notified when the sync is complete. The user can then log onto the web server and view the graphical details of the heart rate recorded at different times. Figure 8 portrays the above discussion.

To ensure the accuracy of the device, heart rate measurements were recorded after different activities in a day like walking, jogging, after having food and at rest. Simultaneously, heart rate was measured manually [10] using stethoscope or counting the radial pulse. Both the recorded and measured results were compared and the device was found to have accuracy above 95%. The same has been illustrated in Table 1.

5 Conclusion

Thus, this work has described and implemented a real-time heart monitoring system. The data has been successfully synced via Wi-Fi to a remote web server. Also, the user-friendly graphs are provided so that the user can focus to improve his/her health, since heart rate contributes to a majority of human health. Each user has been given separate login credentials through which they can check their data separately. The PCB is self-made and thus the overall design is an effective low-cost device.

6 Compliance with Ethical Requirements

Ethical approval: "All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."
Informed consent: "Informed consent was obtained from all individual participants included in the study." Authors have taken the consent from the Department of Electronics, SET, Jain University, to undertake the experiments.

Corresponding author will be solely responsible if any issues arise in future with regard to this.

References

1. Poh, M.Z., McDuff, D.J., Picard, R.W.: Non-contact, automated cardiac pulse measurements using video imaging and blindsource separation. Opt. Express **18**(10), 10766 (2010)
2. Gansloser, J., Seepold, R.: Development of an algorithm and a sensor to monitor the heart rate by volumetric measurement techniques. LNE, vol. 392 (2016)
3. Mallick, B., Kumar, A.: Heart rate monitoring system using finger tip through arduino and processing software. IJSETR **5**(1) (2016)
4. Sun, L., Ma, H., Fang, D., Niu, J., Wang, W.: Lifetime optimization Algorithm with Multiple Mobile Sink Nodes for Wireless Sensor Networks, CWSN (2014)
5. SFH7050 Datasheet, OSRAM Opto Semiconductors
6. MCP602 Datasheet, Microchip Technology Inc
7. Atmega2560 Datasheet
8. Plaza P., Sancristobal E., Fernandez G., Castrom M., Pérez, C.: Collaborative robotic educational tool based on programmable logic and Arduino. In: 2016 Technologies Applied to Electronics Teaching (TAEE), Seville, pp. 1–8 (2016)
9. ESP8266 Datasheet, Express If systems
10. Ramesh, K., Aswin Kumar, S.V.: Efficient health monitoring system using sensor networks. IJSERV **3**(6) (2012)
11. Medical Design Technology. https://www.mdtmag.com/article/2015/01/light-sensors-fitness-tracking

Analysis of Constrained Navigation in a Bounded Smart City

Arihant Sinha and Jimmy Mathew

Abstract Today in the world of smart cities, with growing infrastructure and public demand rising, there is an immediate need to supply the demands at various points by any single organization. In this course of visiting such large number of points, there is a possibility for the organization to cover some avoidable points which has no valuable outcome. This creates the problem of time and effort optimization which is very important with this growing scale. In this paper, we propose the idea to solve the above-stated problem by using the IoT sensors to check which all points need to be covered and eliminate such avoidable points from the course of our path. After reducing the points of travel, the remaining points can be covered optimally using Travelling Salesman Problem algorithm. In this paper, we have illustrated the problem of waste collection in smart cities in which there are multiple waste bins all over the city but only those bins are to be emptied that are suitably filled. This leads to reduction of waste of valuable resources like fuel, time and effort.

1 Introduction

With sprawling urban area and ever increasing urban population, the problem of waste collection is exacerbating [3]. In many parts of the world, this neglect has led to spread of preventable disease and hazardous pests like mosquitos, flies, etc., which kill millions annually. Poor waste management adversely also negatively affects the tourism industry. The current waste management practices are completely manual, opaque and inefficient. Seldom checks are made to ensure waste is collected on time and a lot is left on discretion of the waste collectors. This leads to a highly irregular and poor waste management system in the city.

A. Sinha (✉) · J. Mathew
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore Campus, Vellore 632014, Tamil Nadu, India
e-mail: sinhaarihant@gmail.com

J. Mathew
e-mail: jimmym@vit.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_8

The current work in the field is leveraging *Internet of Things* (*IoT*) to create smart waste management systems. Existing smart bin products are based on ultrasonic sensors to detect the level of waste in a bin and a GPS module to report the exact location, which is sent to the server with the help of a GSM module or a Wi-Fi module. The use of GPS module and GSM module makes capital investment for a bin very high. Hence, this technology is feasible for small-scale area deployment like malls, parks but such a high capital cost makes it infeasible for large-scale deployment like public waste management of a municipality.

With the rising number of smart cities across the globe, this paper proposes to use IoT to create an automated, affordable and efficient waste management system. The proposed bin is equipped with a sensor and Wi-Fi module to communicate its status (full/empty). The garbage truck driver can use an Android application called *SmartBin* app which retrieves the status data of the bins for the centralized server and provides the driver with an optimized path covering only the bins which are full. This system reduces the time, effort and fuel spent on each trip of the garbage truck and makes the process of waste management automated, efficient and transparent.

2 Literature Survey

The garbage dumps and littering make the area a breeding ground for germs and pests like mosquitoes which spread bacterial diseases like malaria. The current practice of waste management across the globe has been detailed by Hoornweg, Daniel and Perinaz Bhada-Tata [4]. The waste management in municipalities has been analyzed by Morrissey et al. [6]. The idea of automation in smart waste collection and real-time waste monitoring has been discussed by Catania et al. [3]. The introduction of IoT has made this idea practical and feasible. The use of IoT for this domain has been proposed and demonstrated by Anagnostopoulos and Theodoros [2] for the data of St. Petersburg City. The idea of using the technology of Wireless Sensor Nodes in this field has been explained by Longhi and Sauro [5].

The SmartBin by SmartBin Corp [8] is a product in this domain which uses ultrasonic sensors with GSM/Wi-Fi module to send data over to the Internet. Smart Bins by GPS Integrated [9] is a product that has a smart bin equipped with an ultrasonic sensor and GPS module and uses Wi-Fi or 3G modules to report the bin status over the Internet. The use of GPS module and GSM module makes capital investment for a bin high. Hence, this technology is feasible for small-scale area deployment like malls, parks but such a high capital cost makes it infeasible for large-scale deployment like public waste management of a municipality.

The development of waste management system is a clear objective and a special focus area for smart cities. The creation of a separate waste management for handling hazardous waste has been analyzed by Anagnostopoulos and Theodoros [1].

3 Proposed Method

This paper focuses on cost reduction with minimal loss of functionality by replacing GPS and ultrasonic sensors with more cost-effective method. The proposed smart bin is equipped with IR transmitter and receiver pair, an operational amplifier (opamp) IC LM741 and a Wi-Fi module ESP8266 Nodemcu version 12E. The IR transmitter and receiver pair is arranged diametrically opposite to each other as shown in Fig. 1. The circuit diagram has been given in Fig. 2. The Wi-Fi module reports the status to the 'ThingSpeak' server through Wi-Fi network. The server stores the status of all bins. The SmartBin app reads the status of the bins from the server. The application queries *Google Maps Distance Matrix* API to create a distance matrix. Using the matrix, the application solves the *Travelling Salesman Problem* (TSP) using dynamic programing recursive solution. The final route is then displayed using *Google Maps Directions API* (Application Programming Interface).

If the level of the waste in a garbage bin is full, then the IR rays are blocked between the IR module pair, which results in the low current in the IR circuit and thus low voltage output. The current in the IR circuit depends on factors like if any obstruction comes in *Line of Sight* (LOS) of the IR pair or distance between the IR pair increases. To define the voltage threshold for the bin to be full or empty, IC741 opamp has been used in comparator mode. In this mode, IC takes two inputs, one from the IR circuit and the other for reference voltage provided from the potentiometer. It checks whether the input from the IR circuit is greater than reference voltage and draws the output voltage high and vice versa. The status for the bin to be full or empty can be calibrated for a given width of the bin by adjusting the reference voltage from the potentiometer; thus, it provides flexibility for the width of the bin. The output of the comparator is taken as input at analog pin of ESP8266. The ESP uploads the data to the ThingSpeak server.

Now the Android application reads the data from the ThingSpeak server and makes a list of full bins. From the list, an adjacency matrix is created with origin, all full bins and destination in that order. If bin number is greater than 8, the adjacency matrix has split into blocks of 10 × 10 matrixes as the Google Maps Distance

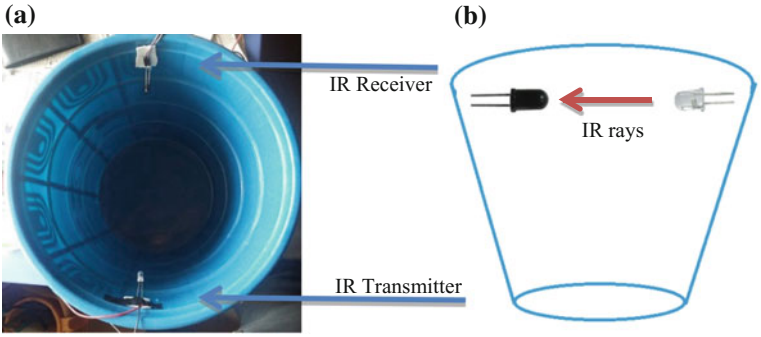


Fig. 1 **a** Top view of the bin and **b** side view of the bin

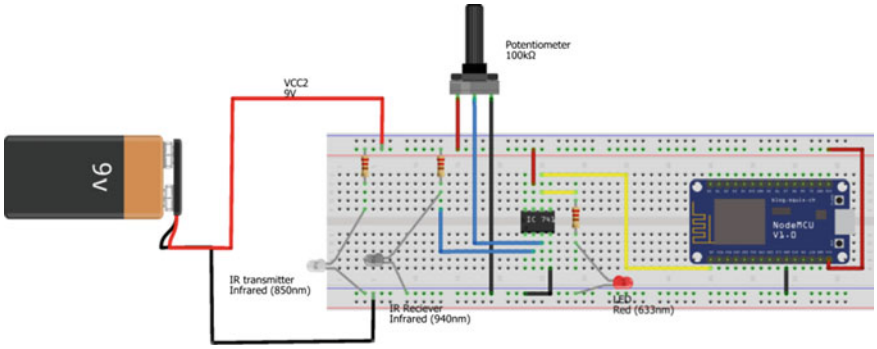


Fig. 2 Circuit diagram

Matrix does not return more than 100 elements in one request. Once the adjacency matrix is populated, the cost of traveling from destination to origin equal has been set to zero. If there are x full bins, then the matrix is of the order (x + 2) * (x + 2):

$$Matrix[x+2][0] = 0$$

Now a TSP is created with an adjacency matrix. The dynamic programming recursive solution is then used to solve the TSP [7]. Once the path sequence is found, Google Maps Directions API is used to display the route. Again Google Maps only allows up to ten waypoints, so the path sequence is split into multiple ten points query, with the destination of the previous query as origin of the next query.

In place of using GPS module, which adds unnecessary cost, a three-step installation process is preferred. First, install the designed smart device in the bin. Second, use the Add Bin activity in SmartBin Android app to add exact GPS coordinates of the bin. Third, when a bin is added, it is given an ID. The basic assumptions in this method are: (1) availability of public Wi-Fi in the smart city (2), no interchanging of bins and (3) protection of designed smart devices from environment factors like rain, etc.

4 Test Setup

The main factors that would affect the smart bin system efficiency would be

(1) **Ratio of number of full bins to number of total bins**: The higher the ratio, the lower is the efficiency offered by the optimization algorithm.
(2) **Bin density**: The number of bins per unit area directly affects the path set to be traverse.
(3) **Bin Distribution**: There are two types of distribution:

a. **Linear**: Bins are located on the same path without much U-turns and without much path change.
b. **Distributed**: Bins are located at different points in a nonlinear arrangement

There are three conditions that have been designed for *Vellore* town in Tamil Nadu, India. In each condition, different distribution patterns are taken to manipulate the bin density with 13 bins in each case to limit the complexity. There are 14 points that are chosen on the map. Origin of the path is 'garbage truck parking' and given coordinates as 0, whereas the destination 'garbage dumping ground' is marked 14. The bins that are distributed all over the area are marked as 1–13. In Conditions 1 and 2, origin and destination are assumed to be the same point as in many towns garbage truck is parked at dumping ground, whereas in Condition 3 they both are different.

Table 1 describes all data differentiating each condition and shows path followed by truck driver without the use of IoT or TSP algorithm. The green markers are bins which are not full and red markers are full bins; blue markers indicate origin and destination. Condition 2 can be found in rural areas with long linear roads and dense settlement among the roadsides, Condition 1 can be found in highways connecting cities where bins are placed on linear roads and Condition 3 will be found in urban areas with many small roads and extremely high population density.

Each case is subjected to three paths that driver follows based on technology used for further optimizing the route:

a. **Normal Path**: The current scenario, i.e., use of neither IoT nor TSP.
b. **IoT path**: In this scenario, the IoT devices are installed in each bin and the path optimization algorithm, the TSP, is not used. Driver uses the same sequence of bins as in normal path but simply skips empty bins, for example if there is a path in the order of bin 1, 3, 2, 5, 7, 4, 6 and bins 2 and 4 are not full, driver simply covers 1, 3, 5, 7, 6 but he optimizes the path to be covered between 3 and 5 (2 skipped) to save time.

Table 1 Case study setup

Case studies	Densely populated and linear arrangement bins	Sparsely populated and distributed arrangement bins	Densely populated and distributed arrangement bins
Bin density (approx.)	13 bins in 32 km^2	13 bins in 43 km^2	13 bins in 30 Km^2
Are origin and destination same?	No	No	Yes
Normal path sequence	0, 1, 2, 3, 4, 5,6, 7, 10, 8, 9, 11, 12, 13, 14	0, 5, 6, 7, 4, 1, 3, 2, 8, 9, 10, 12, 11, 13, 14	0, 10, 9, 11, 12, 8, 6, 4, 5, 7, 13, 3, 2, 1
Normal path time (in sec)	4034	6626	4384

c. **TSP-Optimized path**: In this scenario, TSP algorithm is added in the above IoT path. TSP solves the problem of path optimization that is faced in above case. It covers all the points where bins are full and driver helps to follow an optimized path.

5 Results

The results of the three case studies are:

Condition 1: Bin density is high and bin arrangement is linear
In this condition, there is significant advantage gained by the use of IoT technology but the benefit of TSP optimization after the use of IoT is neither significant nor regular.

In a linear arrangement, even if there one bin on the path is full and the rest are empty, the truck has to still cover the path, for example the third test case of Table 2, despite having two bins which are not full, exactly same path as Fig. 3 is covered by the driver.

Condition 2: Bin density is low and bin arrangement is nonlinear
From Table 3, the use of IoT technology provides significant benefit is most cases but the use of TSP optimization over IoT is significant only in some cases. The savings range from below 3% to above 30%.

In the Sr. No.2, IoT path faces the problem of U-turns on driver's route. As shown in Fig. 4, in normal path truck covers bin 2, takes a U-turn from bin 5 to cover bins 6 to 8 but as bin 7 is empty, the U-turn can be avoided by altering the sequence of bins covered.

Condition 3: Bin density is high and bin arrangement is nonlinear
In this condition, the advantage gained by the use of IoT is great and the advantage of TSP optimization after the use of IoT is also significant and regular unlike in the previous cases (see Table 4).

Table 2 Densely populated and linear arrangement

Sr. No.	Bins not full	Normal Path	IoT path		TSP-optimized path			
			Time	Saving (%)	Order	Time	Saving (%)	Over IoT (%)
1	13	4034	3895	3.4	0, 1, 2, 3, 4, 5, 6, 10, 9, 8, 11, 12, 14	3895	3.4	0.0
2	8, 9, 10, 11	4034	3667	9.1	0, 1, 2, 3, 4, 5, 6, 7, 12, 13, 14	3667	9.1	0.0
3	4, 5	4034	4034	0.0	0, 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 14	4034	0.0	0.0
4	2, 7	4034	3655	9.4	0, 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14	3626	10.1	0.8
5	7	4034	3725	7.7	0, 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 14	3696	8.4	0.8

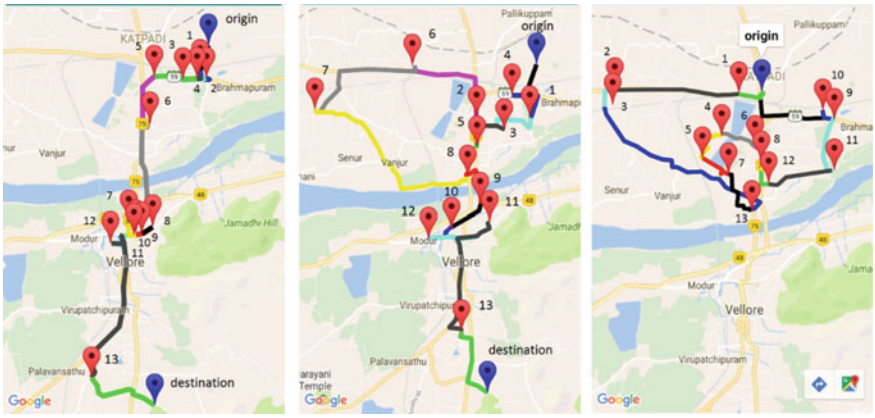


Fig. 3 Bin positions in case 1, 2 and 3 from right

Table 3 Sparsely populated and distributed arrangement

Sr. No.	Bins not full	Normal Path	IoT path		TSP-optimized path			
			Time	Saving (%)	Order	Time	Saving (%)	Over IoT (%)
1	11	6626	5928	10.5	0, 5, 6, 7, 4, 1, 3, 2, 8, 9, 10, 12, 13, 14	5928	10.5	0.0
2	3, 7, 11	6626	5067	23.5	0, 5, 6, 4, 1, 2, 8, 9, 10, 12, 13, 14	4570	31.0	9.8
3	2, 4, 10	6626	6174	6.8	0, 1, 3, 5, 6, 7, 8, 9, 11, 12, 13, 14	6174	6.8	0.0
4	6, 9, 10	6626	6350	4.2	0, 5, 7, 4, 1, 3, 2, 8, 11, 12, 13, 14	6300	4.9	0.8
5	7, 8, 9, 10	6626	4974	24.9	0, 5, 6, 4, 1, 3, 2, 11, 12, 13, 14	4924	25.7	1.0

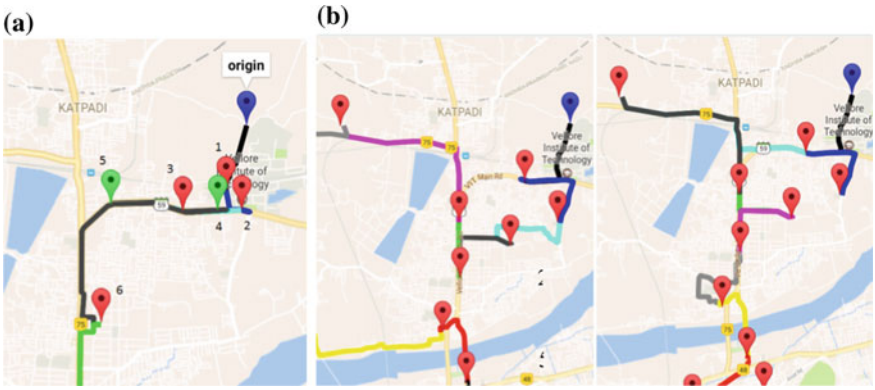


Fig. 4 a. Linear route covers empty bins b. U-turn exception

Table 4 Densely populated and distributed arrangement

Sr. No.	Bins not full	Normal Path	IoT path		TSP-optimized path			
			Time	Saving (%)	Order	Time	Saving (%)	Over IoT (%)
1	4, 13	4384	3609	17.7	0, 10, 9, 11, 12, 6, 8, 7, 5, 3, 2, 1	3468	20.9	3.9
2	6, 8	4384	4248	3.1	0, 10, 9, 11, 12, 13, 7, 5, 2, 3, 1	4037	7.9	5.0
3	2, 3, 6	4384	3898	11.1	0, 10, 9, 11, 12, 8, 13, 7, 5, 4, 1	3537	19.3	9.3
4	2, 3, 13	4384	3347	23.7	0, 10, 9, 11, 12, 6, 8, 7, 5, 4, 1	3309	24.5	1.1
5	4, 7	4384	4208	4.0	0, 10, 9, 11, 12, 6, 8, 13, 5, 3, 2, 1	3970	9.4	5.7
6	4, 5, 7	4384	3760	14.2	0, 10, 9, 11, 12, 6, 8, 13, 3, 2, 1	3760	14.2	0.0

By comparing the three tables, it is evident the best case to go for IoT technology and TSP-optimized path is densely populated region with well-distributed bins like urban part of cities.

6 Discussions

The test cases were designed to access the impact of the different factors and test how the saving achieved by the methods discussed in this paper fares in different real-world scenarios and where they are most appropriate to use.

Some special cases that can arise during the experiments require an explicit mention. In linear bin arrangements, sometimes a case arises where if a single road has 'X' bins and only a single bin is full, yet the truck needs to traverse the road, as the road is the only path to the bin. In such cases, if 'X' bins lie on a single road out of 'N' total bins, effectively there exist only $(N - X + 1)$ bins to cover and the truck has to traverse through $(X - 1)$ bins unnecessarily, if anyone of the 'X' bins is full, for example third test case in Condition 1. Another special case arises when U-turns are required to cover some bins in high bin density area. In such cases, the bin covered after a U-turn is not full; then, the impact of TSP optimization is very high as it can alter the sequence of bins, for example second test case for Condition 2.

The reason for high-density nonlinear arrangement showing larger benefit than low-density linear is that when IoT and TSP optimization are used, there exist many unique paths between the bins in a nonlinear arrangement and high-density areas, but in low-density linear bins there is a higher probability of only one road covering

the bins. So in the sparsely populated distribution, the TSP optimization is unlikely to change the path sequence as there is limited choice but in densely populated region the path sequence can change frequently as there are a large number of smaller roads.

7 Conclusion

The use of IoT-equipped bins with TSP optimization is most suited for urban areas with high bin density and nonlinear bin arrangements. The use of IoT technology results in significant benefit in most cases but the advantage of TSP algorithm is seen in some particular situations where the sequence of pints is altered. The method and technology discussed in the paper are most suited for densely populated areas with high density of bins. Having lower upfront capital cost with nearly same efficiency of the existing solutions makes this method most suited for use in smart cities.

References

1. Anagnostopoulos, T., Kolomvatsos, K., Anagnostopoulos, C., Zaslavsky, A., Hadjiefthymiades, S.: Assessing dynamic models for high priority waste collection in smart cities. J. Syst. Softw. **110**, 178–192 (2015)
2. Anagnostopoulos, T., Zaslavsky, A., Medvedev, A.: Robust waste collection exploiting cost efficiency of IoT potentiality in Smart Cities. Recent Advances in Internet of Things (RIoT), IEEE, (2015)
3. Catania, V., Ventura, D.: An approach for monitoring and smart planning of urban solid waste management using smart-M3 platform. In: Proceedings of 15th Conference of Open Innovations Association FRUCT, IEEE (2014)
4. Hoornweg, D., Bhada-Tata, P.: What a Waste: A Global Review of Solid Waste Management (2012)
5. Longhi, S., Marzioni, D., Alidori, E., Di Buo, G., Prist, M., Grisostomi, M., Pirro, M.: Solid waste management architecture using wireless sensor network technology. In: 2012 5th International Conference on New Technologies, Mobility and Security (NTMS). IEEE (2012)
6. Morrissey, Anne J.: Jim Browne.: Waste management models and their application to sustainable waste management. Waste Manag. **24**(3), 297–308 (2004)
7. Final Report - Solving Traveling Salesman Problem by Dynamic Programming Approach in Java Program Aditya Nugroho Ht083276e Uploaded by Nugroho Aditya. (2010)
8. Smart bin, https://www.smartbin.com/
9. Smart Bins for Smart City, https://www.gpsintegrated.com/smart-bins-for-smart-city

Improving Social Safety with Automobile Pilot Adroitness Analyzer

Bhupesh Singh, Prashanth Balasubramanian, Jimmy Mathew and Binu John

Abstract The number of losses and fatal accidents in developing countries, where traffic rules are not paid much attention to, due to the recklessness of drivers is far too great. Thus, a thorough analysis of driving behavior and driving habits should be performed before issuing a driving license to any driver. The current license issue process involves manual checking of the proficiency of a driver. This kind of process has several loopholes. To avoid such shortcomings, a low-cost system is developed, which digitally monitors a driver's habits. It considers some very important and common metrics that a driver is expected to follow, before qualifying for one's proficiency. A classification is made based on all of the logged data, using various data analysis algorithms. The same can be used to determine whether a driver is fit to drive or not.

1 Introduction

In this paper, we propose an Automobile Pilot Adroitness Analyzer (APAA) system, which will carefully highlight all the positive and negative traits of the driver. The system comprises of several sensors that checks if the most important and basic

B. Singh (✉) · P. Balasubramanian · J. Mathew
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore Campus, Vellore 632014, Tamil Nadu, India
e-mail: bhupesh2193@gmail.com

P. Balasubramanian
e-mail: prash_94@hotmail.com

J. Mathew
e-mail: jimmym@vit.ac.in

B. John
SCMS School of Engineering and Technology, Karukutty, Kerala, India
e-mail: binujohn86@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_9

necessary requirements of being a proficient driver, match with the habits of the driver in the vehicle. This system, if adopted by the government and used throughout the country in driving skill tests, can have profound implications on the society as a whole where inefficient drivers will be off the road. Moreover, the system eliminates all the loopholes in a humanized driving license test as it is completely automated.

This article is organized as follows: In Sect. 2, the related articles published in literature are mentioned. In Sect. 3, the establishment of metrics, calculation and ranking process are explained. In Sect. 4, the design aspects of the proposed system are provided. In Sect. 5, the result obtained is discussed. Finally, concluding remarks are provided in Sect. 6.

2 Related Works

The published articles can be classified into the following divisions:

General approach. Some concepts relating to fuel efficiency to calculate optimal energy management strategies are discussed in [1].

Cognitive based. The ACT-R method in [2] involved psychoanalysis to model driver cognitive behaviour. Our method is different from ACT-R, as the end result is a scaled result after a thorough process of data mining and ranking.

Activity based. In [3], the analysis involved measure of operator activity data, environment data and operator condition data. This was in a manner completely different from the proposed method. The work mentioned in [4] involved behavior on extrinsic and intrinsic behaviors that are similar to the method in [3].

Sensor based. The work in [5] collected its output from just one factor, gripping force on the steering wheel. This is the only metric used here, and whole driver behavior prediction is based on this input. In [6], road based sensors are used to evaluate the performance of a driver. Also, a variety of driver fatigue detection methods using various sensors are mentioned in [7].

Image processing based. In [8], a system named as DBITE identified the behavior of the driver, but not his driving patterns. A very specific analysis for bus driver fatigue is performed in [9]. This too, is based on visual eye openness detected with a camera. In a similar approach, eye movement, yawning and mouth movement were detected in [10]. Another image processing approach to detect head rotation and nodding is performed in [11]. In [12], a combined analysis of head nodding, eye ball movement and electroencephalogram (EEG) is performed to detect drowsiness in drivers, in a simulated environment. An analysis of driver postures is performed in [13], to verify the skill of the driver.

Machine learning based. A Hidden Markov Model (HMM) based stochastic approach is adopted in [14]. In this model, the system is trained with many expert drivers. The system can then assess the skills of a novice driver.

In general, most of the approaches are based on image processing, which is subjected to complex analysis, costly equipment and availability of suitable lighting. Many other factors concerning the performance of the driving are not considered in the previous works.

3 Metrics and Ranking Methods

The most important metrics to analyze the proficiency of a driver is listed as follows:

Close contact with other vehicles or objects. Objects in front, in the rear and to the sides of the car are monitored continually using precise ultrasonic sensors. The threshold distance set for obstacle detection varies dynamically depending on the speed of the vehicle as given in Table 1. Also note that these values have been estimated for standard coupe/sedan vehicles, and not for heavy vehicles.

The frequency of close contact is then calculated, i.e., the number of times the driver comes in close contact with another vehicle, and breaches the threshold. In addition to it, the total time (duration in seconds) the vehicle was in close contact is also calculated as part of the generated report. A general overview of this procedure is shown in Fig. 1.

The threshold breach is not accounted for in the case of the vehicle rear, as other objects or vehicles' presence can hinder proficiency calculation. However, in test environments with two vehicles, the same conditions that are applied to the front of the vehicle, apply to the rear as well, as this is the case where both drivers are undergoing the test, and both vehicles are fitted with the APAA system.

Driver Pulse Analysis. A pulse sensor is attached to the pilot during the test, so that the pulse is monitored to determine if the driver is drowsy, incapacitated or drunk. An upper and lower threshold value for the pulse is set and in the case of a breach, the data is logged to the device memory. These data are used to calculate the

Table 1 Threshold distance ranges

Speed (km/h)	Distance (m)
0–15	1.5
15–30	6
30–45	15
45–60	23
60–75	38.25
75–90 and above	62

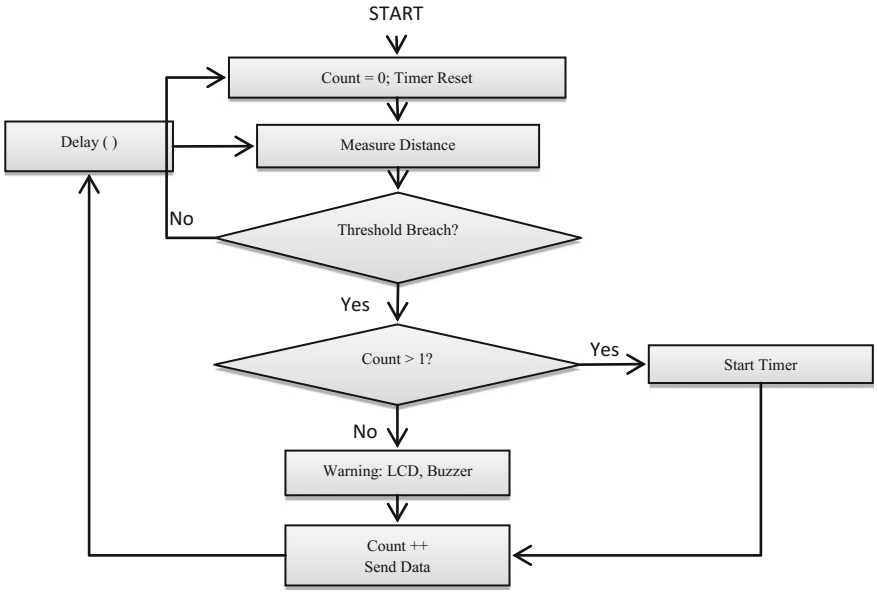


Fig. 1 Flow graph for Metric A

pulse of the driver based on a scaling factor α, and the following approximation is performed in Eq. 1:

$$\mathrm{P} = (1 - \alpha) * \text{sensor\_reading}. \tag{1}$$

Accelerator Pedal and Brake Pedal Diagnostics. A flex sensor attached physically to the accelerator pedal is used to determine the angle of the accelerator pedal at every second. There is no threshold value defined in this case, though the behavior of the driver is monitored through his use of the accelerator pedal. With this arrangement, any sudden application of accelerator can indicate rash driving, or gradual application for a smooth driving, can be detected. Similarly a flex sensor is attached to the brake pedal to determine the angle of the brake pedal at every application instance, to monitor any irregularities in brake usage.

RPM and Average Speed. A Hall Effect sensor is arranged near the wheels, to calculate RPM. With the RPM values at hand, the effective speed of the vehicle (in miles/hour) can be calculated as follows in the Eq. 2:

$$V = (0.00595 \times RPM \times r)/(R1 \times R2) \tag{2}$$

where, RPM = Number of revolutions per minute, r = loaded tire radius (wheel center to pavement), in inches, $R1$ = transmission gear ratio, and $R2$ = rear axle ratio.

4 The Design Aspects

The types of metric used in this system are summarized in Table 2.

The Pulse sensor functions on the basis of a Photoplethysmogram (PPG). PPG is a volumetric measurement of an organ. This is done by illuminating the skin with an IR light and measuring the changes in the intensity of the light, that is, the changes in light absorption.

Measurement of metric A is done by positioning three ultrasonic sensors along the front bumper of the vehicle. There are several cases that can arise while driving, namely those described in Fig. 2. In the first case (Fig. 2a) all three sensors give the exact same reading, within the limits of experimental error. Thus, the mean of three readings is considered for a breach. HC-SR04-Ultrasonic Sensor, Tower Pro SG90 Servo, PPG Pulse Sensor, Flex Sensor, Hall Effect Sensor, ATMEGA 328PU MCU is used in APAA.

In the second case (Fig. 2b), the sensor labelled (501) will have a reading much larger than either of (502) or (503), and is irrelevant to measurement. However, the mean of (502) and (503) will have a high likeliness of inaccuracy. Thus, if the discrepancy between the two values is not too great (± 5), the mean is considered, else the lower value is considered for ranking.

The third case (Fig. 2c) is similar to that of case two. The sensor labelled (603) will have a reading much larger than either of (601) or (602), and is irrelevant

Table 2 Metrics

Metric	Method
A	Based on threshold, number of times of breach and total duration
B	Based on pulse values, threshold breach and total duration
C	Pattern and trend analysis in graph of pedal angle versus time
D	Based on rpm and average speed

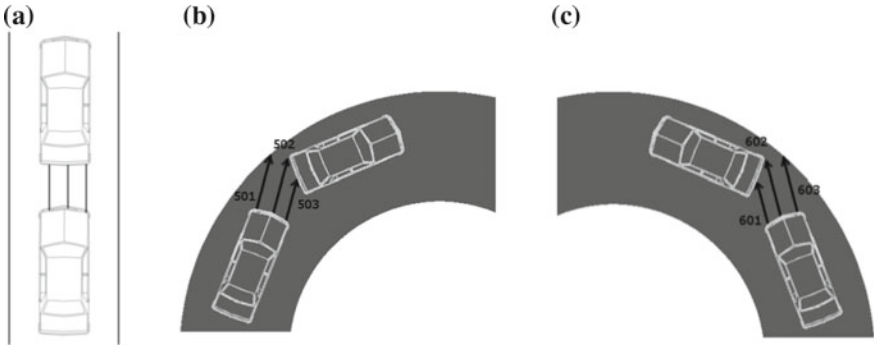


Fig. 2 **a** Case 1. **b** Case 2. **c** Case 3

Fig. 3 Message packet formats

to measurement. Still, the mean of (601) and (602) will have a high likeliness of inaccuracy. Thus, if the discrepancy between the two values is not too great (± 5), the mean is considered, else the lower value is considered for ranking.

In either of the above cases, a count of the number of times distance breached is measured in addition to the duration of each breach, along with a warning indication. This is shown in Fig. 1. In this system, A UDP based custom built protocol is used. Simple message types are used, and no authentication procedure is required. The following are the message types, which are shown in Fig. 3:

i. Enable device message (From server to device)—(801)
ii. Reset device message (From server to device)—(802)
iii. Request device status message (From server to device)—(803)
iv. Re-transmit request message (From server to device and vice versa)—(804)
v. Ping message (From device to server)—(805)
vi. Threshold breach/status message (From device to server)—(806)
vii. Disable device message (From server to device)—(807)

For enabling the device, the code used is 0x01 and for disabling the device, the code used is 0x06. These codes have a predefined meaning. The following are the different codes can be used 0x00—Reset Device, 0x01—Enable Device, 0x02—Request Status, 0x03—Retransmission Request, 0x04—Ping after one minute, 0x05—Threshold Breach Message, 0x06—Device Disconnect/Disable.

Every message contains a device id (1 byte), a message id (2 bytes) and a check sum (1 byte). Every device must have a unique id. The message id is used for tracking, logging and re-transmission purposes. If the check sum does not match with the received check sum, the re-transmission request is sent automatically. Some messages contain a code field (1 byte), which identifies the type of actions to be performed. For example, both enable device message and disable device message have the same message format. Nevertheless, the code used in it will distinguish what action is to be performed.

Figure 4 depicts the complete functioning of the communication state transition. The state diagram begins at the STANDBY state. Here, the device is powered on and is ready to operate. Once the device receives a valid ENABLE message, it enters into the START SENSING state. Here, the sensor values are continuously monitored for the occurrence of any of the following events:

- A reset message received: All sensor readings & counts get reset for fresh start.
- An invalid/error message received: An automatic re-transmission request message is sent to the server.

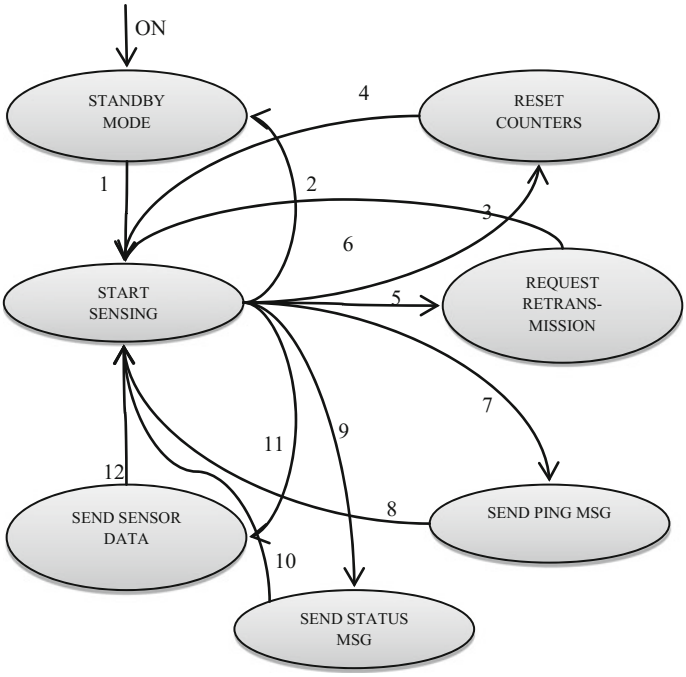


Fig. 4 Communication state diagram. 1. Emergency message received 2. Disable message received 3. Reset message received 4. Return after reset 5. Invalid message received 6. Return from retransmission request 7. One minute period reached 8. Return from ping transmission 9. Start message received 10. Return after status message transmission 11. Threshold breach detected 12. Return from sensor data transmission

- A status request message received: A status message is framed and sent to the server with time stamp.
- A delay of one minute after the last message transmission: A ping message is sent to server to indicate that the device is still active.
- A threshold breach detected: A message is sent to the server with all sensor data and time stamp.
- A disable device message is received: The device is changed to STANDBY state.

Data related to all the aforementioned metrics are sent to the server when a threshold is breached, using the GSM packet formats described above. Now with all the data at hand, a random forest classifier is trained to classify driver proficiency. The features used in APAA are as follows:

x_1 Average frequency of close contact (number of breaches per hour)
x_2 Mean duration of close contact (in seconds)
x_3 Mean heart pulse (in beats per minute)
x_4 Number of threshold breaches
x_5 Mode accelerator pedal angle (in degrees)
x_6 Mean accelerator pedal angle (in degrees)
x_7 Mode brake pedal angle (in degrees)
x_8 Mean brake pedal angle (in degrees)
x_9 Average speed (km/h)

Random forest is an ensemble learning method used for classification, regression. It operates by constructing a multitude of decision trees during training, and outputs the class (classification) or mean prediction (regression) after training. Each decision tree is constructed by a random subset of the training data. After the forest has been trained, each row of real time data can be passed through the classifier to generate a prediction. Such a random forest comprised of binary trees, gives us the basic statistical model that is responsible for visualization and ranking. The system is initially trained with the help of a test set, for which, the ground truth is provided. This data is used to build the model for real time classification.

The APAA system uses nine features for classification, namely, the ones mentioned above. Each of these features is calculated from the data, which is received from the vehicle during a test run. The result from one driving test is called a Labelled Point. Apache Spark is used for back-end cluster computing. Thus, a random forest classifier is trained with instances shown as follows:

$$\begin{bmatrix} fa1 & fb1 & fc1 & \cdots & C1 \\ & \vdots & & \cdots & \vdots \\ fan & fbn & fcn & \cdots & Cn \end{bmatrix}$$

This matrix is called the metric set S. In AAPA system, S is in a nine tuple format,

$$S = \langle x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9 \rangle$$

The data stored in matrix S are collected from all the sensors. Once this data has been stored in the database, the random forest algorithm is applied to break this data set down into random subsets, and then create some de-correlated decision trees for each subset. After the trees have been created, the classification is performed based on various class modules. Based on these ratings, the modules are classified as proficient and non-proficient, as shown in Fig. 5. Once the ensemble is processed, each decision tree gets classified according to these classes. And, the class with the maximum number of occurrences is chosen as the appropriate rating Eq. This entire back-end part of the system is accessible through a web interface, which fetches data from an SQL relational database for whatever record the user desires. The interface is unique and contains secured access to records for each individual or organization.

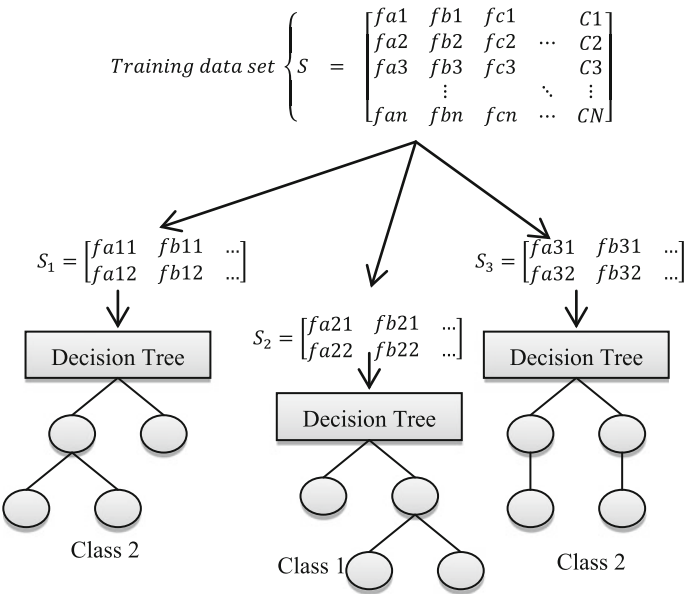


Fig. 5 General random forest classification method. As class 2 is frequent, the result is class 2

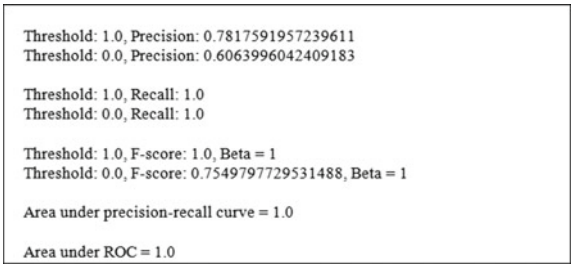


Fig. 6 Classification Results

5 Results

On testing with synthetic data for random forest classification, and splitting the data set as 70% for training and 30% for evaluation, the results obtained are presented in Fig. 6. The model trained in this test case reported promising results for the synthetic data. The random forest ensemble worked best in this particular test case, and achieved a proficiency score of 78%. With sufficient real time data, this score can be improved greatly. Contrariwise, testing with real data requires more preparation and training.

6 Conclusion

Safety is a major concern in any transportation. The majority of the accidents are caused by the road drivers. In the current scenario, there is no mechanism to accurately assess the skill level of an applicant, especially in developing countries. Hence, a novice method to analyze the performance of drivers on the road is presented in this article. In this method, each driver behavior is analyzed by an expert system. With more and more practice data, the system can provide more accurate result. High accuracy in terms of proficiency and rank order is achieved, which can be used as a primary factor in hiring a driver or in issuing a license.

References

1. Opila, D.F., Aswani, D., McGee, R., Cook, J.A., Grizzle, J.W.: Incorporating drivability metrics into optimal energy management strategies for hybrid vehicles. In: 47th IEEE Conference on decision and Control, December 9, pp. 4382–4389 (2008)
2. Liu, Y., Wu, Z.: Multitasking driver cognitive behavior modeling, In: 3rd IEEE International Conference on Intelligent Systems, pp. 52–57 (2006)
3. Remboski, D., Douros, K., Lee, J., Gardner, J.L., Gardner, R.M., Hurwitz, J.B., Leivian, R.H., Nagel, J., Wheatley, D.J., Wood, C.A.: Method and apparatus for vehicle operator performance assessment and improvement, United States patents US 6,925,425. August 2 (2005)

4. Harkness, R.: Advanced Drivers Education Products, Training, Driver training system, United States patent US 6,227,862. May 8 (2001)
5. Burch, L.A.: Driving simulator and method of evaluation of driver competency, United States patent application US 11/903,152. September 20 (2007)
6. Kumar, A., Mudhole, S.S., Lemoine, B.: A smart sensor-based software system for driver evaluation, In 4th Annual Systems Conference, IEEE, April, pp. 472–477 (2010)
7. Clement, F.S.C., Vashistha, A., Rane, M.E.: Driver fatigue detection system, In International Conference on Information Processing (ICIP), December, pp. 229–234 (2015)
8. Bezet, O., Cherfaoui, V., Bonnifait, P.: A system for driver behavioral indicators processing and archiving. In: Intelligent Transportation Systems Conference, ITSC'06. IEEE September 17, pp. 799–804 (2006)
9. Mandal, B., Li, L., Wang, G.S., Lin, J.: Towards detection of bus driver fatigue based on robust visual analysis of eye state. IEEE Trans. Intell. Transp. Syst. (2017)
10. Tang, X., Zhou, P., Wang, P.: Real-time image-based driver fatigue detection and monitoring system for monitoring driver vigilance. In: 35th Chinese Control Conference (CCC), July, pp. 4188–4193 (2016)
11. Wongphanngam, J., Pumrin, S.: Fatigue warning system for driver nodding off using depth image from Kinect. In: 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), June, pp. 1–6 (2016)
12. He, Q., Li, W., Fan, X., Fei, Z.: Evaluation of driver fatigue with multi-indicators based on Artificial Neural Network. IET Intell. Transp. Syst. (2016)
13. Toma, M.I., Rothkrantz, L.J., Antonya, C.: Car driver skills assessment based on driving postures recognition. In: 3rd International Conference on Cognitive Infocommunications (CogInfoCom), IEEE, December, pp. 439–446 (2012)
14. Osgouei, R.H., Choi, S.: Evaluation of driving skills using an HMM-based distance measure. In: International Workshop on Haptic Audio Visual Environments and Games (HAVE), IEEE, October, pp. 50–55 (2012)

Abnormal User Pattern Detection Using Semi-structured Server Log File Analysis

P. V. Sai Charan

Abstract An intrusion can be defined as a group of actions or events that try to compromise the confidentiality, integrity, and availability of a computer system. An intrusion detection system records information about certain events and produces reports for the administrators in the real time by analyzing the data obtained from the events. The objective of this paper is to find abnormal activity patterns of users from a huge amount of semi-structured server log file. The system analyzes the log data by using an open-source framework named Hadoop. At the end, results will be visualized using RStudio. The output plots will help in differentiating between the normal users and the intruders in a particular network. After getting the intruders' data, the network administrators can observe and react accordingly to minimize the further loss in that network.

1 Introduction

In today's digital world, intrusion detection is one of the most important areas of concern for the IT experts as well as research scholars [1]. An intrusion detection system (IDS) records information about certain events and produces reports for the administrators in the real time by analyzing the data obtained from the events. Intrusion detection systems can be broadly classified into two categories.

1.1 Signature-Based Intrusion Detection System

Signature-based intrusion detection system will detect intrusions by scanning the entire network for known malicious patterns [2]. In this technique, the network

P. V. Sai Charan (✉)
TIFAC-CORE in Cyber Security, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham Amrita University, Coimbatore, India
e-mail: pvsaicharan2011@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_10

administrators will search for known malicious patterns. The key advantage with signature-based IDS is that it can be implemented easily but on the downside signature-based IDS cannot detect novel attacks and sometimes they give false positives as they are mostly dependent on regular expressions and pattern matching techniques.

1.2 Anomaly-Based Intrusion Detection System

Anomaly-based intrusion detection systems will search for any abnormal activity patterns in the network by comparing that with normal usage patterns [1, 2]. The main advantage of this system is that it can find new attacks unlike signature-based intrusion detection system. But the main drawback of this technique is that we need to analyze more data and require a lot of knowledge for detecting new threats.

Although both the systems have their own advantages and disadvantages, anomaly-based intrusion detection system has the upper hand over signature-based intrusion detection because anomaly-based system is dynamic and agile in nature. As a result, it can cope up with the modern threats which are of higher priority. Signature-based detection may not give fruitful results for huge corporate networks but on the other hand anomaly-based intrusion detection systems are capable of providing good security to their networks with proper design and implementation.

With the advent of sophisticated tools, cyber attacks and malicious activities are becoming a matter of great concern in the corporate world [3]. Sometimes, the attacker may be a insider, and in this case the time spent on analyzing the incoming traffic is of no use and a complete waste of system resources. Especially, with advance persistency threat (APT) it is very difficult to identify the presence of intruders in corporate networks [4]. One of the best ways to deal with these kinds of attacks is to analyze the server logs in order to trace back to the intruder in the corporate network. But the real problem is that the server log files will be huge in size and mostly in semi-structured or unstructured format [5]; i.e., the schema cannot be actually specified for that particular log file in order to store them in relational database systems for further analysis. The main objective of this paper is to find the abnormal behavior pattern of the users in a particular network by analyzing a huge amount of semi-structured server log file.

This paper is organized as follows. The first section details intrusion overview and types of intrusion detection techniques, followed by the challenges that corporate network face with IDS and how it is tackled. Related works are specified under Sect. 2. Proposed system, general structure of input log file, and Hadoop work MapReduce workflow along with partitioner working are discussed under Sect. 3. Implementation and results are explained in Sect. 4. Conclusion and future work is discussed under Sect. 5.

2 Related Works

Recently, research work on various intrusion detection methods has been statistically increasing. Many hybrid techniques have been developed by integrating both signature and anomaly detection techniques at the core level. Jin proposed virtual machine monitor (VMM) to observe and analyze the cloud data patterns to prevent from novel type of DDoS attacks which deal with a lot of complex data patterns [6]. In addition, Zahra Salek proposed another work to secure the cloud and corporate networks using multilevel IDS which divides the risk level of the users based on the type of service that they are using [7]. S. Anuraj proposed hardware-level intrusion for high-speed interfaces using field-programmable gate array (FPGA) [8]. A. E. Jose proposed gigabit network intrusion detection system using extended broom filters [9]. V. Das proposed a probabilistic approach for network-based intrusion detection system using support vector machine (SVM) [10].

Similarly, there are plenty of works going on in the software industry on Web log mining in order to find out the user behavior and from which location of the world these requests are coming to that server. By using these MapReduce techniques, they are trying to minimize time complexity for the variety of data which are in huge volumes. Abzetdin proposed an example for preprocessing HTTP server logs by using R-environment to filter out the records generated by Web crawlers [11]. SayaJee Narkhede proved experimentally that parallel processing in Hadoop will reduce run time for processing huge log files when compared with traditional database systems [12]. If this proposed system is added to the above-mentioned intrusion detection systems, it will produce a much stronger IDS for large-scale corporate networks.

3 Proposed System

Even though there are many traditional IDSs available in the industry, recent cyber attacks across the globe have left large-scale corporate networks in a bind as to know how to tackle these new attacks [13, 14]. Especially, when APT comes into the picture, the whole scenario changed [15]. Performing traffic analysis on incoming traffic does not work with these kinds of cases to spot intruders in a network. So, careful observations of user's patterns in the network are essential for any corporate network. In this paper, we have proposed a technique to analyze the abnormal behavior of the users by analyzing semi-structured server log files in that particular network. In order to achieve better performance and effective results, MapReduce mechanism is suitable to store and process this kind of semi-structured data as traditional database systems are unable to give fruitful results for these kinds of analyses [16].

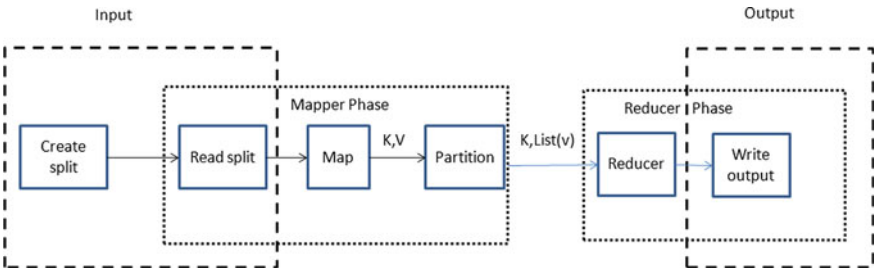


Fig. 1 MapReduce workflow in proposed system

3.1 General Structure of Server Log file

"96.7.4.14 - - [24/Apr/2011:04:20:11 -0400] "GET /cat.jpg HTTP/1.1" 200 12433" General structure of server log file consists of an IP address followed by a time stamp, the type of request initiated by the client and the resource path which the client had accessed, status code sent from the server and the last value indicates that how much data has been transferred to the client [17].

3.2 Hadoop MapReduce Workflow and Partitioner Role

Traditional Hadoop i.e Hadoop(1.x) in a single node cluster setup used for sample log file processing [18]. Generally, Mapper will take values from the Record Reader and generate key-value pairs according to the user's requirement [16]. The number of Mappers will be decided based on the number of input splits. Mapper output will be given to the partitioner if there is any such requirement. Partitioner channels Mapper output to the specified Reducer according to programmers specification [16]. At the Reducer end values will be aggregated and by using RecordWriter Reducer's output will be written to output file as illustrated in Fig. 1.

Partitioner is the key component in proposed system. Partitioner comes into the picture in the Mapper phase. It is used to generate a unique hash value for each month, and based on that hash values 12 reducers will aggregate count for every month as illustrated in Fig. 2.

3.3 Detecting Abnormal Patterns from Log files

Here, a semi-structured server log file of huge size will be processed by using an open-source framework called Hadoop. IP addresses and the month from the time stamp are extracted as a key-value pair <IP Address, Month>by using Mapper. Once

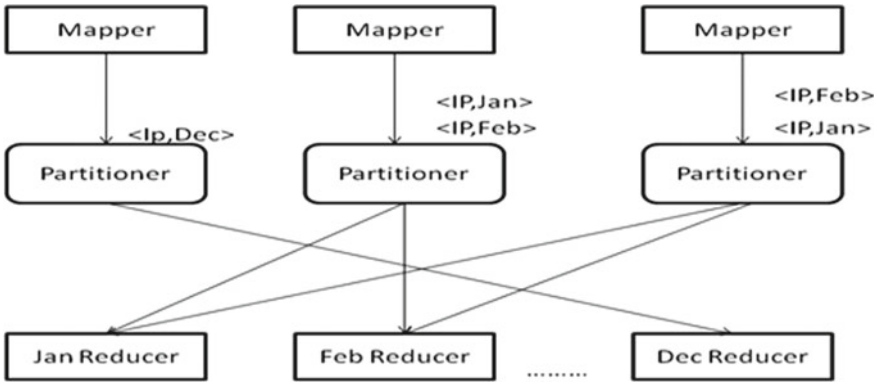


Fig. 2 Hash partitioner workflow

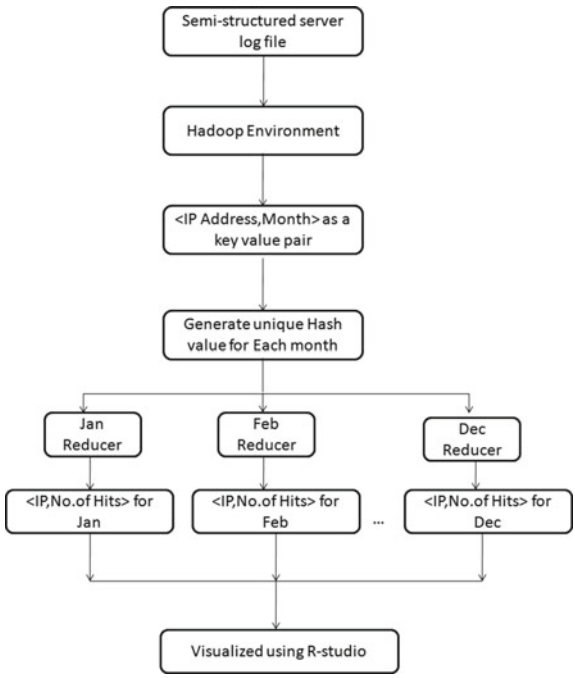
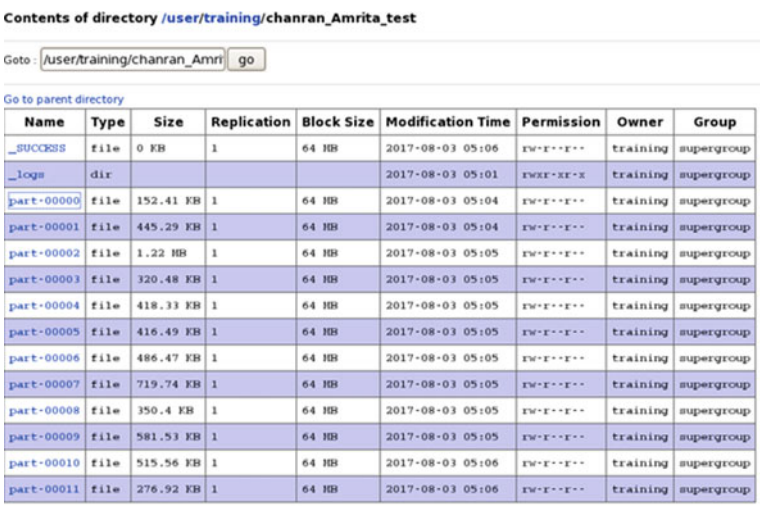


Fig. 3 Proposed system dataflow architecture

the Mapper output is acquired, the key-value pairs are then sent to the hash partitioner to generate a unique hash value for every month. Based on the hash value, we are trying to split all the key-value pairs from Mapper among 12 Reducers (1 for each month). At each Reducer, the count for the number of hits for each IP address is calculated. Finally, 12 outputs from 12 reducers which will represent each and every month in a year. The final output key-value pairs <IP Address, No.of Hits>will be visualized using RStudio as illustrated in Fig. 3.



Contents of directory /user/training/chanran_Amrita_test

Goto : /user/training/chanran_Amri go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
_SUCCESS	file	0 KB	1	64 MB	2017-08-03 05:06	rw-r--r--	training	supergroup
_logs	dir				2017-08-03 05:01	rwxr-xr-x	training	supergroup
part-00000	file	152.41 KB	1	64 MB	2017-08-03 05:04	rw-r--r--	training	supergroup
part-00001	file	445.29 KB	1	64 MB	2017-08-03 05:04	rw-r--r--	training	supergroup
part-00002	file	1.22 MB	1	64 MB	2017-08-03 05:05	rw-r--r--	training	supergroup
part-00003	file	320.48 KB	1	64 MB	2017-08-03 05:05	rw-r--r--	training	supergroup
part-00004	file	418.33 KB	1	64 MB	2017-08-03 05:05	rw-r--r--	training	supergroup
part-00005	file	416.49 KB	1	64 MB	2017-08-03 05:05	rw-r--r--	training	supergroup
part-00006	file	486.47 KB	1	64 MB	2017-08-03 05:05	rw-r--r--	training	supergroup
part-00007	file	719.74 KB	1	64 MB	2017-08-03 05:05	rw-r--r--	training	supergroup
part-00008	file	350.4 KB	1	64 MB	2017-08-03 05:05	rw-r--r--	training	supergroup
part-00009	file	581.53 KB	1	64 MB	2017-08-03 05:05	rw-r--r--	training	supergroup
part-00010	file	515.56 KB	1	64 MB	2017-08-03 05:06	rw-r--r--	training	supergroup
part-00011	file	276.92 KB	1	64 MB	2017-08-03 05:06	rw-r--r--	training	supergroup

Fig. 4 Output of MapReduce job

4 Implementation and Results

Nearly, 1 GB of sample server log file collected from local area network (LAN) environments was processed, which resembles the standard Apache common server log format [17]. Initially, that log files will be uploaded to Hadoop file system. Once the data is processed by MapReduce environment from Hadoop file system, it will produce 12 output files as a result each representing a particular month in a year as shown in Fig. 4.

Each one of the output files will have <IP Address, Count>as a key-value pair as shown in Fig. 5. RStudio is used for visualizing Reducers output data [19]. Based on the results, graphs are plotted for each and every month to exactly trace users with abnormal activity patterns. In the graph, the number of hits taken along X-axis and frequency of the users is taken along Y-axis (e.g., from Fig. 6., the number of users who visited server only one time is more than 15,000).

- Figure 6 illustrates the normal user patterns for the month of November.
- Figure 7 illustrates the abnormal user patterns for the month of November.

An "R" code scripted to identify these abnormal IPs from the results and saved that data in .CSV format for each and every month so that monitoring those IP will be easy for the network administrators. Usually, the lower and upper bounds for the hit count to classify as a normal and abnormal user will be decided by the network administrators. But in this case, we considered 0–1000 hits per month as normal user pattern and rest are considered as abnormal user patterns for sample log output visualization.

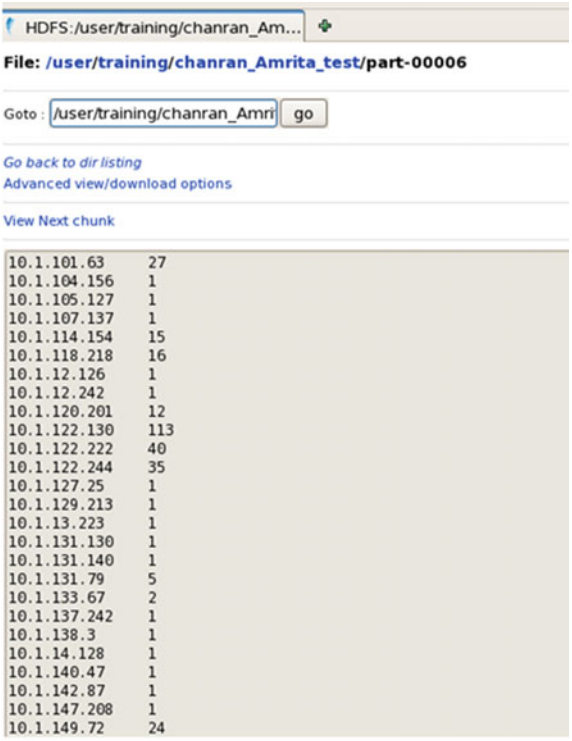


Fig. 5 Processed output for a particular month

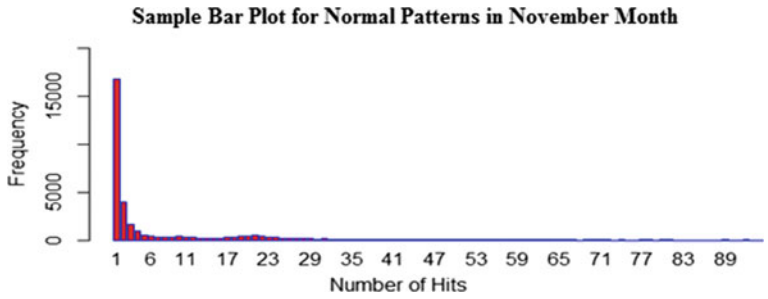


Fig. 6 Normal user pattern for November month

5 Conclusion and Future work

In this paper, we propose a technique to analyze abnormal activity patterns of the users in a network by analyzing the semi-structured server log files by using Hadoop MapReduce framework [16]. From the output plots, we can differentiate the normal users and intruders in that particular network. Once if we get the intruders data, then

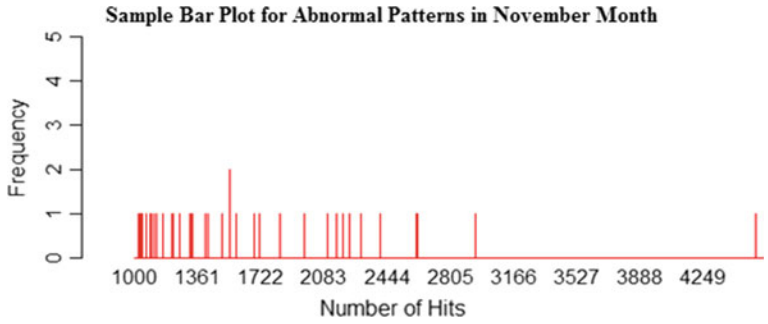


Fig. 7 Abnormal user pattern for November month

we can track them and take appropriate actions to minimize any further harm to the network. This data will be further used for identifying abnormal data usage and activity patterns in that network, and also we can easily track down to the actual vulnerability of those affected systems.

As an additional enhancement to this system, we can further analyze data download patterns of suspected users from server log files. This system will be implemented along with network traffic analysis in order to trace the systems which have been affected by the APT. Once the APT gains entry into our network, the malicious scripts will try to gain access to any sensitive information and they will try to push that data to external servers. So, this system can be extensively used for implementing the existing techniques like Honeypot and Honeynet in a particular corporate network to trap malicious users [20–23].

References

1. Naorem, S., Sharma, A.: An overview of intrsuion detection system. Int. J. Res. Appl. Sci. Eng. Tech. (IJRASET) **3**(VI), (2015)
2. Jyothsna, V., Prasad, V.V.R.: A review of anomaly based intrusion detection systems. Int. J. Comput. Appl. (0975–8887) **28**(7), 26–35 (2011)
3. Neville, S.W. : A research facility for evaluating cyber-security approaches within corporate-scale networks and under operational conditions. In: IEEE Pacific Rim Conference on Communications, Computers and signal Processing, pp. 466–469 (2005)
4. Messaoud, B.I.D.,Karim G., Wahbi, M., Sadik1, M.: Advanced Persistent Threat - new analysis driven by life cycle phases and their challenges. In: International Conference on Advanced Communication Systems and Information Security (ACOSIS), pp. 1–6 (2016)
5. Subramaniyaswamy, V., Vijayakumar, V., Logesh, R., Indragandhi, V.: Unstructured data analysis on big data using map reduce. In: 2nd International Symposium on Big Data and Cloud Computing (ISBCC–15), pp. 456–465 (2015)
6. Jin, H., Xiangn, G., Zou, D. et al.: A VMM-based intrusion prevention system in cloud computing environment. J. Supercomput. pp. 1–19 (2011)
7. Salek, Z., Madani, F.M.: Multi-level Intrusion detection system in cloud environment based on trust level. In: 6th International Conference on Computer and Knowledge Engineering (ICCKE 2016), pp. 94–99 (2016)

8. Anuraj, S., Premalatha, P., Gireesh Kumar, T.: High speed network intrusion detection system using FPGA. In: Proceedings of the Second International Conference on Computer and Communication Technologies: IC3T 2015, vol. 1, pp. 187–194 (2016)
9. Jose, A.E., Gireesh kumar, T.: Gigabit network intrusion detection system using extended bloom filter in reconfigurable hardware. In: Proceedings of the Second International Conference on Computer and Communication Technologies, pp. 11–19 (2016)
10. Das, V., Pathak, V., Sharma, S., Srikanth, M.V.V.N.S., Gireesh, K.T.: Network intrusion detection system based on machine learning algorithms. Int. J. Comput. Sci. Inf. Tech. (IJCSIT) **2**, (2010)
11. Abzetdin, A.: Data mining and analysis in depth. case study of Qafqaz university HTTP server log analysis. In: IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), pp. 1–4 (2014)
12. SayaJee, N., Baraskar, T., Mukhopadhyay, D.: Analyzing web application log files to find hit count through the utilization of hadoop map reduce in cloud computing environment. In: IT in Business, Industry and Government (CSIBIG), pp. 1–7 (2014)
13. Modern cyber attacks. https://www.wired.com/story/2017-biggest-hacks-so-far/
14. Mohurle, S., Patil, M.: A brief study of wannacry threat: ransomware attack. Int. J. **8**(5), (2017)
15. Vukalović, J., Delija, D.: Advanced persistent threats – detection and defense. In: 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1324–1330 (2015)
16. White, T.: Hadoop the definitive guide. O'Reilly media (2015)
17. Apache log file structure. https://httpd.apache.org/docs/1.3/logs.html
18. Hadoop single node cluster. https://www.dezyre.com/article/hadoop-2-0-yarn-framework-the-gateway-to-easier-programming-for-hadoop-users/84
19. R-studio cheatsheet for sample plot visualization. https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf
20. Honeypot. https://ru.wikipedia.org/wiki/Honeypot
21. Egupov, A.A., Zareshin, S.V., Yadikin, I.M., Silnov, D.S.: Development and implementation of a honeypot trap. In: IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), pp. 382–385 (2017)
22. Albashir, A.A.A.N.: Detecting unknown vulnerabilities using honeynet. In: First International Conference on Anti-Cybercrime (ICACC), pp. 1–4 (2015)
23. Aathira, K.S., Nath, H.V., Kutty, T.N., Gireesh, K.T.: Low budget honeynet creation and implementation for Nids and Nips. Int. J. Comput. Netw. Secur. (IJCNS) **2**(8), 27–32 (2010)

Computational Modeling and Parametric Analysis of an Implantable Patch Antenna Using Finite-Difference Time-Domain Algorithm

T. Mary Neebha, M. Nesasudha and Evangelin Chrysolite

Abstract A bio-compatible implantable microstrip patch antenna is designed and analyzed for Implant Communications Services using the finite-difference time-domain (FDTD) method. The space in the patch area of the basic spiral antenna is efficiently used and hence modified. Here, the modified antenna represents a simple structure and is found to have higher gain, lower return loss, lower specific absorption rate (SAR), and a high radiation efficiency of about 73%. The performance characteristics of the antenna have also been evaluated using FEKO software, and the results are compared.

1 Introduction

In the early 1950s, the simulation of electrical pulses aided the heart patients. In 1952, Dr. Paul described the first external pacemaker. With the advent of transistors, came the idea of implantable devices, and in 1958, the first implantable pacemaker was placed in an animal [1]. In 1960, the first successful pacemaker was implanted in humans. Nowadays, the diagnostic procedures aim at reducing the pain or discomfort of the patients. Designing implanted antennas is extremely challenging because of less radiation efficiency, radiation effects of antenna with human tissue, importance of miniaturization, the effect of propagation losses, impedance matching, high tissue conductivity, low-power requirements, and the detuning effects of human body tissues. Usually, implants sense the information from the human body and send the signal to the receiver placed outside the body.

T. Mary Neebha (✉) · M. Nesasudha · E. Chrysolite
Department of Electrical Technology, Karunya Institute of Technology & Sciences, Coimbatore 641114, Tamil Nadu, India
e-mail: maryneebha@karunya.edu

M. Nesasudha
e-mail: nesasudha@karunya.edu

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_11

The United States Federal Communications Commission (FCC) recommended the Medical Implant Communication Service (MICS) band for low power communication with medical implants. A good compromise can then be added between size reduction and power consumption [2].

Quite a lot of antennas have been used for a variety of wireless implantable applications [3]. In this paper, microstrip antenna, which is a low-profile antenna, is used because of its high flexibility in design, direct integration, conformability, and shape. It is less expensive and is easy to fabricate using printed circuit technology [4]. Because of these advantages, they are more involved in modern communication systems.

Implantable antennas are used for cardiac pacemaker [3], defibrillator [1], continuous glucose monitoring [5], satellite and radio telecommunication, wireless biomedical devices [6], modern wireless communication, wireless capsule endoscopy [7], and so on.

To reduce the patch size, a shorting stub can be placed from the radiating section to the ground plane, or a superstrate of high relative permittivity can be added. The former produces the same gain at less than half the size of a similar antenna without the stub, while the latter improves the radiation efficiency of the antenna [5]. The use of superstrate in the design reduces the center frequency and enhances the impedance bandwidth of the antenna. Rogers RO3210/RO3010/6002 is used extensively utilized for implantable antennas, since they have a high permittivity of 10.2. Rogers 3210 is used as a substrate and superstrate in this paper.

A number of parameters such as efficiency and specific absorption rate (SAR) evaluate the antenna performance. Gain (dB) measures the total radiated power from the radiator in a particular direction. Return loss (S11) analyzes the reflected power. It gives a useful measure of electrical power transfer efficiency. It must be less than 10 dB. The energy absorbed by the body when exposed to an electromagnetic field is defined as specific absorption rate (SAR) [8], and its unit is Watts per kilogram (W/kg).

The modeling based on computation has become a vital part of numerous research efforts in key application areas such as chemical, physical, and biological sciences [9]. Computational modeling requires the availability of codes that offer a wide range of capabilities because the information gained is unique. To keep the computational accuracy, computational modeling is used. Finite difference time domain (FDTD) is an electromagnetic computational technique, which is easy to implement the antenna design. In this work, FDTD is used for simulation through MATLAB, and the antenna was designed and simulated with the aid of Feldberechnung bei Körpern mit beliebiger Oberfläche (FEKO).

In this work, a basic spiral microstrip patch antenna design is done and validated for its performance. Then, the planar radiator of the antenna is modified, and the antenna performance was found to be improved. Using finite-difference time-domain (FDTD) computational method, the performance of the modified antenna is verified. The electric and magnetic fields were animated, and the SAR value is tested for its accuracy.

This manuscript is organized as follows. Section 2 provides the simulated FEKO design for the basic spiral microstrip patch antenna and modified spiral antenna (MSA). Section 3 presents the numerical analysis of the MSA, which is done by FDTD through MATLAB. The simulated results from both the sections are discussed in Sect. 4. Finally, Sect. 5 presents discussions and a conclusion to this work.

2 Antenna Design

2.1 *Basic Spiral Microstrip Patch Antenna*

A basic spiral microstrip patch antenna was designed [10] for an input impedance of 50 Ω as shown in Fig. 1.

The antenna was designed for a frequency range of 402–405 MHz. The dimension of the antenna is reduced by using a substrate and superstrate of high permittivity material, Rogers 3210, which has a relative permittivity of 10.2. Coaxial feed is used for this antenna. The patch is located between the substrate and superstrate layers. The patch consists of a uniform width of 4 mm. The antenna is positioned between skin tissues of relative permittivity, 46.7, and conductivity $\sigma = 0.69$ S/m. The size of the antenna is 40 mm × 36 mm × 8 mm. Figure 1a displays the geometry and the specifications for the antenna, and Fig. 1b gives the side view of the antenna.

2.2 *Modified Spiral Antenna (MSA)*

The patch is extended to a length of about 32 mm and then 4 mm from the middle of the basic microstrip patch antenna. This is done for the efficient use of the space

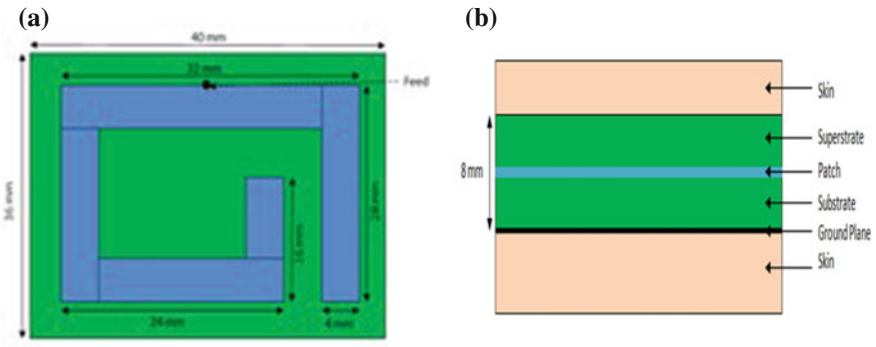


Fig. 1 Basic spiral microstrip patch antenna. **a** Patch of the antenna. **b** Side view

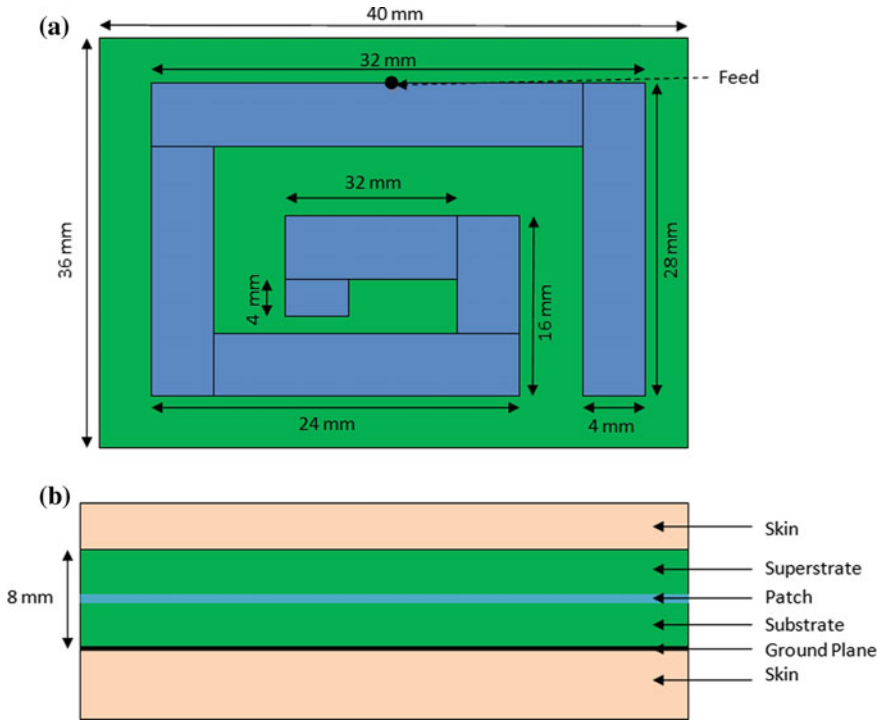


Fig. 2 Modified spiral microstrip patch antenna. **a** Patch of the antenna. **b** Side view

of the antenna, in the middle between the substrate and the superstrate of the antenna, i.e., in the patch area. Figure 2 shows the design of the modified spiral microstrip patch antenna [11]. The antenna is designed for an input impedance of 50 Ω, and coaxial type of feed is used. This antenna is also positioned between the skin tissues, and the results of the modified antenna were found to have higher efficiency compared to that of the basic antenna. For accuracy, the antenna is analyzed numerically by using FDTD, a computational technique, and the results are compared which are explained in Sect. 5.

3 Numerical Analysis

3.1 FDTD Method

FDTD is an effective numerical method for calculating electromagnetic effects when human tissues are present beside electronic devices [12]. It has become the basic method to computationally design many engineering issues. FDTD has the advantage of simplicity of implementation and analysis. As FDTD calculates

the electric and magnetic field components everywhere in the computational domain, it can provide accurate predictions through animated displays of the electromagnetic field movement throughout the antenna.

The original concept of FDTD was described by Yee cell [13]. Figure 3 shows the Yee cube. In order to apply the FDTD method, the antenna is split into Yee cells where each cell unit contains Ex, Ey, Ez, Hx, Hy, and Hz components. Maxwell's equations are converted to point form using the finite-difference approximation that is computed in the time domain. In an isotropic medium, Maxwell's equations are as follows:

$$\frac{\partial D}{\partial t} = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \cdot \nabla \times \mathrm{H} \tag{1}$$

$$D = \varepsilon_r \cdot E \tag{2}$$

$$\frac{\partial H}{\partial t} = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \cdot \nabla \times \mathrm{E} \tag{3}$$

where D is the electric flux density, H is the magnetic field, E is the electric field, ε_0 is the permittivity in free space, ε_r is the relative permittivity, and μ_0 is the permeability in free space.

Equations (1) and (3) produce six scalar equations from which finite-difference approximations are taken. The accuracy of FDTD depends on the cell size. Therefore, the number of spatial cells increases. A cell size of $\lambda/20$ is set for accurate calculations. The timestep calculation [12] is done by

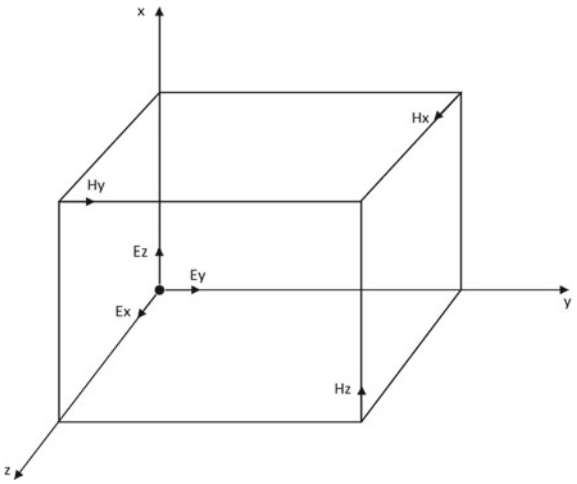


Fig. 3 Yee cube

$$\nabla t \leq \frac{1}{c}\sqrt{\left(\frac{1}{\Delta x^2}+\frac{1}{\Delta y^2}+\frac{1}{\Delta z^2}\right)} \tag{4}$$

where c is the velocity of light in the computational volume of the antenna, and Δx, Δy, Δz are the respective cell sizes.

The electric and magnetic field components are computed and are animated using the FDTD technique. The Ez radiations at timesteps 10, 15, and 20 are shown in Figs. 4, 5 and 6, respectively.

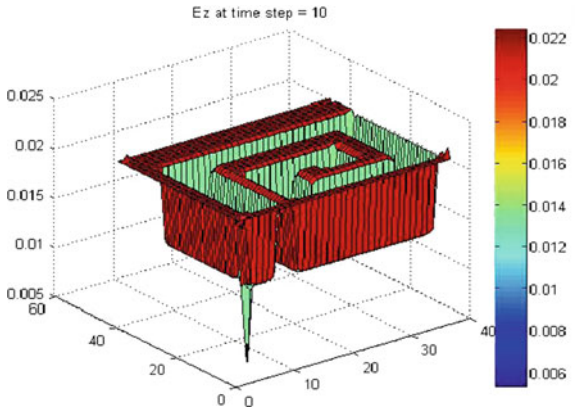


Fig. 4 Ez at timestep = 10

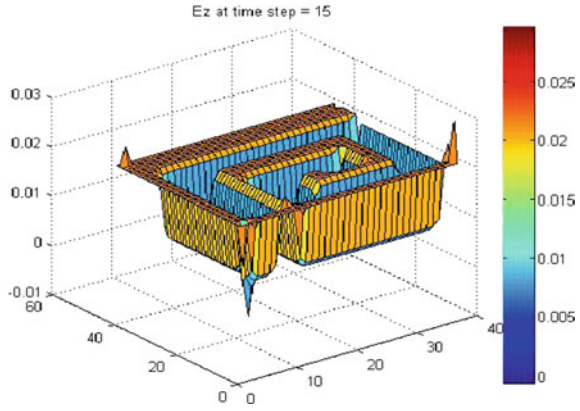


Fig. 5 Ez at timestep = 15

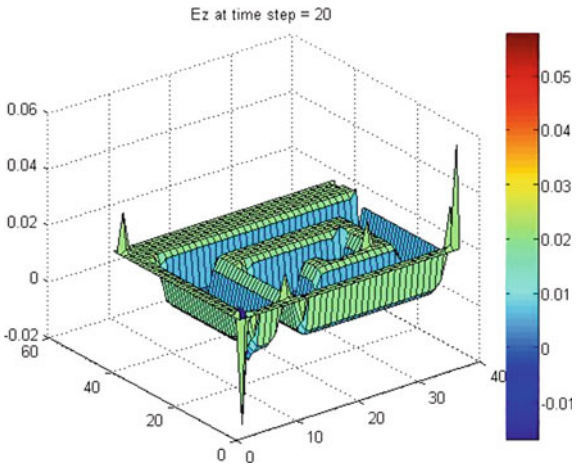


Fig. 6 Ez at timestep = 20

3.2 FDTD Computation of SAR

Implantable devices have power limitations to prevent hazardous heating of biological tissues. The maximum power for any implantable device must comply with the peak spatial-average SAR limitations, i.e., 2.0 W/Kg. SAR is introduced as an evaluation index for the electromagnetic waves in biological tissues. The SAR calculation is estimated using the formula [4] given below

$$SAR(i,j,k) = \frac{\sigma}{2\rho}\left|\vec{E}\right|^2 = \frac{\sigma(i,j,k)\left\{\left|\widehat{E}_x(i,j,k)\right|^2 + \left|\widehat{E}_y(i,j,k)\right|^2 + \left|\widehat{E}_z(i,j,k)\right|^2\right\}}{2\rho(i,j,k)} \tag{5}$$

where $\widehat{E}_x$, $\widehat{E}_y$, and $\widehat{E}_z$ are the peak values of the electric field components; σ (S m^{-1}) is the conductivity of the material used; ρ (kg m^{-3}) is the mass density of the skin.

3.3 FDTD Computation of Return Loss

Return loss (s11) denotes the power reflected from the antenna. Return loss should be less than 10 dB. S11 is computed by finding the ratio of the discrete Fourier transform (DFT) of incident and reflected electric fields [14].

$$S11(f) = \frac{DFT[E_{ref}]}{DFT[E_{inc}]}$$

where E_{ref} is the reflected electric field and E_{inc} is the incident electric field. Return loss is computed in dB by

$$S_{11} = 20\log_{10}(|S_{11}|)$$

The frequency range for computation is in MICS band. The center frequency is the one in which the return loss is the minimum.

4 Results

4.1 Basic Spiral Microstrip Patch Antenna

The basic spiral antenna is simulated inside 16 mm × inf mm × inf mm block of skin. The SAR value obtained is 0.5 W/Kg as shown in Fig. 7. Figure 8 illustrates the return loss of the antenna. As shown, the minimum return loss obtained is −11.8 dB. Figure 9 proves that the gain is 0.70 dB. The radiation efficiency obtained is 27%, which is given in Fig. 10.

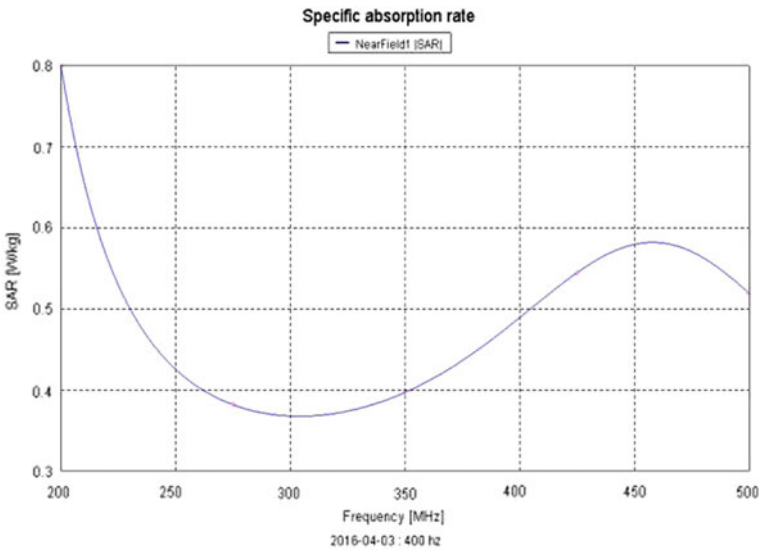


Fig. 7 SAR of the antenna

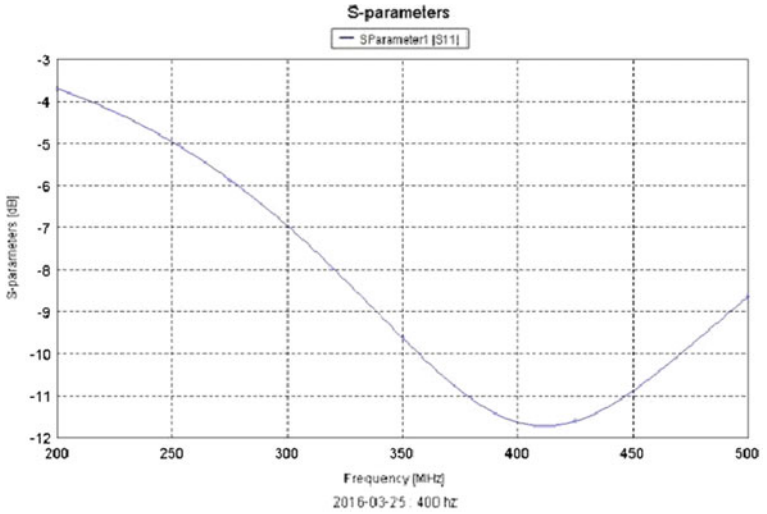


Fig. 8 Return loss

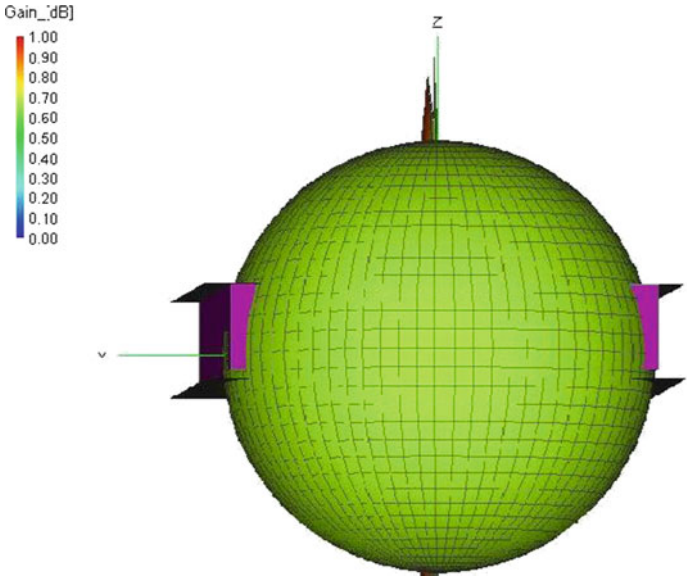


Fig. 9 Gain

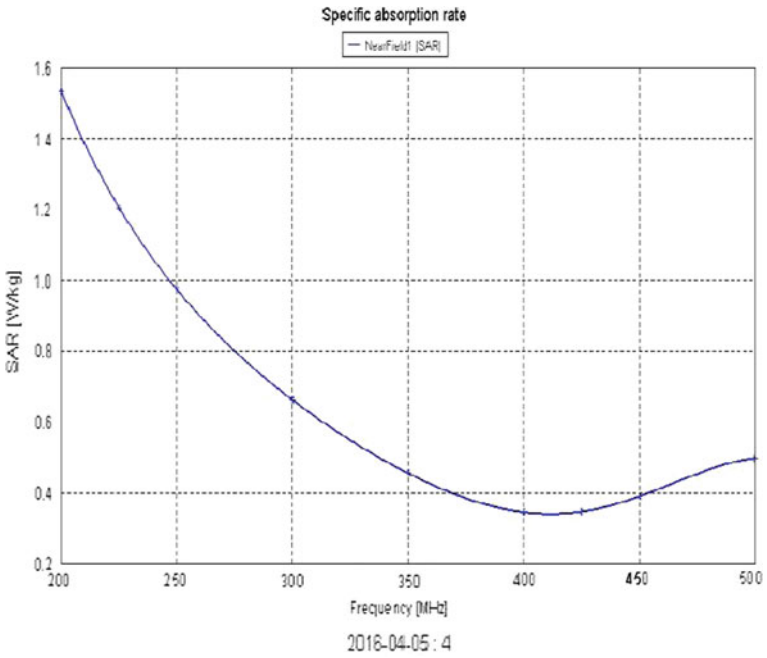


Fig. 10 SAR of the antenna

4.2 Modified Spiral Antenna (MSA)

The MSA is simulated inside 16 mm × inf mm × inf mm block of skin. The SAR value obtained is 0.3 W/Kg which is given in Fig. 10. Figure 11 gives the return loss value of the antenna. As shown, the minimum return loss obtained is −33 dB. Figure 12 shows the gain which is 6.3 dB. The radiation efficiency of the antenna is given in Fig. 13. The radiation efficiency obtained is 73%.

4.3 Computational Modeling

To verify the results obtained from FEKO, we computed SAR and return loss (S11) under the same conditions in MATLAB through FDTD. The value of SAR at resonance frequency is 2.66 W/Kg. The result generated by the FDTD code for the SAR is given in Fig. 14. The return loss obtained from MATLAB through FDTD is represented graphically in Fig. 15. The return loss obtained is −33.5 dB.

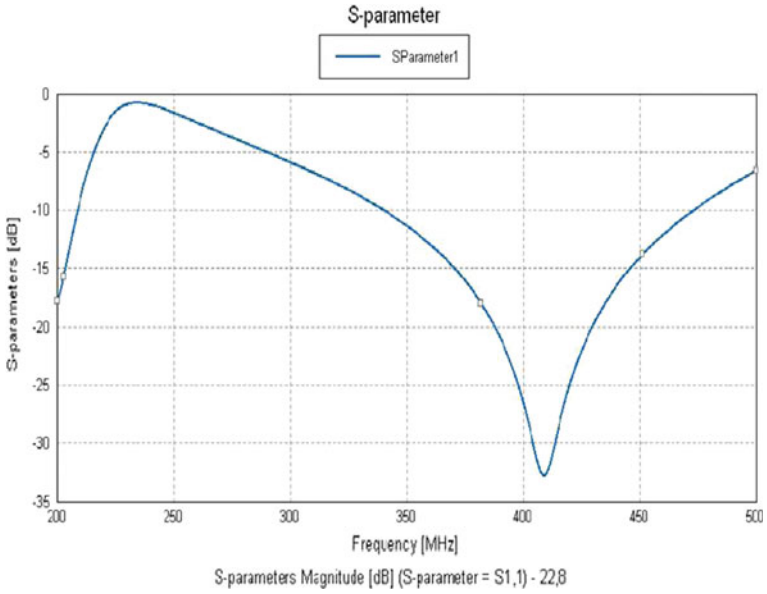


Fig. 11 Return loss versus frequency for the antenna

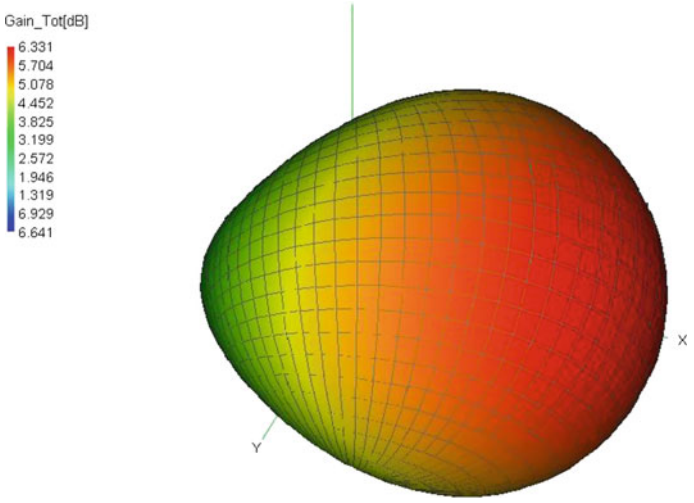


Fig. 12 Gain of the antenna

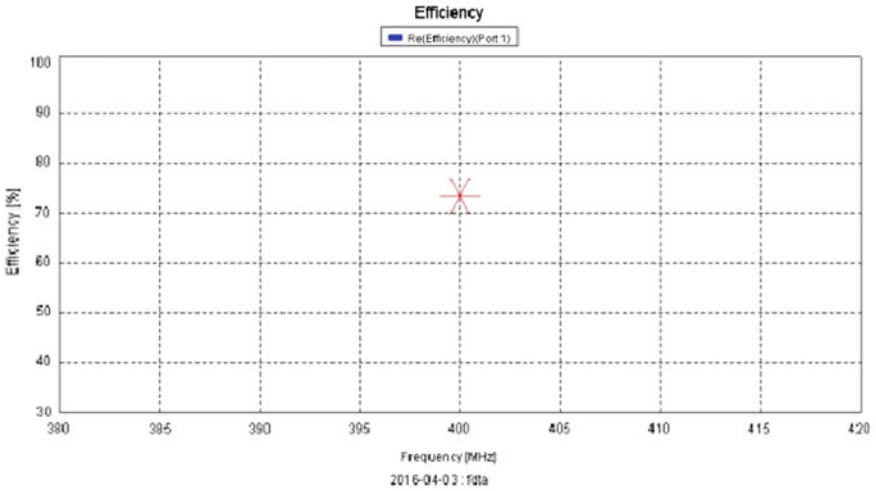


Fig. 13 Radiation efficiency of the antenna

Fig. 14 SAR of the MSA
Return loss of the MSA

```
        50

The value of SAR is

ans =

   2.6663e-004

fx >>
```

Fig. 15 Return loss of the MSA

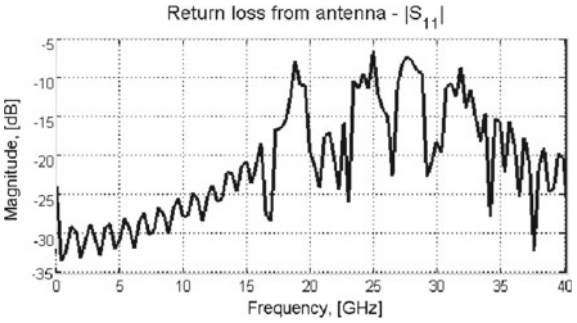


Table 1 Simulated results of the antenna

Antenna	Gain (dB)	Radiation efficiency (%)	Return loss (dB)	Specific Absorption Rate (SAR) (W/Kg)
Basic spiral microstrip patch antenna	0.7	27	-11	0.5
Modified spiral microstrip patch antenna (MSA)	6.3	73	FEKO = -33 FDTD = -33.5	FEKO = 0.3 FDTD = 0.26

4.4 Comparison

Comparisons are made to set a benchmark to the proposed antenna. A comparison in terms of gain in dB, return loss in dB, antenna radiation efficiency, and specific absorption rate in W/kg is summarized in Table 1.

SAR obtained from electromagnetic simulator is 0.3 W/Kg, whereas the value of SAR obtained from numerical FDTD method is 0.26. The magnitude of S11 is −33 dB at 402 MHz computed by the simulation software, while it is −33.5 dB at 402 MHz computed using FDTD. The computed bandwidth in both the cases is almost identical. The results obtained in MATLAB seem to be in close agreement with the simulated results. Thus, the results obtained for the modified antenna is found to have lower specific absorption rate (SAR), lower return loss, higher gain, and higher efficiency.

5 Conclusion and Future Work

A basic bio-compatible implantable microstrip patch antenna was presented to resonate at a frequency in the MICS band. The antenna is modified by efficiently utilizing the space of the planar patches. For accuracy and reliability, the modified antenna was computed numerically. A MATLAB code was developed, and the FDTD spatio-temporal difference equations were simulated. The results were found to be the same as that of the simulated results from FEKO. Our future work involves designing smart antennas for healthcare applications.

References

1. Greatbatch, W., Holmes, C.F.: History of implantable devices. IEEE Eng. Med. Biol. **10**(3), 38–49 (1991)
2. Ibraheem, A., Manteghi, M.: Performance of an implanted electrically coupled loop antenna inside human body. Prog. Electromagn. Res. **145**, 195–202 (2014)

3. Soontornpipit, P., Furse, C.M., Chung, Y.C.: Design of implantable microstrip antenna for communication with medical implants. IEEE Trans. Microw. Theory Tech. **52**(8), 1944–1954 (2004)
4. Percaz, J.M., Chudzik, M., Arnedo, I., Arregui, I., Teberio, F., Laso, M.A.G., Lopetegi, T.: Producing and exploiting simultaneously the forward and backward coupling in EBG-assisted microstrip coupled lines. IEEE Antennas Wirel. Propag. Lett. **15**, 873–876 (2016)
5. Karacolak, T., Hood, A.Z., Topsakal, E.: Design of a dual-band implantable antenna and development of skin mimicking gels for continuous glucose monitoring. IEEE Trans. Microw. Theory Tech. **56**(4), 1001–1008 (2008)
6. Liu, C., Guo, Y.-X., Xiao, S.: A Review of implantable antennas for wireless biomedical devices. In: Forum for Electromagnetic Research Methods Application Technologies (FERMAT)
7. Mary Neebha, T., Nesasudha, M.: Design and analysis of advanced microstrip patch antenna for endoscopic capsules. Wiley Period. Microw. Opt. Technol. Lett. **58**(7), 1762–1767 (2016)
8. Iqbal Faruque1, M.R., Misran, N., Tariqul Islam, M., Yatim, B., Bias, B.: New low specific absorption rate (SAR) antenna design for mobile handset. Int. J. Phys. Scie. **6**(24), 5706–5715 (2011)
9. Valiev, M., et al.: NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. Comput. Phys. Commun. **181**, 1477–1489 (2010)
10. Kim, J., Rahmat-Samii, Y.: Implanted antennas inside a human body: simulations, designs, and characterizations. IEEE Trans. Microw. Theory Tech. **52**(8), 1934–1943 (2004)
11. Evangeline Chrysolite, R., Divya Priya, A., Sushmitha, R.G., Manvitha, R.J., Mary Neebha, T., Nesasudha, M.: Design and analysis of bio-compatible implantable patch antenna. In: National Conference on Signal Processing Information and Communication Engineering (2016)
12. Paker, S., Sevgi, L.: FDTD evaluation of the SAR distribution in a human head near a mobile cellular phone. Turk. J. Electr. Eng. Comput. Sci. (Elektrik) **6**(3), 345–350 1998
13. Yee, K.S.: Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. IEEE Trans. Antennas Propag. **Ap-14**(3), 302–307 (1966)
14. Ali, M.F., Ray, S.: FDTD based SAR analysis in human head using irregular volume averaging techniques of different resolutions at 900 band. Indian J. Radio Space Phys. **43**, 235–242 (2014)

A Survey on Acronym–Expansion Mining Approaches from Text and Web

R. Menaha and VE. Jayanthi

Abstract An acronym is a textual form used to refer an entity and to stress the important concepts. Over the last two decades, many researchers worked for mining acronym expansion pairs from plain text and Web. This is mainly used in language processing, information retrieval, Web search, ontology mapping, question answering, SMS, and social media posting. Acronyms are dynamically growing day by day, and discovering its definition/expansion is becoming a challenging task because of its diversified characteristics. Manually edited online repositories have acronym definition pairs, but it is an overwhelming task to update all possible definitions systematically. To extend the support, different approaches are employed for the automatic detection of acronym definitions from text and Web documents. This paper presents those approaches and also reveals the Web-based methods used for disambiguating, ranking, finding popularity score, and context words of the expansions. The scope for the future work in this research area is also conferred in this paper.

1 Introduction

An acronym is a kind of abbreviation composed from first letter or first few letters of the words in a phrase. It is also called as short descriptors of phrase. It is quietly added as a new linguistic feature in English language during the year 1950s. Since acronym is referred as a word, its definition is called as its expansion. According to its formation, it is characterized into three types—(i) *character-based acronyms*: Generally, it is formed from the initial letter(s) of each word in the phrase

R. Menaha (✉)
Department of Information Technology, Dr. Mahalingam College of Engineering & Technology, Pollachi, Tamil Nadu, India
e-mail: rmenahasenthil@gmail.com

VE. Jayanthi
Department of Electronics and Communication Engineering, PSNA College of Engineering & Technology, Dindigul, Tamil Nadu, India
e-mail: Jayanthi.ramu@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_12

[e.g., ISRO is an acronym for the phrase "Indian Space Research Organization"]. (ii) *Syllable-based acronyms*: Acronyms are formed based on the syllables in the word [e.g., Kg is an acronym for the phrase "Kilogram." (iii) *Combination of character and syllable:* [e.g., RADAR combines the above initial characters and syllable-based acronyms].

Acronyms are widely used in biomedical documents because the names of many diseases, terminologies and procedures can be easily represented by using acronyms. Recognizing the expansions/definitions coupled with an acronym is a significant task in natural language processing (NLP) and information retrieval process. Similarly, acronyms are used very commonly in Web searches; as an example, the user gives NBA as search query instead of giving its full form National Board of Accreditation to reduce the access time. In social media like Twitter, Facebook, the user gives their comments in the form of acronyms to minimize the typing work. Typically, the acronyms are used during online chatting by the users. The usage of acronyms is even more common in mobile devices because acronyms make the typing process easier in such devices and the information is also expressed in a concise way.

To support the extraction of Acronym-Expansion (AE) process, manually collected AE lists are compiled and many are available in Internet as online corpus/repository. For each acronym query, the number of expansions returned by those corpuses varies a lot. As an example for the acronym query CAS, the corpus [1] returns 284 different expansions which are higher than other repositories. For any kind of acronym query, the repositories [1–5] give more number of expansions than the remaining four corpuses [6–9]. However, these repositories are restricted in specific domains or organizations. And the maintenance of acronym with its list of all possible definitions is a big problem because of rapid growth of acronyms.

Due to the above-said limitations found in online repositories, an automatic detection of expansions related to an acronym from free text and Web has been evolved since from last two decades. This paper presents those approaches followed for the detection of AE pairs from the text and Web. This paper also reveals the Web-based methods for expansion disambiguation, ranking, popularity score detection, and the context words identification.

The paper is structured in the following way. Section 2 presents the acronym expansion mining approaches. The Web-based methods for disambiguation, ranking, finding popularity score, and context words of the expansion are tabulated in Sect. 3. The scope of the research work in this research area is obtainable in Sect. 4. Finally, the article is concluded in the Sect. 5.

2 Acronym-Expansion Mining Approaches

Based on the survey analysis, the approaches adopted for recognizing acronym expansion from text are categorized into two ways: (i) heuristics approach and (ii) machine learning approach. The first approach is presented in Sect. 2.1,

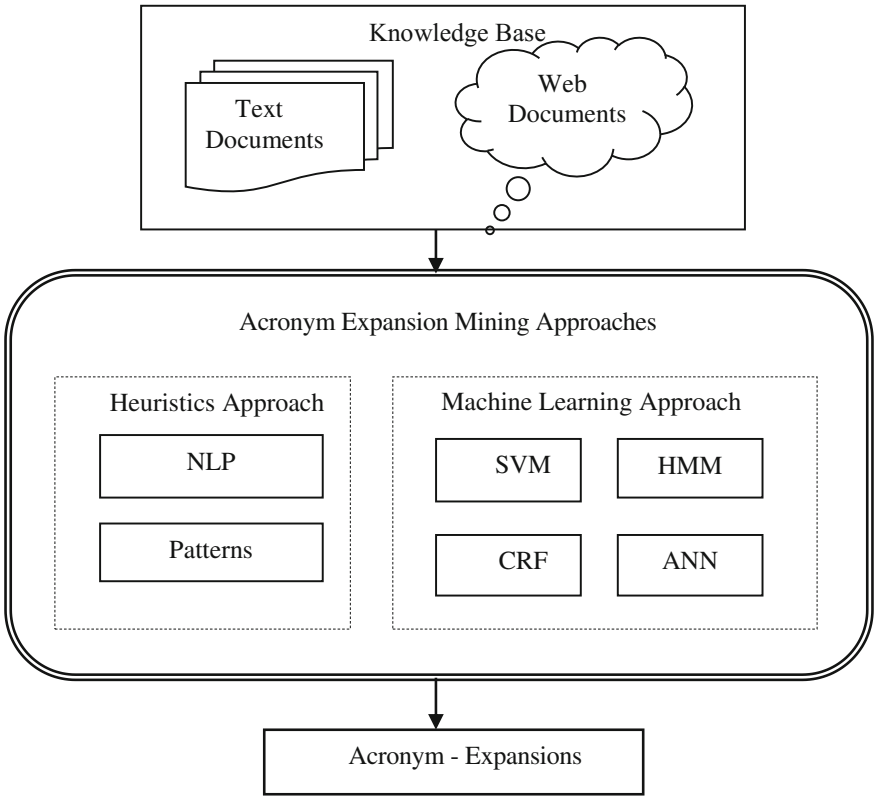


Fig. 1 Acronym-Expansion mining approaches

and the second is presented in Sect. 2.2. An Illustration of acronym expansion mining approaches is given in Fig. 1.

2.1 Heuristic Approach

Heuristics-based approach includes NLP and pattern-based approaches. The NLP-based approach uses the techniques like part of speech (POS) tagging, chunking, relation extraction to discover the expansion associated with an acronym. In pattern-based approach, hand-built rules or regular expressions are applied to the AE extraction process. The rules are manually written by considering the characteristics of acronyms such as ambiguity, nesting, uppercase letters, length, and paralinguistic markers surrounding acronym as features in the text.

A method called AcroMed [10] uses regular expression algorithm and syntactically constrained algorithm for AE extraction process from biomedical documents. A pattern matching algorithm is used for extracting AE pairs from MEDLINE

biomedical abstracts [11]. Dynamic programming is used in [12] for matching acronym and its expansion by maximizing a linguistic plausibility score. An Acronym Finding Program (AFP) is a most primitive acronym expansion identification system from free text [13]. This system applies inexact pattern matching algorithm for mining AE pairs.

A method [14] called three letter acronym (TLA) employs paralinguistic markers such as parenthesis, commas, and periods to extract acronym definition from technical and governmental documents. A database for acronym expansions is developed by applying a large number of patterns and rules on massive Web pages, and this system is known as Acrophile [15]. An algorithm [16] uses the knowledge of pattern-based contraction rules, text markers (parenthesis), cue words (such as, for example, etc.) for AE pairs extraction. In [17], the author applied four scoring rules on the set of paths to locate the most possible definition in a window of text. Pattern matching and five different space reduction heuristics are used for the selection of AE pairs from biomedical text [18].

An approach [19] proposed by the authors solves the problem in [18] by allowing the extraction of AE pairs from textual data even if the acronym and its expansion appear in different lines. In [20], the authors used collocation measures and parenthetical expressions to correlate expansions to an acronym.

In [21], the authors applied a C value measure which combines linguistic and statistical information for automatic term recognition (ATR). And this measure recommends nested terms which appear frequently in the text rather than specific long terms. In [22], the authors build a classifier based on heuristics-based features and user feedback is used to train classifier for recognizing acronym expansions. The summary of heuristics approach is presented in Table 1.

2.2 Machine Learning Approach

Machine learning approach can overwhelm the difficulties available in heuristics approach. Machine learning model can be created by using labeled examples. And labeling the data is simpler than writing rules or regular expressions. Besides that, machine learning model can employ different kinds of facts easily. In this approach, acronym definitions and context they co-occur are denoted by features. Some features are strong evidences, and some are weak evidences, but both can be recognized using machine learning approach easily.

2.2.1 SVM-Based Approach

Support vector machine (SVM)-based approach converts acronym expansion finding to a classification problem. In order to generate a proper acronym expansion pair, SVM is trained with features that are available between acronym and its corresponding definition. An approach uses space reduction heuristics on both

Table 1 Summary of heuristics-based Approaches

Approach and area	Author, year, and performance	Explanation
Approach: Pattern/ rule-based Area: Bio medical Evaluation corpus: MEDLINE abstracts [48] Implementation Language: Java	Schwartz and Hearst [11] *Performance metrics* Precision: 96% Recall: 82% F-factor: 88.4%	• Acronym length should be 2–10 characters • Algorithm fails to extract AE pairs, if there is no accurate character matching between them • The definition which contains comma character and one or two words inside parenthesis are not recognized by the algorithm
	Mohammed and Abdul Nazeer [19] *Performance metrics* Precision: 98.6% Recall: 98.6%	• The method can extract AE pairs which contain digits • It can detect AE pairs even if they appear in multiple lines • AE pairs must be appeared either in acronym (definition) or definition (acronym) in the text
	Rafeeque and Abdul Nazeer [18] *Performance metrics* Precision: 97.2% Recall: 92%	• It uses five different space reduction heuristics for detecting the candidate acronyms and definitions • This system does not recognize acronym with digit and punctuation marks • The acronym and its candidate expansion should be in the same sentence
Approach: Pattern/ rule-based Area: Text mining	Yeates S (2000) *Performance metrics* Precision: 68% Recall: 91% F-factor: 77.8%	• The system is called three letter acronym (TLA) • TLA attempts to find the candidate acronym in each chunk and the candidate expansions in the preceding and following chunks • The abbreviated expressions containing more than one upper case letter from the acronym are not recognized by TLA
	Larkey et al. [15] *Performance metrics* Precision: 87% Recall: 88% F-factor: 87.5%	• The system is called as Acrophile • A massive number of patterns and rules are used to identify AE pairs from the Web documents • The precision of Acrophile is low, due to the existence of complex relationship between AE pairs in Web documents
Approach: NLP-based Area: Text mining Evaluation corpus: MEDLINE abstracts [48]	Taghva K and Gilberth J (1999) *Performance metrics* Precision: 98% Recall: 93% F-factor: 95.4%	• The system is called as acronym finder program (AFP) • Least common subsequence (LCS) is used to extract the expansion • AFP does not support two-letter acronyms (e.g., CA, IP) • It does not allow interior letter matches in the expansion with respect to acronyms
	Pustejovsky et al. [10] *Performance metrics*	• The system is called as AcroMed • It uses regular expression and syntactically constrained algorithms for AE detection

(continued)

Table 1 (continued)

Approach and area	Author, year, and performance	Explanation
	Precision: 98% Recall: 72% F-factor: 83%	• It depends on the results of POS and hard to find the complex patterns
Approach: NLP-based Area: Bio medical	Zahariev M (2004) *Performance metrics* F-factor: 99.6%	• Recognize acronyms only if it occurs sequentially in expansion • The system does not account for acronyms with digits and symbolic characters
Approach: hybrid [NLP + pattern]	Park and Byrd [16] *Performance metrics* Precision: 97% Recall: 94% F-factor: 95.9%	• The first character of the first word must match the first acronym letter • The expression must not contain text markers, stop words at the beginning or at the end

acronym and definition, and SVM to validate AE pairs is presented in [23] by the authors. In [24], the authors used a linear approach called linear regression on a set of features to find the possible alignments between acronym and expansions. The authors [25–28] presented a SVM model, and it uses AE information as features (e.g. length, existence of special symbols, and context) for recognizing acronyms and their expansion from the text.

2.2.2 HMM-Based Approach

Hidden Markov model (HMM) is a statistical method which can be defined by a set of states and transitions among these states, forming a hidden chain. Each state produces a sequence of observation outputs but the state themselves neither or nor known. A HMM model [29] is used for acronym expansion detection. The model uses sequential structure of the sentences to find the acronym expansion pairs, but the given solution is restricted because the expansion and the acronym should be present in close vicinity. In [30], the authors used HMM model to recognize AE pairs from biomedical text. Here, the model is built by considering the alignment between characters or sequence of them in the acronym and expansion.

2.2.3 CRF-Based Approach

Conditional random fields (CRFs) are a class of statistical modeling method used for labeling, segmenting the structured data in the form of sequences, trees, etc.

CRF considers the context into account for the prediction of given input samples. It is an alternative to the HMM model. A CRF-based approach is proposed by the authors [31, 32] to write more effective features for AE that works on a group of neighboring tokens together with the features of individual tokens. They have used nonlinear hidden layers for better representation of input data.

2.2.4 Neural Network-Based Approach

In [33], the authors viewed the acronym expansion detection problem as a sequence labeling task and used a hidden layer using neural networks for modeling the feature selection process. This is commonly known as neural conditional random fields (NCRF). But their model ignores the fine-grained information due to the substructure. A hierarchical latent structure neural structured prediction model [34] is used by the authors for expansion identification. They have introduced latent state neural conditional neural fields [LNCRF] to solve the problem of expansion sequence labeling with nonlinear input features and label sub-structures.

In recent years, the usage of machine learning approach is increasing for acronym expansion detection task. The inferences of both heuristics and machine learning approach are presented in Table 2.

Table 2 Annotations of both heuristics and machine learning approach

Approach	Inferences
Heuristics	• Creating and fine-tuning rules for all kind of acronym expansion is a time consuming process • Developing patterns manually can limit the usage of information because only strong evidences can be included in patterns • Manually created features are often noisy and needless • In heuristics-based techniques, mostly the acronyms are selected with a small number of false positives, so the precision is high, but the recall is low • Some of the regular expressions are adolescent because it selects the acronyms with high precision but miss a lot of known patterns. And some regular expressions are too wide, but it selects too many false-positive acronyms
Machine learning	• In SVM-based approach, each token is classified as either positive or negative samples without considering the features of neighboring samples • The HMM-based approach requires expert prior knowledge • The performance of CRF is heavily depends on the quality of the input features, but it is difficult for the human to build these features. Each class label is having the complex sub-structures, but it ignores the intermediate substructure information that lead to lose important information • In neural network-based approach, latent state variables are used to capture the granular structures of each class and learn the high-level representation of complex input features

3 Web-Based Acronym Expansion Approaches

The research works which are related to mining acronym definitions based on Web resources, disambiguating, and ranking the expansions. Finding the popularity score and context words of the expansions is discoursed in this section.

A method [35] recommends a most appropriate expansion of an acronym. This is done through a dataset built from Wikipedia disambiguation pages [36] by using simple syntactic patterns. In [37], the authors used decision tree learning program as classifier for disambiguating expansions. An unsupervised method [38] is employed by the authors to extract acronym and its definition from Web. They have used patterns and constraints for configuring their model as domain and language independent. In [39–41], the authors used paralinguistic features and statistical measures for extracting the acronym expansion from Web documents.

The heuristic approach is used for checking uppercase and parenthesis in text. And linguistic features are employed for identifying phrase structures to identify candidate sentence [42]. The system [43] discovers the expansions of acronym from query-click log files of a search engine. It also employs methods for finding popularity score and context words of each expansion. A pattern-based approach called AcroMiner [44] used for extracting acronym expansion from Web documents and computed the rank of each expansion.

The authors [45–47] focused on generating knowledge map by recognizing acronym expansion from two large-scale unstructured data sets; they are (i) Wikipedia (collective intelligence) and (ii) NDSL (scholarly database). From Wikipedia, acronym expansion pairs are extracted from disambiguation pages; synonyms of expansions are identified via URI redirection page. From NDSL, acronym expansion is extracted from free text. The NN type and NP type features are used to train naïve Bayesian classifier to recognize correct acronym expansion pairs. The summary of above-said approaches are presented in Table 3.

4 Scope for the Future Work

The survey result suggests that: (i) To extract AE pairs from text documents like biomedical documents, the researchers can use either heuristics or any one of the machine learning approach. (ii) To disambiguate the expansion, finding the ranking score of an expansion, and identifying the context words related to an expansion, the researchers can employ the Web resources like Search Engine Results Page (SERP), log files of search engine, and Web-crawled documents.

As an extension of the survey, we have planned for detecting list of definitions of an acronym from SERP of Google by using machine learning approach. Few of the web-based acronym expansion approach uses snippets as resource from SERP for the detection of expansion. But we have planned to use titles from SERP for detecting expansions, because the availability of expansions in title part is higher

Table 3 Web-based methods for exploring acronym expansions

Author and year	Work and corpus	Implementation
Do-Heon Jeong et al. (2014), (2015) and (2016)	Work: generating knowledge map for acronym expansion. *Evaluation corpus* • Wikipedia and NDSL abstracts.	• Constructed a gold standard source from NDSL for AE disambiguation • For disambiguating, the expansion with highest frequency is selected as right definition for the acronym
Sumita and Sugaya [37]	Work: disambiguation of expansion *Evaluation corpus* • Wikipedia • www.acronymsearch.com	• Focused on disambiguating acronyms by employing decision tree learning program as classifier • To train the classifier, top N frequent words obtained from snippets of "acronym AND definition" query are used as features
Donjin Choi et al. [35]	**Work**: recommending most appropriate expansion using Wikipedia *Evaluation corpus* • Wikipedia extended abstracts	• Recommend most relevant expansion words for an acronym • Linguistic approach is used to detect AE pairs from Wikipedia • WUP similarity measure is used to discover the most appropriate expansion • It cannot detect expansion if the order of the capital letter of successive words is not same as the given acronym
Sanchez D, Isren D, (2011)	Work: mining acronym expansion from Web. *Evaluation corpus* • Web search engine results	• Iterative query expansion algorithm is used in Web search to retrieve more number of definitions • Domain and language-independent approach
Mathieu Roche, Violin Prince (2010)	Work: disambiguate biomedical acronym expansion. *Evaluation corpus* • Biomedical documents provided by AcroMiner. • 102 Acronym AE pairs identified.	• Provided nine quality measures for appropriate definition prediction. And those measures are based on mutual information, cubic MI, and Dice's coefficient • Difficult in building a context from the Web page and context extraction relies on words frequency
Mathieu Roche, Violin Prince (2008), (2014)	Work: deals with extraction of acronym/expansion and disambiguating of these definitions. *Evaluation corpus* • www.sigles.net	• Paralinguistic markers and statistical measures are used for AE extraction • Alta vista search engine and Java used for implementation • Since more global patterns are used, a lot of noisy data are returned

(continued)

Table 3 (continued)

Author and year	Work and corpus	Implementation
Alpa Jain et al. [42]	Work: web-based AE extraction and ranking the expansions for an acronym. *Evaluation corpus* • Web search engine results (live search)	• AE pairs are extracted from three sources (i) Crawled Web documents (ii) Search engine logs and (iii) Web search results • They compared each resources performance in terms of precision and recall
Xiaonan J (2008)	Work: mining, ranking, and using acronym patterns *Evaluation corpus* • V.E.R.A acronym dictionary [49].	• The system is known as AcroMiner • Two strategies are used, namely lower level and upper level for recognizing AE patterns • Rank score is controlled by three factors; they are (i) Pattern popularity (ii) Gap between the acronym (iii) Mapping score of AE.
Bilyana Taneva et al. [43]	Work: mining acronym expansion, finding popularity score, and context words for each acronym. Data Source: query clicks logs of bing 2010 and 2011. *Evaluation corpus* Wikipedia disambiguation pages [36].	The system performs four important tasks with acronym expansion. They are as follows: • Candidate expansion identification • Acronym expansion clustering • Enhancement for tail meanings • Canonical expansion, popularity score computation, and context words identification

than the snippet part of SERP. Moreover, the text content subject to mine for detecting expansion is lesser in this proposed idea.

Sequence labeling is an ideal method for this acronym expansion detection process because writing rules or regular expressions for all kind of expansions is a huge task. The machine learning approach is well suited for this sequence labeling task. In recent years, few authors [31, 32, 34] used this sequence labeling method in their work for detecting expansion from text documents. The statistical models like HMM, CRF, and ANN are appropriate for this acronym expansion identification task.

Basically, acronym is an ambiguous one; i.e., it can have multiple definitions. In order to retrieve the maximum possible definitions of an acronym, the researcher can use Web search engine result pages. To find the rank of each expansions, the researcher can utilize fuzzy systems by framing rules in such a way that the rank of the expansions will come under desired categories.

5 Conclusion

This survey reveals the approaches followed for mining acronym expansion from free formatted text and Web. To recognize acronym and its expansion from text and Web documents, heuristics and machine learning approaches are mainly employed by different authors in the past two decades. Heuristics approaches uses NLP and pattern matching concepts. Statistical models like SVM, CRF, HMM, and neural networks are used in machine learning approaches. Both the approach has its own merits and demerits. To assess the performance of those approaches, the researchers predominantly used three metrics namely precision, recall, and F-score. And it is observed that the recall value is lesser in heuristics approach than the machine learning approach. In web-based acronym expansion detection approach, few authors have been focused on disambiguation of acronym expansions, generating knowledge map for AE pairs, finding popularity score, and context words of the expansion. Mostly used Web resources by themselves are Wikipedia, Web-crawled documents, log files, and search engine results.

References

1. http://www.acronymfinder.com
2. http://www.abbreviations.com
3. http://www.acronymslist.com
4. https://acronyms.thefreedictionary.com/
5. http://www.special-dictionary.com/acronyms/
6. http://www.acronymsearch.com
7. https://www.allacronyms.com
8. http://acronyms.silmaril.ie
9. http://acronym24.com
10. Pustejovsky, J., Castano, J., Cochran, B., Kotecki, M. Morrell, M.: Extraction and disambiguation of acronym meaning-pairs in MEDLINE. In: Proceedings of 10th Triennial Congress of the International Medical Informatics Association, pp. 371–375. MEDINFO, IOS Press, London, (2001)
11. Schwartz, A., Hearst, M.: A simple algorithm for identifying abbreviation definitions in biomedical text. In: Pacific Symposium on Biocomputing, vol. 8, pp. 451–462 (2003)
12. Zahariev, M.: Efficient acronym – expansion matching for automatic acronym acquisition. In: International Conference on Information and Knowledge Engineering, pp. 32–37 (2003)
13. Taghva, K., Gilberth, J.: Recognizing acronyms and definitions. Information Science Research Institute, University of Nevada, Technical Report TR, pp. 191–198 (1999)
14. Yeates, S.: Automatic extraction of acronyms from text. In: Proceedings of third New Zeland Computer Science Research Student's Conference, pp. 117–124, University of Waikato, New Zealand (1999)
15. Larkey, L.S., Ogilvie, P., Price, M.A., Tamilio, B.: Acrophile: an automated acronym extractor and server. In: Proceedings of 5th ACM Conference on Digital Libraries. Association for Computing Machinery, pp. 205–214 (2000)
16. Park, Y., Byrd, R.J.: Hybrid text mining for finding abbreviations and their definitions. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, EMNLP, pp. 126–133. Intelligent Information System Institute, Pittsburgh (2001)

17. Adar, E.: S-RAD: a simple and robust abbreviation dictionary. HP Lab. Bioinform. **20**(4), 527–533 (2004)
18. Rafeeque, P.C., Abdul Nazeer, K.A.: Text mining for acronym -definition paris from biomedical text using pattern matching method with space reduction heuristics. In: Proceedings of 15th International Conference on Advanced Computing and Communications, pp. 295–300. IEEE, IIT Guwahati, India (2009)
19. Saneesh Mohammed, N., Abdul Nazeer, K.A.: An improved method for extracting acronym–definition pairs from biomedical literature. In: International Conference on Control Communication and Computing (ICCC), pp. 194–197. IEEE (2013)
20. Liu, H., Friedman. C.: Mining terminological knowledge in large in biomedical corpora. In: Proceedings of 8th Pacific Symposium on Biocomputing, PSB Association, Lihue, pp. 415–426 (2003)
21. Okazaki, N., Ananiadou, S.: A term recognition approach to acronym recognition, In: Proceedings of the COLING – ACL'06, pp. 643–650. ACM, Sydney (2006)
22. Yarygina, A., Vassilieva, N.: High – recall extraction of acronym – definition pairs with relevance feedback. In: BEWEB, pp. 21–26, ACM, Berlin (2012)
23. Nadeau, D., Turney, P.: A supervised learning approach to acronym identification. In: Proceedings of 18th Conference of the Canadian Society for Computational Studies of Intelligence, pp 319–329. Springer, Berlin (2005)
24. Chang, J.T., Schutze, H., Altman, R.B.: Creating an online dictionary abbreviation from MEDLINE. J. Am. Med. Inform. Assoc. **9**(6), 612–620 (2002)
25. Xu, J., Huang, YL.: A machine learning approach to recognizing acronyms and their expansion. In: International Conference on Machine Learning and Cybernetics, IEEE, China (2005)
26. Xu, J., Huang, Y.L.: Using SVM to extract acronyms from text. Soft Computing, pp. 369–373. Springer, Berlin (2006)
27. Ni, W., Xu, J., Huang, Y., Liu, T., Ge, J.: Acronym extraction using SVM with uneven margins. In: Proceedings of the 2nd IEEE Symposium on Web Society, pp. 132–138. IEEE, Beijing (2010)
28. Gao, Y.M., Huang, Y.L.: Using SVM with uneven margins to extract acronym expansion. In: Proceedings of the 8th International Conference on Machine Learning and Cybernetics, pp. 1286–1292, IEEE, Baoding (2009)
29. Taghva, K., Vyas, L.: Acronym expansion via hidden Markov models. In: Proceedings of International Conference on Systems Engineering, IEEE, pp. 120–125 (2011)
30. Osiek, B.A., Xexeo, G., de Carvalho, L.A.V.: A language - independent acronym extraction from biomedical texts with hidden Markov models. IEEE Trans. Biomed. Eng. **57**(11), 2677–2688 (2010)
31. Nautial, A., Sristy, N.B., Somayajulu, D.V.L.N: Finding acronym expansion using semi-Markov conditional random fields. In: Compute 2014, India, pp. 16:1–16:6. ACM, (2014)
32. Liu, J., Chen, J., Liu, T., Huang, Y.: Expansion finding for given acronyms using conditional random fields. In: WAIM, pp. 191–200 (2011)
33. Liu, J., Liu, C., Hu, Q., Huang, Y.: Fine – grained acronym expansion identification using latent-state neural structured prediction model. In: Proceedings of International Conference on Machine Learning and Cybernetics, pp. 259–264. IEEE, Guangzhou (2015)
34. Liu, J., Liu, C., Huang, Y.: Multi-granularity sequence labeling model for acronym expansion identification. Inf. Sci. **38**, 462–474 (2017)
35. Choi, D., Shin, J., Lee, E., Kim, P.: A method for recommending the most appropriate expansion of acronyms using wikipedia. In: Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IEEE, pp. 217–220 (2013)
36. https://en.wikipedia.org/wiki
37. Sumita, E., Sugaya, F.: Using the web to disambiguate acronyms. In: Association for Computational Linguistics (ACL), pp. 161–164. New York (2006)
38. Sanchez, D., Isren, D.: Automatic extraction of acronym definitions from the Web. J. Appl. Intell. **34**(2), 311–327 (2011)

39. Roche, M., Prince, V.: A web-mining approach to disambiguate biomedical acronym expansions. Informatica **34**, 243–253 (2010)
40. Roche, M., Prince, V.: Managing the acronym/ expansion identification process for text - mining applications. Int. J. Softw. Inf. **2**(2), 163–179 (2008)
41. Roche, M.: How to exploit paralinguistic features to identify acronyms in text. In: International Conference on Language Resources and Evaluation, Reykjavik, Iceland, pp 69–72 (2014)
42. Jain A., Cucerzan, S., Azzam, S.: Acronym-expansion recognition and ranking on the web, In: Proceedings of the IEEE International Conference on Information Reuse and Integration (IRI 2007), pp. 209–214 (2007)
43. Taneva, B., Cheng, T., Chakrabarthi, K., He, Y.: Mining Acronym Expansions and their Meanings Using Query Log. WWW 2013, pp. 1261–1271. ACM, Brazil (2013)
44. Ji, X., Xu, G., Bailey, J., Li, H.: Mining, ranking, and using acronym patterns, In: Proceedings of the 10th Asia-Pacific Web Conference on Progress in WWW Research and Development, pp. 371–382 (2008)
45. Jeong, D.H., Gim, J., Jung, H.: Incremental discriminating method for acronyms in heterogeneous resources. Int. J. Adv. Soft Comput. Appl. **7**(1), 59–67 (2015)
46. Jeong, DH.., Hwang, M.G., Kim, J., Jung, H. Sung, W.K.: Acronym- expansion recognition based on knowledge map system. Int. Inf. Inst. (Tokyo). Inf. Koganei **16**(12), 8403–8408 (2013)
47. Jeong, D.H., Hwang, M.G., Sung, W.K.: Generating knowledge map for acronym– expansion recognition. In: International Conference on U-and E-Service, Science and Technology, (UNESST), pp 287–293 (2011)
48. http://www.ncbi.nil.nih.gov [MEDLINE Abstracts]
49. http://www.delorie.com/gnu/docs/vera/vera_toc.html [V.E.R.A]

Deformable Facial Fitting Using Active Appearance Model for Emotion Recognition

Lakshmi Sarvani Videla, M. R. Narasinga Rao, D. Anand, Hima Deepthi Vankayalapati and Shaik Razia

Abstract In facial emotion recognition, facial features of the person have to be detected first. For this, active appearance models (AAMs) are used in this work. The main drawback of AAM is that it cannot generalize to unseen faces. To overcome this drawback, the training images are pre-processed before using them for model construction in this work. For automatic initialization of the model, the output of Viola–Jones face detector is used. In literature, principal component analysis (PCA) is used to capture the main variations of the training data in AAM. As the variance in PCA is fixed, they produce large models which are difficult to optimize. In this work, fisher face method is used in which PCA is first used to reduce the dimensions of the feature space so that the resulting within class scatter matrix is non-singular and then apply the linear discriminant analysis (LDA) to reduce the dimensions still further. The experimental results on unseen and seen faces show that fitting accuracy is better and takes less time to fit when compared with existing AAMs.

L. S. Videla (✉) · M. R. N. Rao · D. Anand · S. Razia
Department of Computer Science, K L E F, Vaddeswaram, Guntur, Andhra Pradesh, India
e-mail: sarvani.mtech2014@gmail.com

M. R. N. Rao
e-mail: manda.ramanarasingarao@gmail.com

D. Anand
e-mail: ananddama92@gmail.com

S. Razia
e-mail: razia28sk@gmail.com

H. D. Vankayalapati
Department of Electronics and Communication Engineering, VelTech University, Chennai, Tamil Nadu, India
e-mail: vhdeepthi83@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_13

1 Introduction

Emotions play an important role in human's everyday activities such as communication, task solving, concentration, learning, and decision making. Emotions are expressed through speech, facial expressions, gestures, and other non-verbal clues. But, the psychological studies indicated that facial expression is the most natural and primary cue for communicating emotions contributing to 55% [1].

Facial Action Coding System (FACS) [2, 3] and Facial Animation Parameters (FAPs) [4] are two systems that measure facial expressions by analyzing relations between muscle(s) contraction and changes in the face appearance caused by them. In order to identify the emotion from a facial image, outline of eyebrows, mouth, eyes, and face outline is required. To get these outlines from a face image, a model has to be created from the annotated training face dataset, which contains facial images that are annotated around the outlines of facial features. As these annotated training faces are high dimensional data, it is difficult to train a model using this dataset. To overcome this problem, many linear and nonlinear dimensionality reduction techniques were proposed such as PCA, LDA, independent component analysis (ICA), linear projective maps (LPP), etc. In recent years, Moin A et al. has used weighted PCA for dimensionality reduction for fusion of multimodal medical images in [5]. Weighted PCA enabled them to choose the most discriminative principal components. Krishna et al. [6] used 2D ridgelet transform and then PCA for fusion of medical images. They achieved improvement in quality of the images. PCA was used in ridgelet and wavelet domain in [7]. The proposed work in this paper uses fisher face method [8] that combines PCA and LDA. As PCA captures variations among all training images producing large models which are difficult to optimize, LDA is used as a technique after PCA to reduce the dimensions still further. The rest of the paper is organized as follows: The proposed methodology is discussed in Sect. 2. Experimental results are shown in Sect. 3. Conclusion and future work are presented in Sect. 4.

2 Proposed Methodology

The existing active appearance models are domain specific and cannot easily generalize to unseen faces. Hence, some preprocessing of training appearances has to be done. For proper initialization and initial scale of the model, the output of Viola and Jones face detection algorithm is used.

To train AAMs, annotated training images are required. As we are training AAM for facial emotion recognition, we require annotated training images of faces with various expressions. These kinds of data are not available, and so training shapes and appearances are taken by selecting the peak emotion images and their corresponding landmarks from the Extended Cohn-Kanade CK + database [9]. To avoid the background linking problem of AAMs, the training appearances are prepared appropriately.

The proposed work uses the CK + database [9]. In CK + , all videos follow the temporal dynamic sequence: neutral- > onset- > apex. There are 593 sequences across 123 subjects which are FACS coded at the peak frame. Only peak frame images are used in this paper. Like the FACS files, there is only 1 emotion file for each peak frame. Emotions are numbered as 0 = neutral, 1 = anger, 2 = contempt, 3 = disgust, 4 = fear, 5 = happy, 6 = sadness, 7 = surprise.

The whole work is divided into five modules and they are preparing training data (appearances) module, learning shape model module, learning appearance model module, fitting module, and facial expression recognition module.

2.1 Preparing Training Data (Appearances) Module

Steps involved in preparing the training data are as follows

1. Read each input image in the training set.
2. Detect the face in the input image using Viola and Jones face detection algorithm.
3. Get the coordinates of the bounding box around the detected face.
4. All the pixel intensities except the area bounded by the bounding box are made black.
5. Add virtual structuring element (VSE) to the images.

The facial images after adding virtual structuring element result in faces bounded by an octagon. All other pixel intensities are made black except the area bounded by octagon.

The block diagram of the proposed work is as shown in Fig. 1.

2.2 Learning Shape Model Module

Procrustes Analysis for Aligning Shapes
In order to align all the training shapes Procrustes analysis [10] is used

1. Arbitrarily choose a reference shape (typically by selecting it among the training shapes)
2. Superimpose all shapes to current reference shape
3. Calculate the mean shape of the current set of superimposed shapes
4. If the Procrustes distance between mean and reference shape is above a threshold, set reference shape to mean shape and go to step 2.

Procrustes distance is the square root of sum of squared distances (SSD) between corresponding points and the output after applying Procrustes Analysis are aligned shapes. In this work, the threshold used is 0.001.

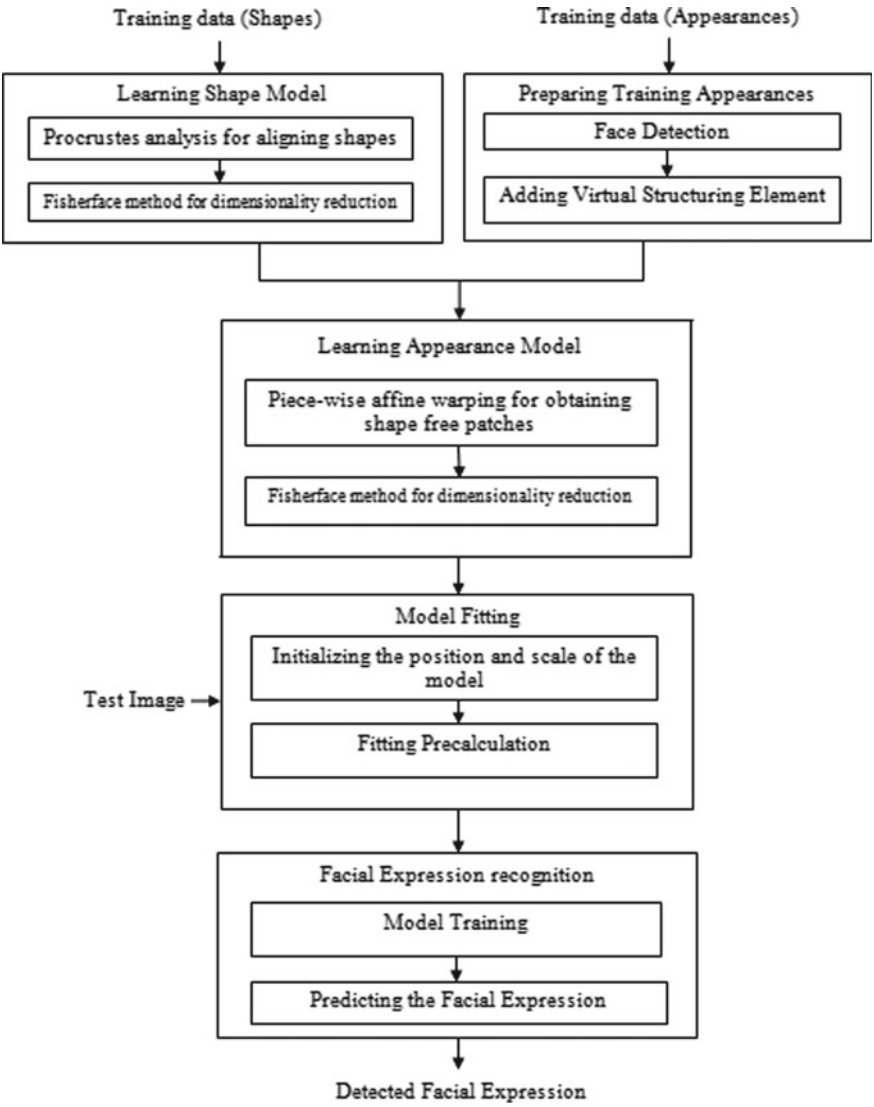


Fig. 1 Proposed method for deformable facial model fitting for emotion recognition

Fisher Face Approach

1. Compute global mean of all the training shapes using Eq. (1)

$$M = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

2. Calculate the covariance matrix using Eq. (2) [8]

$$\text{cov} = \sum_{i=1}^{n} (X_i - M)(X_i - M)^T \tag{2}$$

3. $d \leftarrow \max\left(\text{cov}, \text{cov}'\right)$
4. Compute eigen vectors and eigen values of d.
5. Sort the eigen vectors and eigen values in descending order.
6. $\text{evalues} \leftarrow \left(\frac{1}{\sqrt{\text{eigen values}}}\right)$
7. $\text{evectors} \leftarrow \left(X' * \text{eigen vectors}\right) * \text{evalues}$
8. $H_i = \sqrt{cn_i} * \left(\frac{1}{cn_i}\left(\sum_{k=1}^{cn_i} \text{evec}_i\right)\right)$
   where cn_i are the number of evectors belonging to ith class.
   and evec_i are the evectors of ith class
9. Let the eigenvectors and eigenvalues of H be eigvector and eigvalues after discarding smallest eigenvalues and their corresponding eigenvectors.
10. eigvectors ←(evectors*(evalues*eigvector))
11. Normalize the eigvectors.
12. Output the eigvalues and normalized eigvectors obtained in step 11.

Mean shape s_0 is obtained by Eq. (1), and eigenvectors are obtained from fisher face approach.

Shape model is obtained by Eq. (3) [10]

$$s = s_0 + \sum_{i=1}^{n} p_i s_i \tag{3}$$

where n are the first n largest eigenvalues, s_i are the corresponding eigenvectors, and p_i are the coordinates of shape in lower dimensions.

2.3 *Learning Appearance Model Module*

Piecewise affine warping is used to warp training appearances to mean shape as follows:

1. Find the coordinates of vertices of all the triangle in s_0.
2. For each pixel within s_0, for each triangle, if the pixel x is inside the triangle, then compute α, β, γ. If $0 \leq \alpha, \beta, \gamma \leq 1$, then counter = counter +1 as α + +γ = 1, and if the pixel is not inside the triangle, then α = β=γ = 0.

3. Update the warp table entries. Each warp table entry contains the following:
   a. i, pixel's x coordinate value,
   b. j, pixel's y coordinate value,
   c. The triangle in which the pixel x at (i, j) is present,
   d. α, β, γ which are the relative position of pixel w.r.t vertices of the triangle in s_0 and are computed using Eqs. (4)–(6) [11]

$$\alpha = 1 - (\beta + \gamma) \tag{4}$$

$$\beta = \frac{yx_3 - yx_1 - y_1x_3 - y_3x + x_1y_3 + xy_1}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2} \tag{5}$$

$$\gamma = \frac{xy_2 - xy_1 - x_1y_2 - x_2y + x_2y_1 + x_1y}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2} \tag{6}$$

4. For each pixel, x in the Warp table, find the corresponding pixel W(x; p) in the training shape s using Eq. (7) [11]

$$W(x;p) = x' = \alpha x_1' + \beta x_2' + \gamma x_3' \tag{7}$$

where x_1', x_2', x_3' are the vertices of the triangle in which pixel x' lies.

5. Find the intensity of pixel x' by using bilinear interpolation from corresponding appearance of s and set the intensity at x to that value. Apply fisher face method on the shape-free patches to obtain eigenvalues, λ_i, and eigenvectors A_i.The output of this module is appearance model, A_0.

2.4 Fitting Module

Fitting of the shape and appearance model to the test image is done using inverse compositional image alignment. Here, fitting an AAM to an image consists of minimizing the Eq. (8) [12]

Update the entire warp by composing the current warp with the computed incremental warp.

$$\sum_{x \in s_0} \left[\sum_{i=1}^{m} \lambda_i A_i(x) - I(N(W(x;p);q) - A_0(x) \right]^2 \tag{8}$$

For any pixel x in s_0, there is corresponding pixel W(x;p) in test image. At pixel x, the AAM has appearance $A = A0 + \sum_{i=1}^{m} \lambda_i A_i(x)$ and the test image has intensity

I(W(x;p)). We want to minimize the sum of square differences between these two quantities, and the sum is calculated overall pixels in base mesh s_0. Update the entire warp by composing the current warp with the computed incremental warp.

2.5 Facial Expression Recognition Module

LDA is used as transformation for classification; the input to the classifier is the training data (shape) files and their corresponding class value based on the emotion displayed. The fitted shape to the test image is fed to the trained classifier to predict the class or the emotion of the test image.

3 Results

Mean shape obtained after aligning all training shapes using Procrustes analysis [10] is as shown in Fig. 2.

All the training appearances are warped to mean shape, and the resulting appearances are used for learning appearance model.

The proposed method performs better than PCA in terms of accuracy and time taken to fit. Figure 3 shows the comparison of the existing and proposed technique on IMM dataset [13]. IMM dataset [13] consists of 240 frontal images. The dataset can be freely used for education and research.

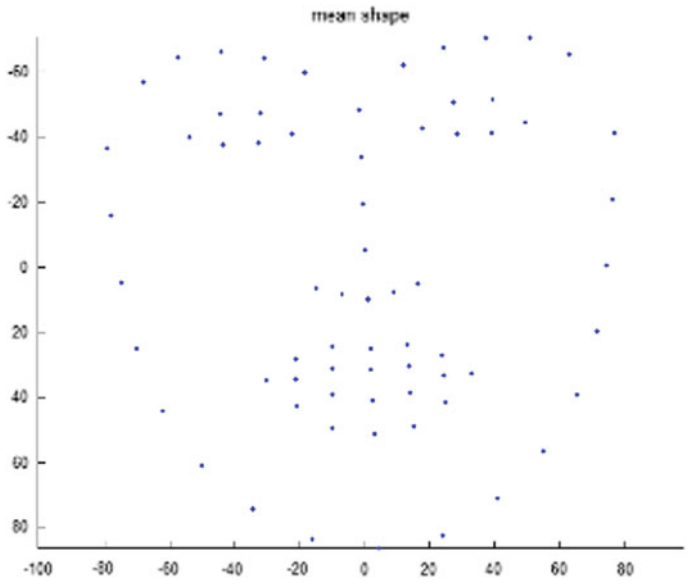


Fig. 2 Mean shape

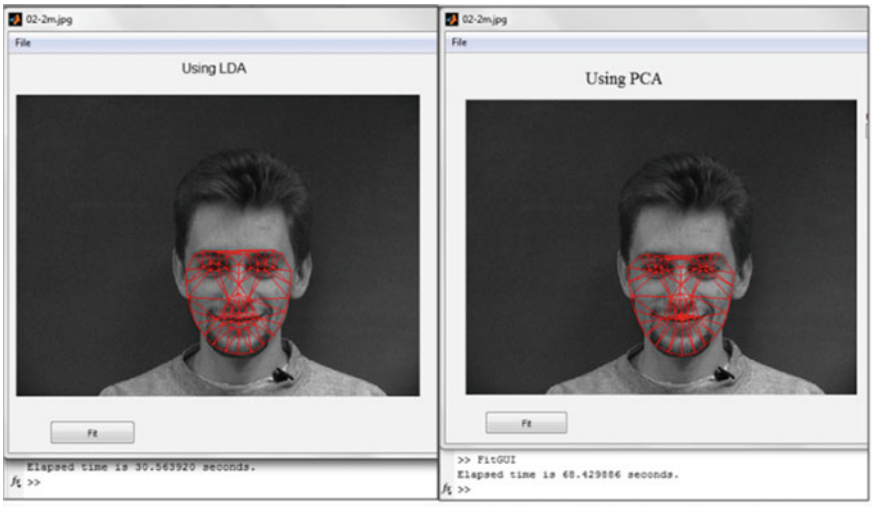


Fig. 3 Comparison on IMM dataset [13]

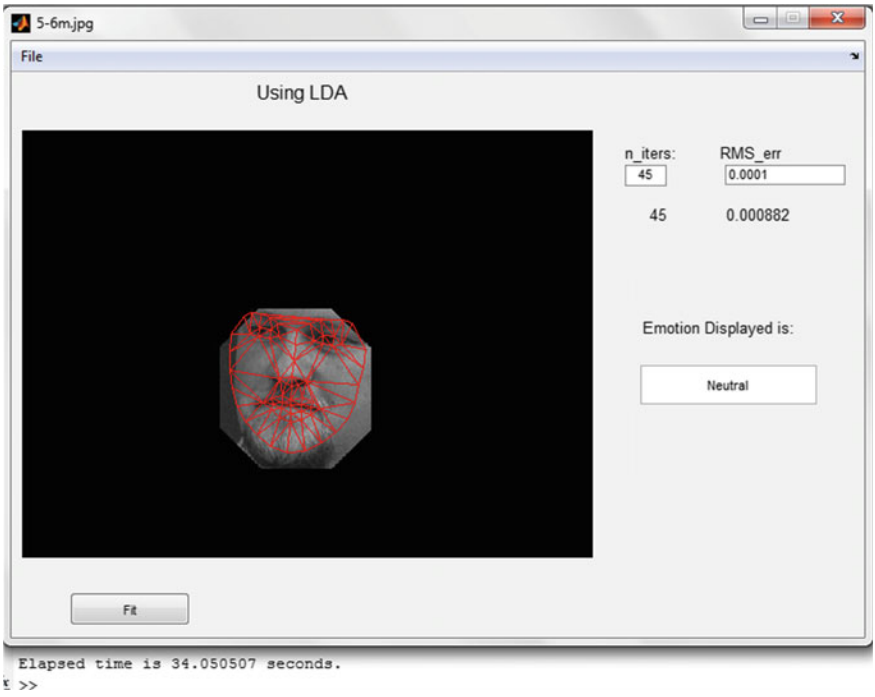


Fig. 4 Fitting on tilted face taken from IMM dataset [13] for emotion recognition

The proposed method also fits tilted faces approximately. The time taken to fit the tilted face taken from IMM dataset is 34 s as shown in Fig. 4. It also displays the emotion based on the fitted shape.

Table 1 Elapsed time comparison of proposed AAM-based emotion recognition system

Dataset	Elapsed time in seconds	
	Existing method	Proposed method
Trained images	29.80	19.32
CK + dataset	31.61	25.14
IMM dataset	68.42	30.56
Real-time images	55.21	40.92

Average time taken by the proposed method and existing method to detect the emotion of 20 images taken from trained, CK + dataset, IMM dataset, and real-time images is shown in Table 1. Twenty images are taken from each dataset.

The proposed approach is an enhancement of the existing approach. The proposed approach shows better-recognized results than the results of existing approach and also works in different lighting condition and with glasses. For PCA to capture 90% variations, 24 eigen shape vectors and 40 eigen appearance vectors were required when there are 84 training images. The proposed method requires only 13 shape and appearance eigen vectors for 84 training images.

4 Conclusion

Fisher face method (PCA and LDA) is used instead of PCA in AAM. The proposed method not only reduces the dimensions but also achieves better fitting accuracy and takes less time. Even though the training images are all frontal faces with different expression and without glasses, the proposed method fits approximately to faces with glasses and tilted faces. Our proposed method has significantly improved the fitting of model to unseen faces using a model with less number of dimensions and time.

References

1. Mehrabian, A.: Communication without words. Psychol. Today **2**(4), 53–56 (1968)
2. Ekman, P., Huang, T., Sejnowski, T., Hager, J.: Final Report to NSF of the Planning Workshop on Facial Expression Understanding (1992)
3. Ekman, P., Friesen, W.V.: Manual for the Facial Action Coding System. Consulting Psychologists Press, Palo Alto (1977)
4. MPEG Video and SNHC: Text of ISO/IEC FDIS 14 496-3: Audio. In: Atlantic City MPEG Mtg (1998)
5. Moin, A. et al.: Weighted-PCA based multimodal medical image fusion in contourlet domain. In: Proceedings of the International Congress on Information and Communication Technology 2016, pp. 597–605. Springer, Singapore
6. Krishn, A. et al.: PCA based medical image fusion in ridgelet domain. In: Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014 2015, pp. 475–482. Springer, Cham

7. Bhateja, V. et al.: Medical image fusion in wavelet and ridgelet domains: a comparative evaluation. Int. J. Rough Sets Data Anal. (IJRSDA) **2**(2), 78–91 (2015)
8. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
9. Lucey, P., Cohn, J.F., Saragih, J., Matthews, I., Ambadar, Z.: The extended Cohn-Kanade database: a complete facial expression database for both facial action units and emotion detection. In: IEEE CVPR4HB (2010)
10. ACTIVE APPEARANCE. MODELS: Theory, Extensions & Cases, 2nd edn. Mikkel Bille Stegmann. LYNGBY (2000)
11. Baker, S., Matthews, I.: Lucas-kanade 20 years on: a unifying framework. Int. J. Comput. Vis. **56**(1), 221–255 (2004)
12. Matthews, Iain, Baker, Simon: Active appearance models revisited. Int. J. Comput. Vis. **60**(2), 135–164 (2004)
13. Stegmann, M.B., Ersboll, B.K., Larsen, R.: FAME - a flexible appearance modeling environment. IEEE Trans. on Med. Imaging **22**(10), 1319–1331 (2003)

Grey Wolf Optimization-Based Improved Closed-Loop Speed Control for a BLDC Motor Drive

Devendra Potnuru and Ayyarao S. L. V. Tummala

Abstract Brushless DC motor (BLDC) is increasingly being used in many practical applications, and the speed tracking performance of closed-loop control has been an important issue for high-performance applications. In spite of many available nonlinear and adaptive controllers, the PID controller is still being used as speed controller because of its simple structure and easy to implement. However, the tuning of PID gains is a tedious process to achieve the desired performance. This paper focuses on the design of speed controller, and tuning its gains is grey wolf optimization. An objective function is formulated based on integral square error of speed error so as to improve the transient and steady-state performance. The objective function is minimized using grey wolf optimization (GWO). The proposed idea is evaluated by simulating the system under various operating conditions, and further, the experimental results show the supremacy of the controller.

1 Introduction

Permanent magnet brushless DC motor has been used in many day-to-day applications. There are two types such as PMSM and BLDC motor based on its construction. However, BLDC motor has its own advantages such as good control flexibility and lesser cost. The BLDC motor drive essentially consists of inverter, rotor position speed and position sensor. As it has been run in self-control mode, the measuring of the rotor angular position is essential, and hence, there are two categories of speed control such as sensored and sensorless control [6]. However, for closed-loop speed of BLDC motor drive with PID controller is robust when we design the speed controller with proper PID gains. Hence, many researchers

D. Potnuru (✉) · A. S. L. V. Tummala
GMR Institute of Technology, Rajam, Andhra Pradesh, India
e-mail: devendra.p@gmrit.org

A. S. L. V. Tummala
e-mail: Ayyarao.tslv@gmrit.org

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_14

proposed PID gain tuning approaches such as Ziegler–Nichols approach, PID gain scheduling, fuzzy gain scheduling [1].

PID controller is one of the most popular and industry-opted linear controllers. Now finding the suitable values for gains of proportional, integral and derivative gains is called as PID tuning. However, this approach is not that simple, and it takes more time for some systems for getting the desired performance. Some of the approaches are Ziegler–Nichols tuning method, manual tuning, fuzzy-neuro approaches and optimization-based methods. Sometimes rule-based tuning might not support certain higher order plants and systems with little time delay. Nowadays many researchers suggesting the tuning of PID controller by nature-inspired optimization method and considering the minimization of integral square error as an objective function.

Many methods such as PSO, GA, DE are applied to speed control of BLDC motor [2, 3, 8]. However, the tuning of PID controller using grey wolf optimization algorithm for BLDC motor is not yet implemented for the efficient tuning of the PID gains and enhances the speed control performance of the drive. This algorithm is developed based on the group hunting nature of one of the most successful predators, grey wolves [4, 5]. This paper proposes optimal speed control design by tuning the parameters of PID controller using GWO. These nature-inspired algorithms are used for tuning of PID controllers.

The paper is organized as follows: Section 2 elaborates the BLDC drive operation. Objective function formulation and optimal control design are elaborated in Sect. 3. Section 4 analyses the experimental results and conclusions are given in the last section.

2 The Drive Scheme of BLDC Motor

The three-phase BLDC motor drive scheme is shown in Fig. 1. It consists of inverter, BLDC motor with speed sensor and hall signals and dSPACE DS 1103 controller which supports the rapid prototyping. In the drive operation, the angular speed of motor is compared with the command speed and error is fed to PID controller. The torque command is obtained based on the speed error. The output of the PID controller is scaled down using the motor torque constant to obtain the reference currents of the hysteresis current controller. The current controller generates the required PWM pulses to the inverter switches.

The block diagram for PID tuning for speed control of BLDC motor drive using grey wolf algorithm is shown in Fig. 2. For more details of the drive scheme, one can refer [7].

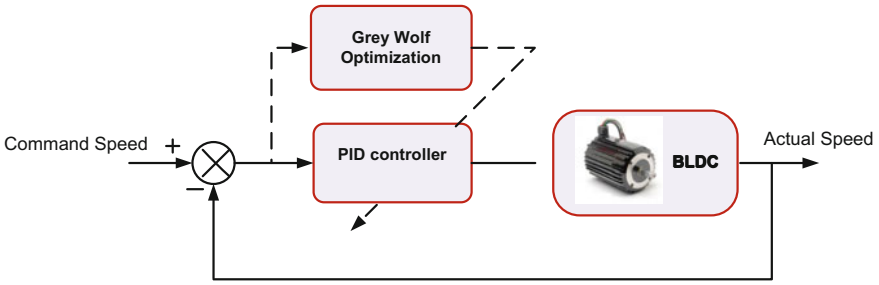
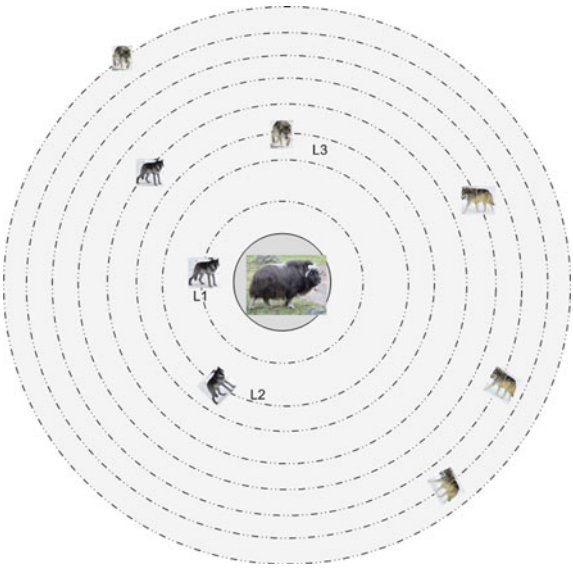


Fig. 1 Speed controller of BLDC motor using GWO algorithm

Fig. 2 Grey wolves attacking the prey



3 Grey Wolf Optimization

In the year 2013, another innovation took place in the field of optimization by Mirjalili and developed nature-inspired algorithm based on hunting nature followed by grey wolves [4, 5]. A group of around ten wolves will be led by a spearhead lead, and the group respects the decision of the lead1. Second rank leaders lead2 of the group govern low-rank wolves. Third rank leaders of the group are lead3. The remaining wolves are the working members of the group.

Grey wolves identify the prey and hunt the prey. Later they surround the prey and harass the prey and finally attack the prey. Let X_k represent the positions (solutions) of grey wolves, and k represents current iteration. X_T is the target position. Then the updated positions (solutions) of the grey wolves are given by (1)

$$X_{k+1} = X_T - \xi|\theta X_T - X_k| \tag{1}$$

where $\theta = 2.rand()$, $\xi = 2\delta.rand() - \delta$ with a is constant value.

The grey wolf algorithm is explained below

A. Initialization:

The population of grey wolves is 30
Maximum iteration number is 500
Dimension of the problem is set to 3 as there are three parameters to be tuned.
Lower bound of the parameters is set to `[30 2 0.005]`
and upper bound of the parameters is initialized to `[80 6 0.1]`
Lead1, lead2 and lead3 values are initialized to infinity.

B. Main loop

if iteration < maximum iteration.

C. Generate positions and evaluate the fitness

Generate random positions of grey wolves
Evaluate the fitness value of every grey wolf using (2)

$$\mathbb{F} = \int_0^{t_{sim}} (\Delta\omega)^2 d\tau \tag{2}$$

D. Update lead1, lead2 and lead3 values

If the fitness value of grey wolf is less than lead1, then lead1 value is updated to the fitness of the grey wolf, and record the position of Lead1.

If the fitness value of grey wolf is greater than lead1 and less than lead2, then lead2 value is updated to the fitness of the grey wolf, and record the position of lead2.

If the fitness value of grey wolf is greater than lead2 and less than lead3, then lead3 value is updated to the fitness of the grey wolf, and record the position of lead3.

E. Update the positions of lead1, lead2, lead3 and positions of grey wolf.

The positions of lead1, lead2 and lead3 are updated following Eq. (3). Let the updated positions of lead1, lead2, lead3 are XL1, XL2 and XL3, respectively. Then the updated position of grey wolf is given by (12)

$$X = \frac{X_{L1} + X_{L2} + X_{L3}}{3} \tag{3}$$

Increment the iteration value.
End of the main if loop.

The BLDC motor drive is simulated from the dynamic equations using MATLAB/Simulink. GWO algorithm is written in MATLAB file which runs BLDC Simulink file, and the fitness value is returned to main MATLAB file.

4 Hardware Implementation

Hardware implementation of the drive is as shown in Fig. 4, and the snap shot of the experimental test bed is shown in Fig. 3. It consists of inverter, BLDC motor and dSPACE DS 1103 controller board. The proposed work is simulated many times and obtained the suitable PID gains. Later these PID gains are used for testing in real-time speed control of BLDC motor [7].

The board has equipped with a slave DSP processor, TMS320F240. Now the already MATLAB/Simulink file is connected with RTI blocks of DS1103 and dumped into DSP enabled code. For more details of real-time implementation, one can read in [7].

5 Result Analysis

The hardware experimentations have been conducted to know the performance of the proposed algorithm. The closed-loop speed control using grey wolf algorithm is tested for two reference speed as test cases for the best PID gains with kp = 48.55, Ki = 11.31 and kd = 0.318. These cases opted to know the dynamic performance of the drive during acceleration and deceleration. The first case consists of triangular and ramp references as target speed. Figure 5a shows the performance of the closed-loop speed control using Ziegler–Nichols method, and Fig. 5b shows speed

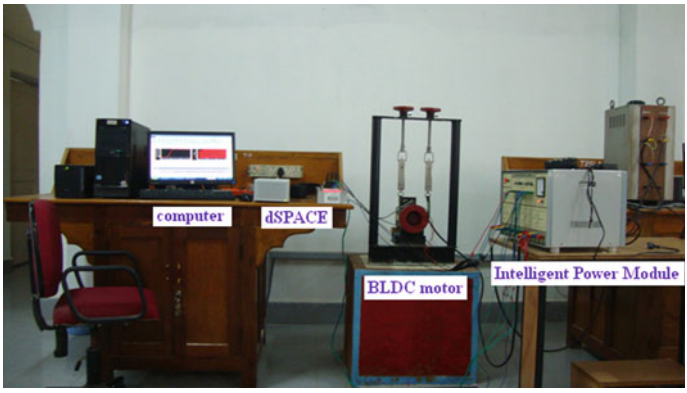


Fig. 3 Hardware implementation of the speed controller for BLDC motor

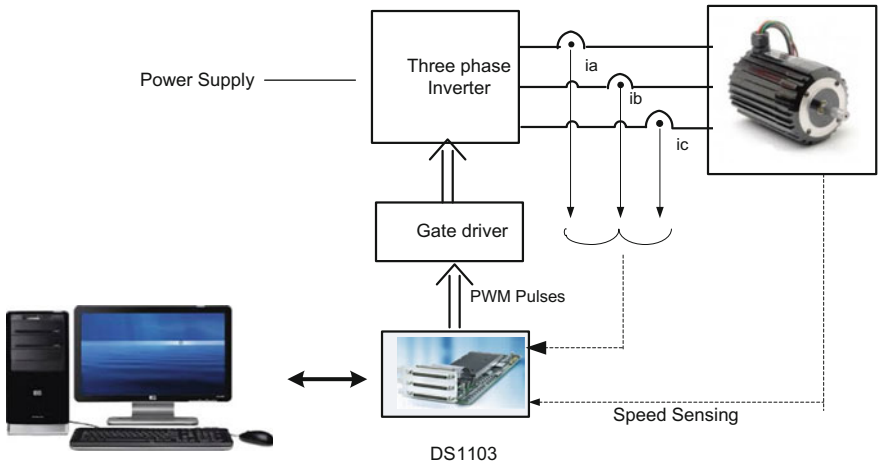
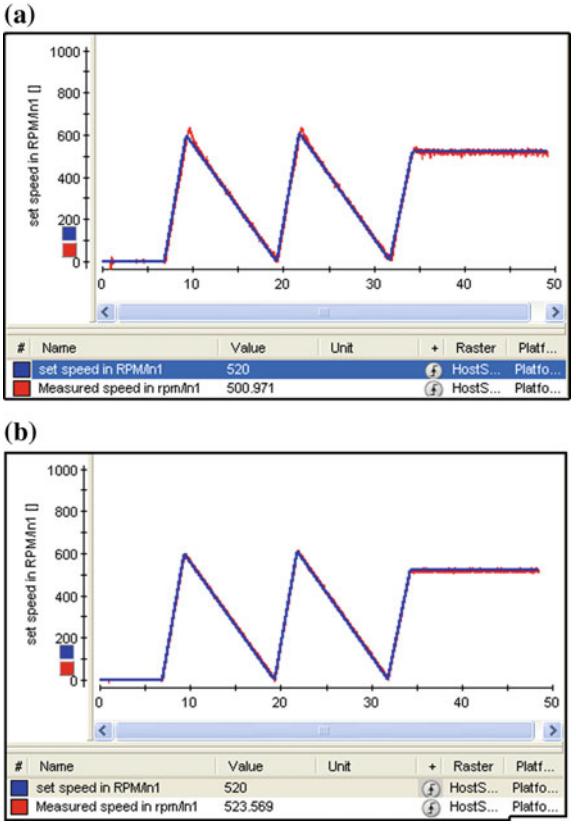


Fig. 4 Hardware implementation of the speed controller for BLDC motor

Fig. 5 **a** closed-loop speed control using ZN method. **b** closed-loop speed control using GWO method



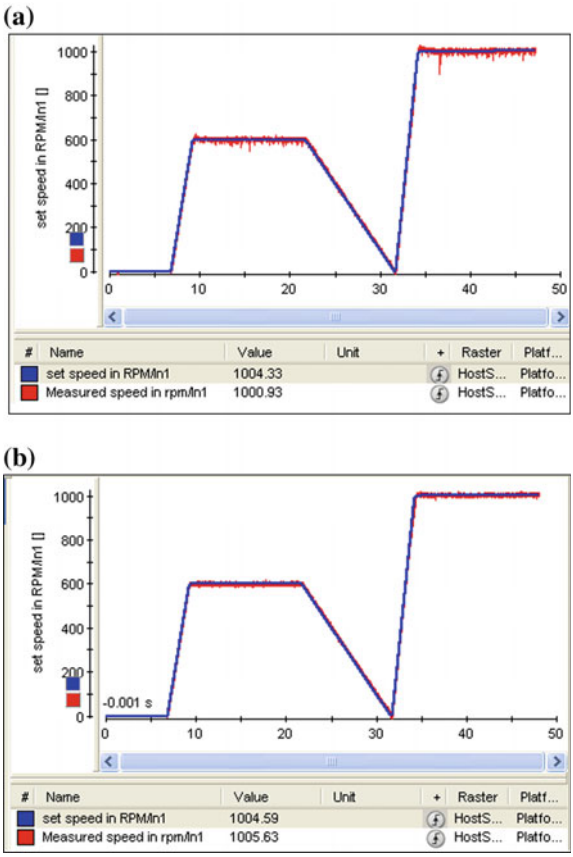


Fig. 6 **a** speed control of the BLDC motor drive with ZN method. **b** speed control of the BLDC motor drive with GFO method

control using GWO method. In GWO method, motor tracks the given target speed exactly and speed error is negligibly small.

In case 2, the reference speed is taken in a different way as shown in Fig. 6a, b, wherein Fig. 6a shows the speed control using ZN method and Fig. 6b shows the speed control using GFO algorithm. One can note that the closed-loop speed control performance during acceleration and deceleration and as well as at different speeds including low and high speeds.

6 Conclusions

The closed-loop speed control of BLDC motor has been implemented successfully by using the PID gains obtained using the grey wolf optimization technique and has been compared with conventional ZN method. The optimal developed system is tested under transient and steady-state conditions, and the experimental results

proved that the speed deviations are negligible for given reference speed, and closed-loop speed control is superior to the proposed approach.

References

1. Ahn, K.K., Truong, D.Q.: Online tuning fuzzy PID controller using robust extended Kalman filter. J. Process Control **19**(6), 1011–1023 (2009)
2. El-Wakeel, A.S., et al.: A hybrid bacterial foraging-particle swarm optimization technique for optimal tuning of proportional-integral-derivative controller of a permanent magnet brushless DC motor. Electr. Power Compon. Syst. **43**(3), 309–319 (2015)
3. Kumar, S., Henry, J.: A high performance direct torque control of PMBLDC motor using AI. J. Electr. Syst. **1**, 27–35 (2015)
4. Mirjalili, S., et al.: Grey wolf optimizer. Adv. Eng. Softw. **69**, 46–61 (2014)
5. Mirjalili, S., et al.: Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization. Expert Syst. Appl. **47**, 106–119 (2016)
6. Pillay, P., Krishnan, R.: Modeling, simulation, and analysis of permanent-magnet motor drives. II. The brushless DC motor drive. IEEE Trans. Ind. Appl. **25**(2), 274–279 (1989)
7. Potnuru, D. et al.: Design and implementation methodology for rapid control prototyping of closed loop speed control for BLDC motor. J. Electr. Syst. Inf. Technol. 4–16 (2017)
8. Premkumar, K., Manikandan, B.V: Neurocomputing adaptive neuro-fuzzy inference system based speed controller for brushless DC motor. Neurocomputing 1–11 (2014)

Energy-Aware Workflow Scheduling Algorithm for the Deployment of Scientific Workflows in Cloud

S. Balamurugan and S. Saraswathi

Abstract Recently, there is a fast increase in the use of cloud computing and its assets for the execution of scientific workflows [1] (Xu et al. IEEE Trans Cloud Comput 4(2);166–179, 2016). All the scientific applications are very broad in scale, so the fruitful execution of the application can be in fact finished by expanding in a cloud platform. A single scientific workflow more often contains a huge number of tasks; to execute all these tasks, it requires more number of resources. Only cloud infrastructure can provide such a huge number of computing resources. All the processing assets are given as virtual machines inside a cloud platform. More amount of computational energy is spent to execute complex scientific applications, so it becomes very critical to use the cloud virtual machines in an energy conservational way. So, using the cloud computing resources in more energy efficient manner has gained more attention among the analyst. But even today, executing a scientific workflow in an energy-aware manner in the cloud platforms is a challenging task. Considering all these issues, this paper proposes an energy-aware workflow scheduling with task migration (EAWSTM) algorithm which is proposed for the deployment of various scientific workflows in cloud platforms.

1 Introduction

Cloud is an Internet-based computing platform which provides us very large-scale processing modules which are distributed and located in a way such all the computing resources and its services are accessible to clients on the demand. The important idea behind the cloud computing is virtualization. By utilizing a virtualization innovation, we can ready to change over a solitary physical server to

S. Balamurugan (✉)
Perunthalaivar Kamarajar Institute of Engineering and Technology, Karaikal, India
e-mail: balaprof@gmail.com

S. Saraswathi
Pondicherry Engineering College, Puducherry, India
e-mail: swathi@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_15

run a few virtual machines all the while. Workflow is characterized as a gathering of tasks which are executed in a particular request to relating to any application. Scientific workflows allow the clients to effectively characterize multi-step computational assignments, for example, recovering information from any product or a database, reorganize the information, analyze information, and run. Scientific workflow processes have emerged to challenge the issue of inordinate multi-faceted nature in logical examinations and applications. Many number of tasks can be performed in a single workflow. Each task is in charge of a small piece of functionality with the end goal that many tasks should be anchored in a pipeline in an order to get a workflow that plays out some valuable assignment.

Any scientific workflow will have an overall deadline and also each task will have its own deadline called sub-deadline. With the substantial scale, usage of cloud data centers and computing servers, the amount of energy consumed has turned into a major test for the cloud specialist. So, using the cloud computing resources in more energy-efficient manner has gained more important amount the analyst in the field of cloud computing for deploying scientific workflows.

The remaining sections of the paper are organized as follows. Section 2 briefly discusses the existing energy-efficient scheduling algorithms that are commonly used for the execution of scientific workflows in cloud. Section 3 describes the energy model of cloud computing and its related equations for finding the energy consumption of a virtual machine. In Sect. 4, a new algorithm called energy-aware workflow scheduling with task migration (EAWSTM) is proposed and discussed. Section 5 presents the detailed illustration of the proposed work. Finally, Sect. 6 concludes the paper with the future work.

2 Related Works

This section discusses various energy-efficient workflow scheduling algorithms proposed by many researchers.

An energy-aware resource allocation method called EnReal [1] is proposed by Xiaolong Xu et al. to reduce the total energy consumption of any workflow. In this method, the virtual machines are deployed dynamically as on when the need arises for scientific workflow executions. This work focuses on lessening the energy utilization which incorporates the physical machine base-level energy utilization; the energy spent by all the idle virtual machines, the energy spent by all the idle physical machines (PM) and any data transfer operations.

An energy-efficient task scheduling algorithm [2] is proposed by Zhuo Tang et al. for executing the workflow in cloud where DVFS is enabled. To accomplish energy efficiency, this author proposed an algorithm called DVFS energy-efficient workflow task scheduling (DEWTS). In this algorithm, the generally imprudent virtual machine processors are merged by recovering the idle time. After few intermittent merges, it can use the useful slack time. DEWTS utilizes HEFT algorithm for mapping the task assigned to it. The test simulation result proves that

DEWTS has decreased the aggregate power utilization by up to 46.5% for different applications.

Huangke Chen et al. developed an energy-efficient online scheduling algorithm called ENOS [3] for real-time workflows. The computing resources can be scaled up and scaled down using the three policies proposed by the author. These policies can integrate into ENOS to balance weighted square frequencies of hosts.

Khadija Bousselmi et al. proposed an scheduling scheme called workflow partitioning for energy minimization (WPEM) [4] and cat swarm optimization (CSO) algorithm for the execution of scientific workflows. This algorithm lessens the total energy utilization of the workflow and the energy spent in data transfer parallel execution of tasks. This algorithm uses CSO schedule the created partitions such a way to limit the total energy utilization of the workflow and its execution time. The experimental outcomes demonstrate that their proposed algorithm has decreased the energy utilization with the tested workflows.

In [5], an energy-efficient task scheduling algorithm is proposed. This algorithm focuses on executing the workflow within a given deadline using frequency scaling technique. It is the job of ESFS (Stepwise Frequency Scaling) algorithm to find the frequency of the processor to use when the job is executed on that machine. Also the algorithm has to take care of the deadline of the task so that the deadlines are met. The algorithm constantly changes the processor frequency for each job in such a way that the overall energy consumption is reduced. The simulation results proved that there is reduced energy consumption when compared to other algorithms like HEFT and EES.

In [6], the author presented yet another Pareto-based multi-targeted workflow scheduling algorithm. It is the augmentation of prevalent best in class heuristic algorithms which are capable of computing a set of tradeoffs most favorable solutions in terms of energy efficiency and workflow makespan. This approach depends on experimental models which catch the real behavior of energy consumption in the heterogeneous parallel framework.

3 Energy Model of Cloud

Normally, cloud will offer different types of virtual machines with different computing performance. There are many characteristics of virtual machine; for example, computing performance denoted as C is calculated in million instructions per second, and maximum power is denoted as P. If a virtual machine has higher computing performance, then it consumes more CPU power, and it is given by Eq. 1.

$$\text{If } C_k > C_{k+1} \text{ then } P_k > P_{k+1} \tag{1}$$

The energy consumption of any virtual machine is of two types: static and dynamic energy consumption. Let E_{dyn} denotes the dynamic energy consumption

and E_{idle} denotes the idle state energy consumption. P_k denotes the power consumption of any virtual machine vm_k, and it is calculated by Eq. 2.

$$P_k = V_k^2 * f_k \tag{2}$$

Here, V_k is the voltage level of virtual machine vm_k and f_k is the execution frequency virtual machine vm_k. Now, the energy consumption can be defined as the product of power consumption of virtual machine and time. At any given time 't', the dynamic energy is calculated by Eq. 3.

$$E_{dyn} = P_k * t_e \tag{3}$$

Static energy consumption is calculated by Eq. 4.

$$E_{idle} = P_k * t_{idle} \tag{4}$$

Then, the total energy consumed by any virtual machine is given by Eq. 5.

$$E(vm_k) = E_{dyn} + E_{idle} \tag{5}$$

4 Proposed Work

In this section, energy-aware workflow scheduling with task migration (EAWSTM) Algorithm is proposed and the operational activities are discussed. This algorithm consists of two phases. In the first phase, each task is considered as independent and the tasks are selected and assigned to an optimal virtual machine based on the concept called power utility. In the second phase, identify all the task which can be merged and migrated and do task migration based on the criteria discussed in Sect. 4.2.

4.1 EAWSTM Task Assignment Scheme

Before proceeding with the algorithm, a concept called power utility is introduced. Power utility is defined as the ratio of workload to the total energy consumption of task t_i on virtual machine of type k. The task in a workflow is mapped to an appropriate virtual machine based on the concept of power utility [7]. Power utility of any virtual machine is given by Eq. 6.

$$PU\left(t_i^k\right) = \frac{W_i}{E\left(vm_k\right)} \tag{6}$$

Now, the power utility set called PU (t_i) for task t_i is formed which includes all the power utilities when that specific task is executed in different virtual machines. Likewise, for all the tasks present in the workflow, a power utility set is formed and it is represented as given in Eq. 7.

$$PU(t_i) = \{pu(t_i^1), pu(t_i^2), pu(t_i^3), \ldots\ldots\ldots, pu(t_i^j)\} \tag{7}$$

The power utility set of task t_i which has all the power utilities are arranged in decreasing order of their values and first value denotes the best power utility for executing the task t_i on the virtual machine and the corresponding virtual machine is denoted as vm_{opt}.

4.1.1 Implementation Steps

Step 1: For the given task t_i and if t_i can be finished within the deadline, then assign t_i to vm_{opt}. This means task t_i is mapped to vm_{opt}.
Step 2: If the feasibility for mapping the task, t_i to vm_{opt} is not possible. Then, find out the next optimal virtual machine from the power utility set. Find the all the tasks which are unable to assign and put in the new set called SUB_OPT_TASK_SET.
Step 3: For each task in the set SUB_OPT_TASK_SET, select the next optimal power utility from the power utility set. And set that virtual machine as vm_{opt}, map the task t_i to vm_{opt}, and remove the task from SUB_OPT_TASK_SET.

4.2 EAWSTM Task Migration Scheme

If two tasks t_i and t_{i+1} have already been mapped to Virtual machine of type vm_i and vm_{i+1}. If it is found that task t_{i+1} comes next to task t_i. Then, the energy consumed by the two tasks is denoted by E (t_i) and E (t_{i+1}). Then, these two tasks can be migrated to new virtual machine of type vm_k and the execution time of these two tasks can be computed using Eq. 8.

$$T(t_i t_{i+1}) = W_i + W_{i+1}/C_k + D_{Trans} \tag{8}$$

Here, D_{Trans} denotes the data transfer time, if two tasks are in the same virtual machine vm_k, we need not to consider the data transfer time. W_i and W_{i+1} denote the workload, and C_k denotes the computing performance.

From the task set, find if any task t_{i+n} can be finished within the deadline of task t_i as given in Eq. 9, if there exist any such two tasks then check whether the total

energy spent in combined execution of tasks t_i and t_{i+n} is less than the energy spent when the tasks are executed in two different virtual machines as given in Eq. 10. If both the criteria satisfied, then create a virtual machine of type k and migrate those two tasks to virtual machine k.

4.2.1 Implementation Steps

Step 1: Move all the tasks into the new set called task set.
Step 2: For each task in the TAST_SET, check whether the following conditions are met as given in Eqs. 9 and 10. If the conditions are satisfied, then perform task merging and migration; otherwise, the tasks are not migrated and removed from the task set. Repeat the above step till the task set is empty.

$$\text{Deadline}\left(t_{i+n}\right) < \text{Deadline}\left(t_i\right) \tag{9}$$

$$E\left(t^{k}_{i,i+n}\right) \leq E\left(t^{j}_{i}\right) + E\left(t^{j+m}_{i+n}\right) \tag{10}$$

5 Illustrations

Figure 1 illustrates the functionality first phase of EAWSTM [8] algorithm called task assignment phase. Let 'k' denotes the number of virtual machines and 'n' refers to the number of tasks extracted from the workflow DAG. For a given scientific workflow, first identify the set of tasks and create the necessary virtual machines (VM) based on the application given. Now, the task set and VM set are given to the power utilization module. The power utilization module calculates the power utility for each virtual machine based on Eq. 6 and a power utility set is formed for each task in the task list.

Based on the PU set, each task is assigned to an optimal virtual machine by the task assignment module. The first value in the power utility set denotes the best power utility for executing the task t_k on the virtual machine and the corresponding virtual machine is denoted as vm_{opt}. In this method, finding the optimal virtual machine for the execution of task is done prior to assigning the task. Prior estimation is done on the virtual machine based on the nature of the incoming task, specification of the virtual machine and historical data. So, this method can guarantee that the given task will be executed on the optimal virtual machine, so that the overall energy efficiency can be improved.

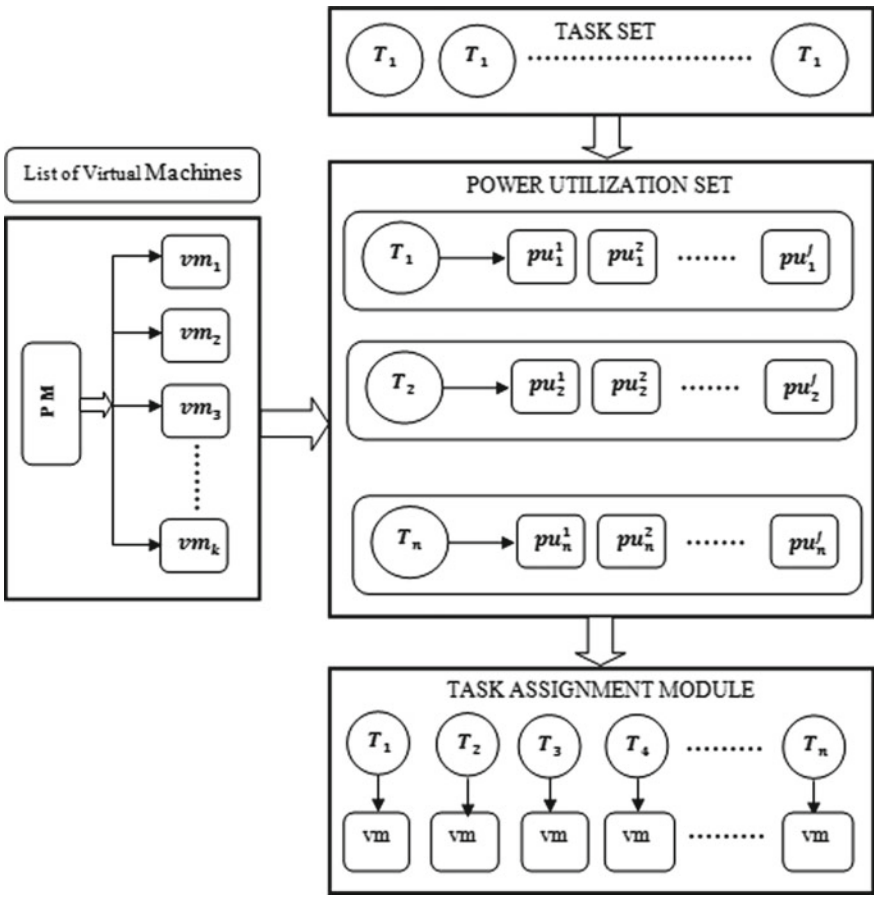


Fig. 1 Task assignment in EAWSTM

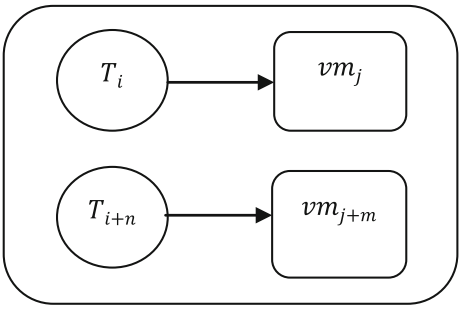


Fig. 2 Individual task running on VMs

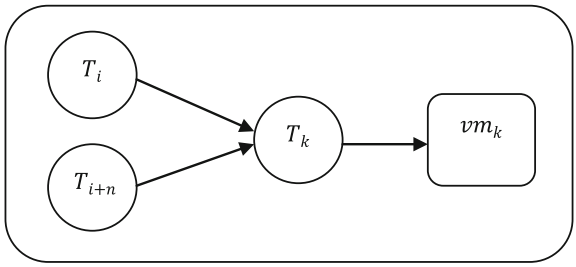


Fig. 3 Task merging and migration

After the completion of task assignment module, check whether there is any possibility for task migration. Figure 2 shows the task assignment for task t_i and t_{i+n} on vm_j and vm_{j+m}, respectively. Figure 3 shows the migration of two tasks based on the deadline and total energy.

From the TASK set find if any task t_{i+n} can be finished within the deadline of task t_i as given in Eq. 9, if there exists any such two tasks then check whether the total energy spent in combined execution of tasks t_i and t_{i+n} on any virtual machine of type k, is less than the energy spent when the tasks are executed in two different virtual machines as given in Eq. 10. If both the criteria are satisfied, then create a virtual machine of type k and migrate those two tasks to virtual machine k. If it is possible for us to merge and migrate more number of tasks onto a fewer virtual machines, we can able switch off one or more physical machines which do not have any virtual machines. So, it is possible to conserve more energy when we execute the workflow on cloud.

5.1 EAWSTM Algorithm

In this algorithm, the nature of the application selected is maintained as a DAX file. DAX [9] is the description of any abstract workflow in XML format and it all the information pertaining to the task nodes and its dependencies. The DAX file is parsed to get the task list and its sub-deadline. The input to the algorithm is DAX file, list of virtual machines as VM_LIST and TASK_SET generated by DAX Parser.

Algorithm: Energy Aware Workflow Scheduling with Task Migration (EAWSTM)

Inputs: 1. DAX File
2. List of Virtual Machines as VM_LIST

```
String daxFile = "<…/<DAXFile.xml>"
int k = <no. of Virtual Machines>
vmlist0 = createVM(k) // Create the list of virtual machines
VM_LIST = {vmlist0}
TASK_SET = daxParser (daxFile)
int noOfTasks = getSize(TASK_SET) // get the no. of tasks in the task set
for each in the TASK_SET do
    PU (t_i^k) = W_i / E (vm_k)
    //Arrive power Utility Set for each task
    for each in theTASK_SET do
        PU(t_i) = {pu(t_i^1), pu(t_i^2), pu(t_i^3), …… , pu(t_i^j)}
    end for
    for each element in PU(t_k) do
        SORT (PU(t_i))
    end for
    for each task in the TASK_SET do
        for each element in PU(t_i)
            if (free(pu(t_i^j)) // check if virtual machine is free
                vm_opt ← pu(t_i^j)  // mark as optimal virtual machine
                vm_opt ← t_i // map the task to the virtual machine
            end if
        end for
    end for
end for
for each in the TASK_SET do
    if (Deadline (t_i+n ) < Deadline (t_i) && E (t_i,i+n^k) ≤E (t_i^j) + E (t_i+n^j+m))
        t_k ← merge (t_i , t_i+n)  // merge two tasks as a single task
        vm_k ← t_k // map the merged task to a new virtual machine
        Update (t_k ,vm_k)  // update the information to the Power Utility set
    end if
end for
wfEngine.submitJob(VM_LIST, noOfTasks, TASK_SET, PU);
```

6 Conclusions

This paper focuses on the issue of higher energy utilization of virtual machines when a scientific workflow is executed in cloud. After the in-depth analysis of existing algorithm, a new algorithm called energy-aware workflow scheduling with task migration (EAWSTM) is proposed. In this algorithm, prior energy estimation is done for the virtual machine based on the nature of the incoming task, specification of the virtual machine, and historical data. Also, it is possible for us to merge and migrate more number of tasks onto a fewer virtual machines and idle physical machines which do not have any virtual machines are switched off. So, this method can guarantee that the workflow is executed on the optimal virtual machines to conserve more energy. In future, it is planned to simulate the algorithm using simulator called workflowSim and demonstrate the observed results for a variety of scientific applications.

References

1. Xu, X., Dou, W., Zhang, X., Chen, J.: EnReal: an energy-aware resource allocation method for scientific workflow executions in cloud environment. IEEE Trans. Cloud Comput. **4**(2), 166–179 (2016)
2. Tang, Z., Qi, L., Cheng, Z., Li, K., Khan, S.U., Li K.: An energy-efficient task scheduling algorithm in DVFS-enabled cloud environment. J. Grid Comput. Springer Sci. (2015). https://doi.org/10.1007/s10723-015-9334-y
3. Chen, H., Zhu, X., Qiu, D., Guo, H., Yang, L.T., Lu, P.: EONS: minimizing energy consumption for executing real-time workflows in virtualized cloud data centers. In: IEEE International Conference on Parallel Processing Workshops (2016). https://doi.org/10.1109/icppw.2016.60
4. Bousselmi, K., Brahmi, Z., Gammoudi, M.M.: Energy efficient partitioning and scheduling approach for scientific workflows in the cloud. In: IEEE International Conference on Services Computing (2016). https://doi.org/10.1109/scc.2016.26
5. Ilia, P., Sakellariou, R.: Energy-aware workflow scheduling using frequency scaling. In: Parallel Processing Workshops (ICCPW), 43rd International Conference, pp. 104–113. IEEE Press (2014)
6. Durillo, J.J., Nae, V., Prodan, R.: Multi-objective energy-efficient workflow scheduling using list-based heuristics. Future Gener. Comput. Syst. **36**, 221–236 (2014)
7. Li, H., Zhu, H., Ren, G., Wang, H., Zhang, H., Chen, L.: Energy aware scheduling of workflow in cloud center with deadline constraint. In: 12th International Conference on Computational Intelligence and Security, IEEE (2016)
8. Balamurugan, S., Saraswathi, S.: A model for energy efficient scheduling of scientific applications in cloud. Int. J. Electron. Electr. Comput. Syst. (IJEECS) **6**(8), 48–53 (2017)
9. Pegasus Documentation and workflow gallery. https://pegasus.isi.edu/workflow_gallery/

A Framework for Evaluating Performance of MADA-AODV Protocol by Considering Multi-dimensional Parameters on MANET

Thati Balamuralikrishna and Mohammed Ali Hussain

Abstract The study is purely related to the problem of energy management in mobile ad hoc network (MANET). The augmentation in MANET can optimize the performance of the devices and minimize the consumption of the power of the battery. It is correct that the nodes shape such set of connection is mobile, consequently no instant renew of battery is essential. Energy-aware design and evaluation of network procedure require knowledge of the energy consumption behavior of actual wireless interfaces. Even if lots of power-aware routing scheme have been proposed for wireless ad hoc network, it is not suitable for network with assorted power supplies, here nodes will run by battery or connected to the key lattice network. This paper investigates energy-aware routing algorithms for energy optimizing technique in the ad hoc networks. The metrics are evaluated by varying packet dropping reliability, end-to-end delay, throughput, renovied energy, and total energy with different input parameters taken for AODV [1] and MADA-AODV [2] algorithms, for number of nodes and probability of packets processing locally on various performance metrics are discussed. In this paper, we furthermore present outcome by showing the saved power consumption at most maximum possible in the same network when a superlative power control plan is functional.

T. Balamuralikrishna (✉)
Department of Computer Science, Bharathiar University, Coimbatore, India
e-mail: balu.thati9@gmail.com

M. A. Hussain
Department of Electronic Computer Science & Engineering, KL University, Vijayawada, India
e-mail: dralihussain@kluniversity.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_16

1 Introduction

Regularly, the directional antennas are traditionally worn to increase the capability of the network in cellular wireless networks. So energy-aware steering is the successful method to extend the life point of strength-conditioned node points in wireless ad hoc networks. While transmitting packets from source node to destination node, route paths are identified by taking into account the power charge. The resultant could produce knowing routes by putting away a very small amount of energy for forwarding packets or nodes in the routes may have more remaining battery power. Due to so many constraints in research positions, we can use wireless ad hoc networks. We can aim for reducing energy consumption or achieve self-configuration and self-organization because there are various proposed protocols in these complex networks. In addition, these networks may have applications in real time by reaching sent messages. Here, all the movable nodes (devices) are mostly reliant on power of the battery. Almost all researches of power utilization methods have focused wireless networks which are prearranged nearby base stations and central servers, which do not have limitations and are connected with little and portable or little nodes (devices).

2 Related Research

The packet radio networks were initially projected to require directional transmissions; in this place, a slotted acknowledgment packet radio network was measured. Afterward, varieties of studies were taken place on coming up with Macintosh protocols in mobile unexpected networks for mistreatment of sectorized and beamforming directional antennas. The swap over of request-to-send and clear-to send packets for protocols bestowed were to impede near nodes from transmittal in several directions once that will interfere with a supposed directional information transmission whereas permitting them to transmit on alternative directions.

Before, the convention portrayed in the present paper was dealt with as virtual transporter detecting by utilizing the RTS and CTS trade and executed the undertaking on a part premise fitting customary omnidirectional premise [10]. This system portion vector (NAV) is kept up in each part independently and furthermore permits prompt transmission of control bundles on those divisions. These segments are perfect and safe in the transmission. The more propelled shaft shaping and receiving wires by supplanting with sectorized directional radio wires in specially appointed systems in supportive for the advancement [11]. The creator puts stock in the earlier learning of area data and utilizes directional transmission of TRS and CTS bundles. The examination displays the crash of two shirking plans as one is more effective than the other to square transmission from posting hubs that accomplish a RTS or a CTS bundle. The idea of directional NAV is additionally acknowledged at whatever point pillar shaping reception apparatuses are utilized [12].

Heuristic methodologies are a unit to be used to understand problematic answers for troublesome problems with expansive size. Elie Wiesel et al. delineate a helpful calculation known as broadcast improved power (BIP). During this calculation, new gadgets area unit added to the tree utilizing a base monetary value heuristic [3]. Imprints displayed a transformative approach utilizing hereditary calculations along with techniques for making introductory arrangements [6]. Das et al. planned AN insect state framework approach and a close-by hunt heuristic known as r-contract system for enhancing arrangements got utilizing fast problematic calculations in remote systems, for instance, (Broadcast Improved Power) BIP.

As a matter of reality, a mobile ad hoc network is a combination of nodes which share communication together by firmly a multi-hop radio network [4]. These nodes move and the network topology can be dynamic. Due to this, there is no centralized administration. In this process, any node can connect to the network and disconnect itself at any time. In general, the battery cannot be recharged so as to guaranty the mobility of nodes [5]. Any node with an enhanced battery is treated as dead and remained useless. In recent times, efficient routing method with the help of energy-efficient parameter has been presented by the present author [7, 8, 9].

Bhaskar S has explained AODV [1]; as a matter of fact, reactive protocols attempt to find a way to establish routes on-demand. In the process of transmission, when a node desires to start communication with another node that has no route, the routing protocol will attempt to initiate the route. The task of AODV [1] is to find out the way in which constituent parts are interrelated or arranged and transmit the intended message from one node to another on-demand.

Ali Hussain and the Author stated MADA-AODV [2]. In the process of transmission, the source node begins to identify the route to the destination node. At the same time to rectify this problem, any intermediate node will be sent a routing request message to measure the current speed of the node by using its current and earlier location subsequently the node will solve the neighbor's problem by counting the neighbors. If the condition is failed, the node will not forward the packet. Therefore, the source node will get a path through intermediate nodes to the destination node. Finally, the nodes are removed from the route which is having more speed. Now the recognized paths in MADA-AODV can be constant when we compare the routes in common AODV.

3 Model Description

Using the network simulator tool a network is established, and source and destination nodes are identified and started propagating data packets through routes as per the two techniques that we are comparing. Those are AODV and MADA-AODV protocols. The typical RREQ propagation between source and destination (in both AODV and MADA-AODV) is explained in Fig. 1.

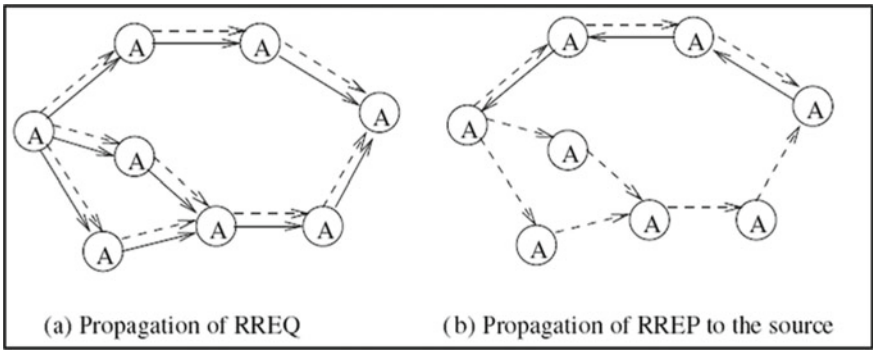


Fig. 1 Broadcast of RREQ and RREP in both AODV and MADA-AODV

In the constructed network, the battery power is compared with mains power while data packets are passed through one established root from basis to the target. Here, the power consumption is observed in the simulation environment by measuring the packet dropping reliability, end-to-end delay, throughput. We obviously get renovied energy and total energy. The comparison of battery power between nodes is shown in Fig. 2.

Here, six completely different ways of power consumption and network routing protocols area unit are studied. The primary one is concerning IEEE Mackintosh protocol. It is to use carrier-sense multiple transmission to collision rejection with the power to interchange the RTS and CTS management packets. The second is concerning medium access management victimization directional antennas. It is

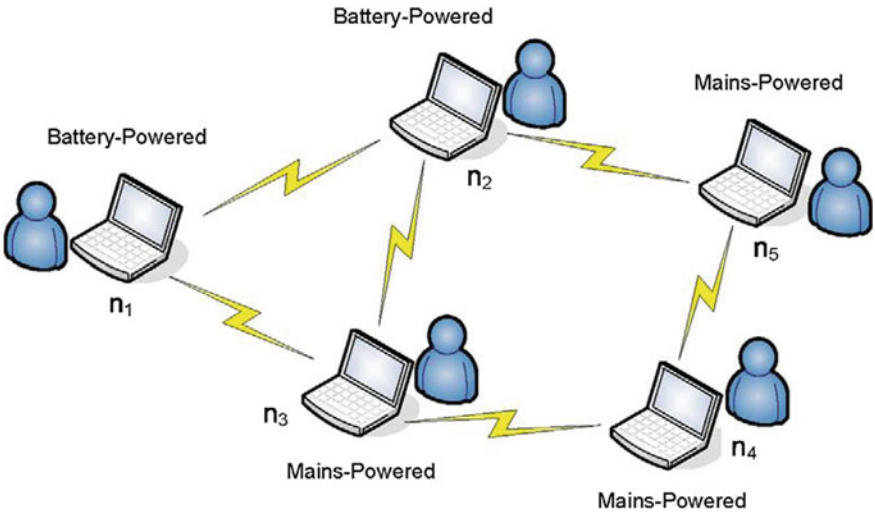


Fig. 2 Comparison of battery power between nodes

applied at the mobile nodes to realize two of the many edges reminiscent of the next outturn saving within the power consumption one explained in elaborated analysis. The third one is minimum energy routing; it is truly the routing metric and is utilized to reduce the energy consumption within the routing [14]. That is, the consumption of energy by per hop is minimized, similarly the consumption power needed by each node within the route is commencing to the ending purpose [15]. Here, the whole power level for route is measured because the ad of all power levels from setting out to finish the purpose within the route [13]. The fourth one is capacity-aware routing; here, we tend to take into consideration that each one node within the network area unit equally vital, no node ought to be used for routing additional typically than alternative nodes. Therefore, the remaining battery capability of a node needed to define a routing metric that captures the expected amount of your time of a node, and so, the amount of time of the network. The fifth one is AODV which is mobile unexpected networks, and it is a topology and is just transmitted by nodes on-demand [1]. Once a node needs to transmit traffic to a bunch of routes and it will generate a route request (RREQ) message that may be flooded into an exceedingly restricted things to alternative nodes. The sixth one is MADA-AODV which is a protocol in mobile unexpected networks, and it is the method implemented and involves two phases [16]. In the first phase, the evaluation of performance of routing protocols are explored to the possibility of improving one of the selected protocol. In the second phase, the aim is designing a new and efficient algorithm that is mobility- and density-aware routing protocol. Here, we considered two protocols AODV and MADA-AODV [2] for comparing with five different parameters as we said in the above methodology. Distances of any two nodes are

$$P_r = P_t G_t G_r \lambda^2 / (4\pi)^2 d^2 L. \tag{1}$$

P_r Receiving power
P_t Transmitting power,
G_t Gain of transmitting antenna
G_r Gain of receiving antenna
λ is wavelength
L System loss factor which has nothing to do with the transmission
d Distance between antennas.

Distance between two nodes that are equipped for transmitting antennas is defined as

$$D = \sqrt{\frac{P_t G_t G_r \lambda^2}{(4\pi)^2 L P_r}}. \tag{2}$$

The power consumed by node to transmit a packet is defined as

$$\mathrm{ECTij} = \left(\mathrm{Pt} + \frac{\mathrm{Pij}}{\mathrm{K}}\right)\mathrm{T} \tag{3}$$

The energy consumed by node to receive a packet is defined as

$$\mathrm{ECRij} = (\mathrm{Pr})\mathrm{T} \tag{4}$$

ECTij Energy consumed by a transmitting node from source i to the destination j.
Pt The power required to transmit a packet in the wireless interface from source node.
Pij The transmission power from node i to the node j.
K The power efficiency of power amplifier of transmitter.
T The time required to transmit a packet.
Pr The power required to receive a packet in the wireless interface to destination node

3.1 Algorithm for Evaluating the MADA-AODV Protocol

Step 1: For each source node i to destination j of a path in the MANET.
Step 2: Transmit packet by calculating the distance between one-hop nodes first and the two-hop nodes using (1).
Step 3: Calculate distance between nodes that are equipped by transmitting antennas using.
While sending packet of size L bits from source i to destination j in a time T using (2).
Step 4: Measure power consumed by transmitting node using (3).
For every source node i via neighbor node to destination node j;
Step 5: Similarly measure power consumed by receiving node using (4).
For every source node i via neighbor node to destination node j;
Step 6: Finally, estimate throughput and end-to-end delay in a network.

4 Presentation Valuation of the Planned Algorithm

The presentation of the planned map-reading algorithms is evaluated by imagining the nodes are shared equally in the network. The stretches are generated between chosen basis and target nodes in random-based widening spread intervals of time with its suggested value. Every node may have recognized a number of sessions to special targets, or it can be alive the target for number of stretches at the equal point. We compare various algorithms to reduce the variability. We suppose that basis nodes produce information packets with stable rate packets, and nodes have the

same original battery energy. Every time in our simulation, outcome is carried by captivating the moderate above standards secured in simulation runs. For every simulation run, the stretches are produced at random while for every random production of network as well. Table 1 contains the default values of the simulation parameters of the ns-2 tool environment.

The following two scenarios are considered and taken print from NS-2 tool, Fig. 3 is one screenshot in the beginning of the network construction, and Fig. 4 is another screenshot after completion of the network simulation model.

The outcomes of a range of presentation benchmarks that is to say packet dropping reliability, end-to-end delay, throughput, renovied energy, and total energy with different input parameters taken for AODV AND MADO = AODV algorithms with values and charts measured through NS-2 tool.

Packet Dropping Reliability

Here, we considered two algorithms: one is the popular one AODV and the second new algorithm is MADA-AODV to measure the packet dropping reliability using simulator. The comparison results are as mentioned as in Table 2.

In Table 2, if we observe, the MADA-AODV is showing better PDR and same is represented in Fig. 5 to have better view.

Table 1 Default values of the simulation parameters

Parameters	Values
Tool used	NS-2.35 V
Simulation area	1000 m/1000 m
Simulation time	300 s
Protocols used	AODV, MADA-AODV
Energy levels	1, 2, 3, 4, 5
Number of nodes	20
Mobility model	Random way

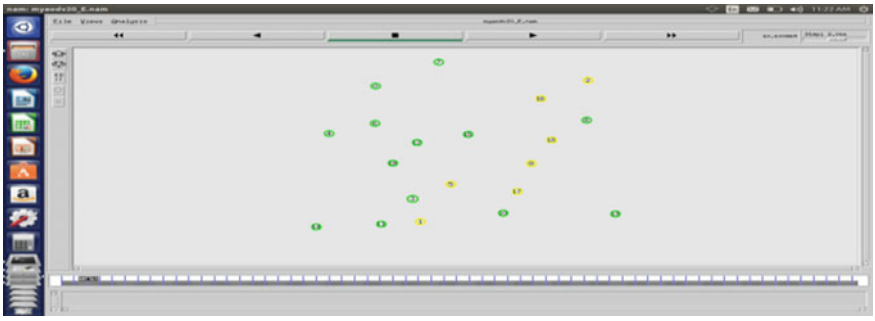


Fig. 3 Before the operation of the simulation with an initial node 20, certain speed, and path time

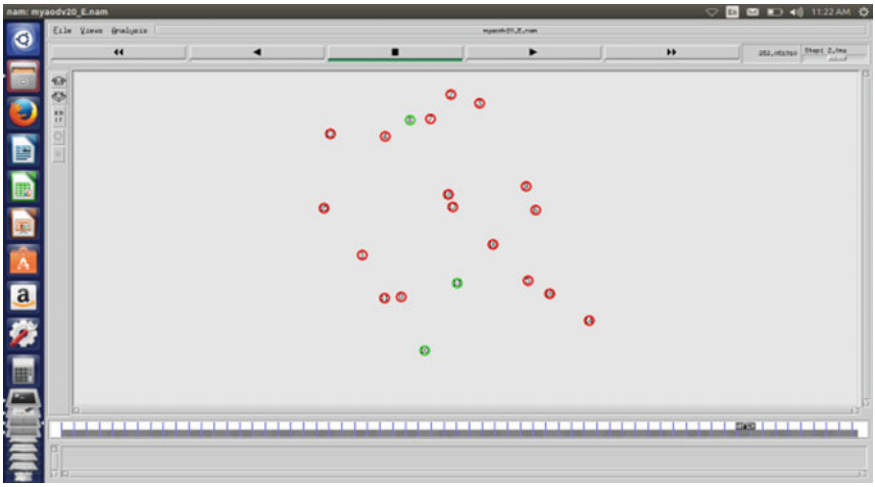


Fig. 4 After the operation of the simulation, the position of nodes with calculation of link breakage

Table 2 Comparison of AODV and MADA-AODV for PDR

Initial energy	AODV	MADA-AODV
1	90.1018	97.561
2	63.5223	96.8982
3	61.3371	85.1562
4	61.7451	70.7547
5	60.9001	61.5979

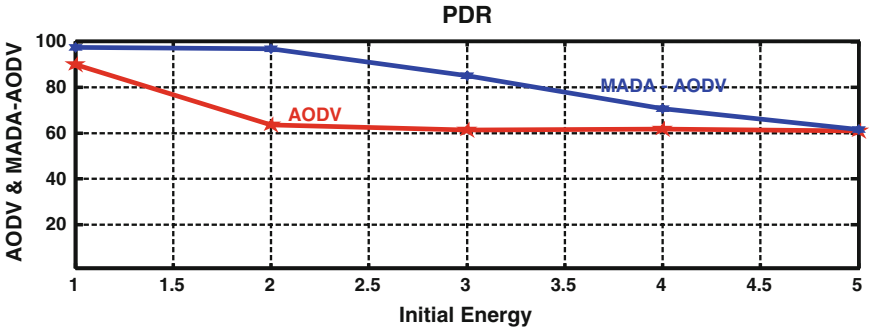


Fig. 5 MADA-AODV improved packet dropping reliability

End-to-End Delay

Here, we consider two algorithms one a popular one AODV and second new algorithm MADA-AODV to measure the end-to-end delay using simulator. The comparison results are mentioned as in Table 3.

In Table 3, if we observe, the MADA-AODV is showing reduced end-to-end delay and same is represented in the Fig. 6 to have better view.

Throughput

Here, we considered two algorithms one a popular one AODV and second new algorithm MADA-AODV to measure the throughput of using simulator. The comparison results are as mentioned as in Table 4.

In Table 4, if we observe, the MADA-AODV is showing better throughput and same is represented in Fig. 7 to have better view.

Renovied Energy

In this case too, we compare the two algorithms using simulator. Here also MADA-AODV shows better results, i.e., saved renovied energy. The results are mentioned in Table 5.

Table 3 Comparison of AODV and MADA-AODV for end-to-end delay

Initial energy	AODV	MADA-AODV
1	0.03715	0.0182
2	0.03686	0.0119
3	0.56724	0.0122
4	0.60035	0.088
5	0.80038	0.012

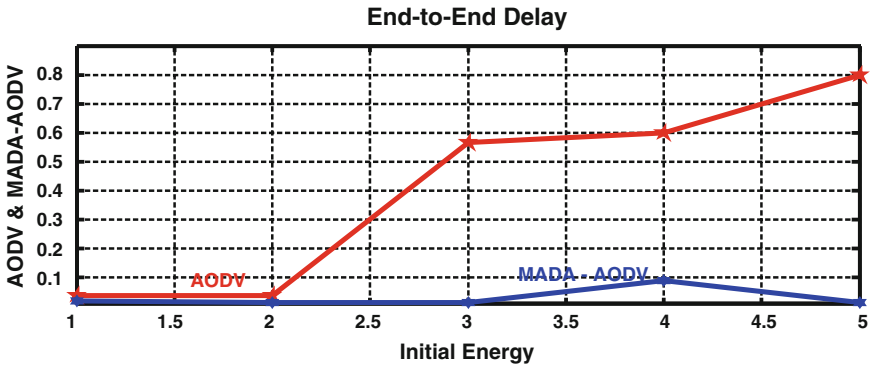


Fig. 6 MADA-AODV reduced end-to-end delay

Table 4 Improved throughput of MADA-AODV

Initial energy	AODV	MADA-AODV
1	8052.4	11826.9
2	12787.7	14935.1
3	18072.5	23515.3
4	24311.3	23613.5
5	28483.4	29669.4

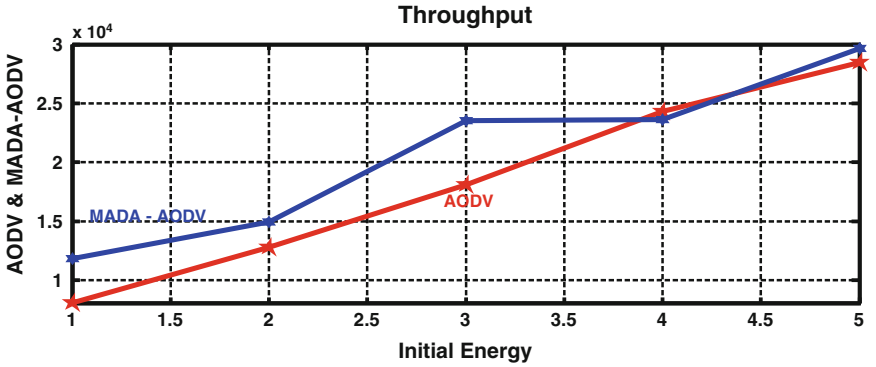


Fig. 7 MADA-AODV improved throughput

Table 5 Better results of MADA-AODV using simulator

Initial energy	AODV	MADA-AODV
1	2.2279	1.9563
2	3.4182	3.2703
3	3.8693	2.8837
4	3.6338	3.6222
5	3.2603	2.1022

In Table 5, if we observe, the MADA-AODV is showing saved renovied energy and same is represented in Fig. 8 to have better view.

Total Energy

In this case too, we compare the two algorithms using simulator, Here also MADA-AODV shows better results, i.e., saved total energy. The results are mentioned in Table 6.

In Table 6 if we observe, the MADA-AODV is showing saved total energy put and same is represented in Fig. 9 to have better view.

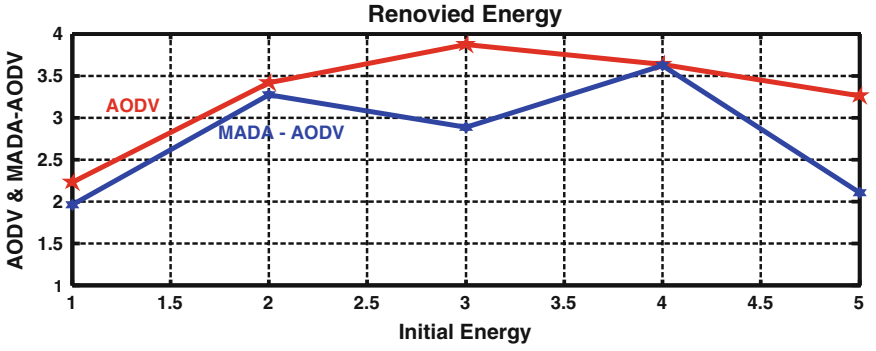


Fig. 8 Saved energy graphically in MADA-AODV

Table 6 How MADA-AODV saves the energy

Initial energy	AODV	MADA-AODV
1	13.2236	12.9838
2	29.6715	23.0148
3	48.6562	47.1892
4	67.3917	50.2277
5	86.5435	84.9718

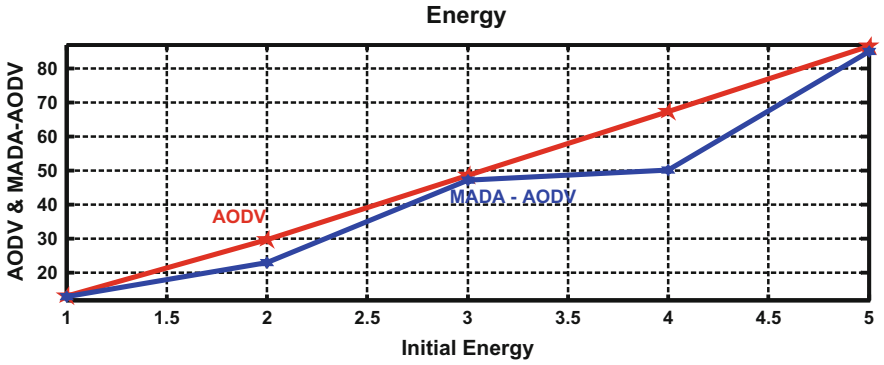


Fig. 9 MADA-AODV gives better result by saving the total energy

5 Conclusion

The study has proposed an algorithm and recent techniques over AODV to improve the packet dropping reliability and how to reduce end-to-end delay, improved throughput, and energy saving mechanism by using MADA-AODV protocol. A MADA-AODV protocol tackles the number of nodes and possibility of packets dispensation in the neighborhood on various performance metrics and also solves the trouble in the networks that is of harmonizing energy management with communication and quality.

This study has proposed various types of protocols that manage power management in routing for networks. The planned protocol able to identify power management efficient routes dynamically by directing the traffic to undead nodes of the network to avoid they find energy-efficient routes which dynamically direct the traffic to mains powered nodes of the network in order to avoid relaying over battery-powered nodes.

The route selection in this framework is a criteria decision-making problem. Reducing the power management and improving the quality packet delivery with optimal price. The AODV by the side of with the planned MADA-AODV protocol with mobility and node density is providing better performance with lesser maintenance in almost all the cases.

References

1. Balamuralikrishna, T., Hussain, M.A.: A stochastic analytical model for evaluating routing performance of AODV protocol for MANETs. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **6** (3) (2016). ISSN: 2277 128X
2. Balamuralikrishna, T., Hussain, M.A.: Design and development of an efficient routing protocol for mobile ad hoc networks. Int. J. Emerg. Eng. Res. Technol. **3**(4(A)) (2015)
3. Schiex, T., Fargier, H., Verfaillie, G.: Valued constraint satisfaction problems: hard and easy problems. In: Proceedings of IJCAI'95, Montréal, Canada. 1995 march; 4(1), 4202–4212
4. Rishiwal, V., Yadav, M., Verma, S.: Power aware routing to support real time traffic in mobile adhoc networks. In: Proceedings of 1st International Conference on Emerging Trends in Engineering and Technology. 2008 July; 1(1), pp. 206–217
5. Nagy, A., El-Kadi, A., Mikhail, M.: Swarm congestion and power aware routing protocol for manets. In: Proceedings of 6th Annual Communication Networks and Services Research Conference. 2008 May; 6(3), 1802–1817
6. Brougham, N., Song, Y.: A new routing metric for satisfying both energy and delay constraints in wireless sensor networks. J. Signal Process. Syst. **51**(2), 137–143 (2008)
7. Heni, M., Bouallegue, R.: Power control in reactive routing protocol for mobile ad hocnetwork. Int. J. Wirel. Mobile Netw. (IJWMN) **3**(3), 220–232 (2012)
8. Nasipuri, A., Li, K., Sappidi, U.R.: Power consumption and throughput in mobile AdHoc networks using directional antennas. In: Proceedings of the IEEE International Conference on Computer Communication and Networks (IC3 N'02). November; 8(5), 331–342 (2002)
9. Energy-aware routing algorithms for wireless ad hoc networks with heterogeneous power supplies Javad Vazifehdan Computer Networks. **55**(2), 3256–3274 (2011)
10. Abdellah Idrissi, How To Minimize The Energy Saad.T.H. E.Fadil: Queuing Approach to Estimate the MANET's Optimal Number of Nodes. Oriental Journal of Computer Science & Technology **5**(2), 205–214 (2012)
11. Madhavi, G., Kaushik, M.K.: Queuing Methodology Based Power Efficient Routing protocol for Reliable Data Communications in Manets **4**(1), 277–289 (2013)
12. Bisnik, N., Abouzeid, A.A.: Queuing network models for delay analysis of multi hop wireless ad hoc networks. Adhoc Netw. (Elsevier) **2**(2), 79–97 (2009)
13. Gaurav, K., Prasanna, G., Hota, Ch.: Probabilistic routing using queuing theory for MANETs. Int. J. Wirel. Mobile Netw. **3**(4), 144–158 (2011)
14. Vinoth Kumar, K., Bhavani, S.: An efficient secured localization based optimized energy routing for MANET 2Indian. J. Sci. Technol. 8(35) (2015)
15. Sumathi, A., Vinayaga Sundaram, B.: An ANN approach in ensuring CIA triangle using an energy based secured protocol E-AODV for enhancing the performance in MANETS. Indian J. Sci. Technol. **8**(34),17485- IPL0821 (2015)
16. Balamurugan, N.M., Applaud Alias Balamurugan, S.: Performance analysis of AD-HOC on-demand distance vector routing protocol. Indian J. Sci. Technol. (2016)

A Comparative Analysis of Breast Cancer Data Set Using Different Classification Methods

M. Navya Sri, J. S. V. S. Hari Priyanka, D. Sailaja and M. Ramakrishna Murthy

Abstract Patterns or models in data can be found using data mining algorithms. This is a knowledge discovery process in which data mining is involved. It is a scientific method which is intended to examine massive data, so as to find out the systematic relationships and consistent patterns among variables and further check for the accuracy of the findings. This can be done by taking new subsets of data and applying the detected patterns to them. The core part of the data mining techniques is classification. In classification, in order to develop a model which will categorize the population of records, we make use of a set of pre-classified examples. The techniques of classification use the model which is built on basis of training data and apply it to test data. "Breast cancer Wisconsin data set is used as a training set." There is an open source data mining tool named WEKA, which consists of implementation of data mining algorithms. By making use of WEKA we have compared the well-known classification algorithms that are decision tree and Bayesian algorithms. It is concluded that decision tree classification algorithm got high accuracy compared to Bayesian classification algorithm.

M. N. Sri (✉) · J. S. V. S. H. Priyanka · D. Sailaja · M. Ramakrishna Murthy
Department of Information Technology, Anil Neerukonda Institute of Technology and Sciences, Sangivalasa, Bheemunipatnam, Visakhapatnam
Andhra Pradesh, India
e-mail: navyasrimullapudi@gmail.com

J. S. V. S. H. Priyanka
e-mail: priyapatnaik.hari@gmail.com

D. Sailaja
e-mail: dadla.sailaja25@gmail.com

M. Ramakrishna Murthy
e-mail: ramakrishna.malla@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_17

1 Introduction

In this new era, we have data everywhere. Data are the facts which are collected for analysis. In the past twenty years, a massive enlarges in the amount of data being accumulated. The user aims to have more sophisticated data. This rises up and about the demand of data mining. Data mining refers to dig out or "mining" knowledge from huge amount of data [1]. The purpose of data mining is to concentrate on mining of information from bulk collection of data and make over it into an effortless understandable formation for further purpose. The various methods of data mining are clustering, classification, association rules, prediction. In the course of classification algorithms, the well-known classification algorithms are decision tree classification and Bayesian classification. Data mining has a wide range of applications in the real life; those are customer segmentation, financial banking, criminal investigation, research analysis, bioinformatics, fraud detection, intrusion detection, and customer relationship management [2]. Improvement of health systems is one of the significant applications of data mining. It uses data and analytics to recognize superlative practices that reduce expenses and develop care. In urbanized nations, breast cancer has turned into the crucial reason of death in females. Breast cancer is a heterogeneous disease to one in every eight women which affects during their lives. It forms in the cells and develops from breast tissues with "high degree of diversity between and within tumor's as well as among cancer-bearing individuals" being a factor together to identify the risk of disease progression and therapeutic resistance [3]. Accurate prediction of breast cancer cells at earliest possible could help in decreasing the cost of health and enhances the time that is required for a patient to take the required treatment as well. Therefore, this paper presents an advanced approach which allows analyzing and classifying tumors at an unprecedented depth. A lot of research has been made for becoming aware of survivability of cancer. In this paper, several research mechanisms carried out using Bayesian classification and decision tree classification. These algorithms play an important role to identify the patient functioning.

The enduring of this manuscript is ordered as follows. Section 2 is regarding the survey of the authors experimented with various data mining techniques that are applied to the examination of breast cancer. The classification algorithms are used for the analysis of breast cancer and the methodology are explained in Sect. 3. Section 4 shows the experimental results of comparative study between decision tree and Bayesian classification. The entire survey work is concluded in Sect. 5.

The remaining portions of this paper are structured as follows: Section 2 specifies the literature survey. Section 3 specifies different classification algorithm. Section 4 provides the outcomes of this experiment. Section 5 concluded and mentions the future work of this paper. Finally, lists the references used in this paper.

2 Literature Survey

According to B. Padmapriya [4], a comparative analysis has been made between ID3 and C4.5 classification algorithms to detect breast cancer in female and proved that C4.5 classification algorithm gives efficient results to diagnose breast cancer compare to ID3 classification algorithm. They used SEER data set for breast cancer analysis and discussed the various applications of breast cancer applications. Rostemmennor [5] discovered a drug to use machine learning algorithms for big data analytics on top of map reduce for the breast cancer protein receptor. In this, they have concentrated on protein of breast cancer, so they have selected 4JLU receptor which is a critical structure. Chandresharya [6] discussed different expert systems designed for last twenty years. Especially, these systems are used to detect breast cancer tumor in human. They experimented on mammogram images to find out the abnormality in patient. In this, Wisconsin breast cancer data set is used and experimented with different classification approaches; those are support vector machine algorithm, neural network algorithm. Xinagchumet all [7] analyzed various statistical procedures to detect breast cancer tumor in women. In this research work also mammogram images are applied and got the precision from 68 to 79%. In this, a statistical technique that is principal component analysis is used to identify abnormalities in patient.

As per the above, a lot of research work is accomplished on the detection of breast cancer tumor in women. This paper provides the best algorithm and its accuracy to find out the abnormality tumor in women. In this, the comparative analysis has been made between the well-known classification algorithms which are the decision tree and Bayesian algorithms.

3 Classification Algorithms

A progression of building a class model from a data set which encloses class labels is described as classification. In this, the well-known UCI machine learning repository that is breast cancer Wisconsin (Diagnostic) data set is used for experimentation. This data set features are extracted from digitized representation of a fine needle aspirate (FNA) of a breast dimension. These extracted features are applied as input for both decision tree and Bayesian classification algorithms.

3.1 Decision Tree Classification (J48)

A researcher named J. Ross Quinlan in 1980 build up an algorithm known as decision tree classification also known as ID3 (Iterative Dichotomiser 3) [8]. An extension of ID3 algorithm is j48 implemented by the WEKA project team. This

algorithm builds the classification model in tree structure form. The data set is divided into subsets up to leaf nodes. The leaf nodes of a tree are the decisions or class labels. This algorithm is implemented as j48 in WEKA tool. The following steps are fundamental steps of this algorithm which are (i) the instances which belong to same class of the tree labeled with the same class name. (ii) For every attribute, the gain (potential information) has to be calculated, and (iii) Calculate entropy which is a measure of the data disorder. In the data sets, a number of instances are not well defined from remaining instances. So to decrease the classification errors the pruning is applied for reducing classification errors [9]. This is one of the reasons that decision tree algorithm gives better accuracy because of pruning technique. After pruning applied, then the data set is given as input for decision tree algorithm, so that the classification errors can be reduced.

3.2 Bayesian Classification

It is a statistical and supervised learning method for training the data sets. It is a probabilistic model to determine the likelihood of outcomes. This algorithm is used to solve diagnostic health diseases [10]. This algorithm is proposed on Bayes theorem. It provides the prior knowledge to assess many challenges. Many algorithms are not capable of classifying noise data or null data exist in the data set, but even though some data sets have noise data, Bayes algorithm is robust to train noise in input instances [11]. To train the text documents in sentiment analysis or customer feedback analysis, Bayesian algorithm is one of the best-known algorithms to classify the text documents. In this paper, we selected Bayes algorithm because of its robust nature.

4 Experimental Results

The experiments are conducted on WEKA tool. The input breast cancer data set is given as input for decision tree and Bayesian classification algorithms. This data set has 32 attributes with the number instances 286. The characteristic of this data set is multivariate.

Methodologies:

1. Correctly classified instances (CCI): Correctly classified instances show the accuracy of decision tree classification algorithm compared to Bayesian.
2. Kappa statistic (KS): It computes to examine the same data by two independent tools.
3. Mean absolute error (MAE): It is a quantity used to measure how close estimates or predictions are to the eventual outcomes.

4. Root mean square error (RMS): It is one of the most widely used statistics in GIS. We use RMSE in a variety of applications when we need to compare two data sets.
5. Relative absolute error: It is calculated as the mean absolute error separated by the error of the ZeroR classifier.
6. Root relative squared error: It is considered in relation to what it would have been if a simple predictor had been used. More absolutely, this simple predictor is just the average of the actual values. Thus, the error takes the total squared error and normalizes it by separated by the total squared error of the simple predictor. By considering the square root of the relative squared error one makes smaller the error to the same dimensions as the quantity being predicted.

4.1 Performance Evaluation of Decision Tree Classification

Correctly classified instances (CCI): 217
Incorrectly classified instances (ICI): 69
Kappa statistic (KS): 0.2899
Mean absolute error (MAE): 0.3658
Root mean squared error (RMS): 0.4269
Relative absolute error (RAE): 87.4491
Root relative squared error (RRS): 93.4017
Total number of instances (TI): 286

4.1.1 Confusion Matrix of Decision Tree Classification

Classified as	A	B
non-recurrence events	194	7
recurrence events	62	23

4.2 Performance Evaluation of Bayesian Classification

Correctly classified instances (CCI): 214
Incorrectly classified instances (ICI): 71
Kappa statistic (KS): 0.3693
Mean absolute error (MAE): 0.3012
Root mean squared error (RMS): 0.4278
Relative absolute error (RAE): 72.0082
Root relative squared error (RRS): 93.6095
Total number of instances (TI):286

4.2.1 Confusion Matrix of Bayesian Classification

Classified as	A	B
non-recurrence events	174	27
recurrence events	44	41

4.3 Comparative Analysis

	CCI	ICI	KS	MAE	RMS	RAE	RRS	TI
J48	75.875%	24.12%	0.2899	0.3658	0.4269	87.4491	93.4017	286
Bayesian	75.17%	24.82%	0.3693	0.3012	0.4278	72.0082	93.6095	286

4.3.1 Detailed Accuracy by Class

	True positive rate	False positive rate	Precision	Recall	F-measure	Roc area
J48	0.759	0.523	0.76	0.759	0.716	0.639
Bayesian	0.752	0.404	0.74	0.752	0.743	0.76

The decision tree classification algorithm identified correctly 217 instances, whereas Bayesian classification identified 214 classified instances. Particularly, mean absolute error is the probability estimates which is high for decision tree compared to Bayesian classification. From the confusion matrix, we can say that one instance of class is recurrence event and other is non-recurrence events. In decision trees confusion matrix, 194 instances belong to class non-recurrence event and in Bayesian confusion matrix, 174 instances belong to non-recurrence event.

5 Conclusion and Future Work

Various techniques are reviewed on breast cancer tumor problems in this study. This scenario is mostly investigated under decision tree and Bayesian classification algorithms and decision tree got high accuracy in assessment to Bayesian. A diverse research has been done on breast cancer diagnosis, and it is notified that source and symptoms associated to each occurrence with the evidence correlated to each patient so that up to some extent it is possible to prevent from the cancer. The executions of both algorithms are investigated on WEKA tool. In future, the expert systems of breast cancer may be useful for accomplishing high classification rate.

References

1. Tomar, D., Agarwal, S.: A survey on data mining approaches for healthcare. Int. J. Bio-Sci. Bio-Technol. **5**(5), 241–266 (2013)
2. Sujatha, G., Usha Rani, K.: A survey on effectiveness of data mining techniques on cancer data sets. Int. J. Eng. Sci. Res. **04**(1), 1298–1304 (2013)
3. Utomo, C.P., Kardiana, A., Yuliwulandari, R.: Breast cancer diagnosis using artificial neural networks with extreme learning techniques. Int. J. Adv. Res. Artif. Intell. **3**(7) (2014). http://dx.doi.org/10.14569/IJARAI.2014.030703
4. Padmapriya, B.: A survey on breast cancer analysis using data mining techniques. IEEE. ISBN No: 978-1-4799-3975-6, 2014
5. Mennour, R.: Drug discovery for breast cancer based on big data analytics techniques. IEEE (2015)
6. Arya, C.: Expert system for breast cancer diagnosis: a survey. ICCCI. ISBN No: 978-1-4673-6680-9 (2016)
7. Xiangchun, K.X.: Analysis of breast cancer using data mining & statistical techniques. IEEE. ISBN No: 0-7695-2294-7 (2005)
8. Sivagami, P.: Supervised learning approach for breast cancer classification. Int. J. Emerg. Trends Technol. Comput. Sci. **1**(4), 115–129 (2012)
9. Rajesh, K., Anand, S.: Analysis of SEER dataset for breast cancer diagnosis using C4.5 classification algorithm. Int. J. Adv. Res. Comput. Commun. Eng. **1**(2), 72–77 (2012)
10. Babagholami-Mohamadabadi, B., Jourabloo, A., Zarghami, A., Kasaei, S.: A bayesian framework for sparse representation-based 3-d human pose estimation. IEEE Signal Process. Lett. **21**(3), 297–300 (2014)
11. Ramakrishna Murty, M., Murthy, J.V.R., Prasad Reddy, P.V.G.D: Text document classification based on a least square support vector machines with singular value decomposition. Int. J. Comput. Appl. (IJCA) (indexed by DOAJ, Informatics, ProQuest CSA research database) **27**(7), 21–26 (2011). ISBN 978-93-80864-56-6, https://doi.org/10.5120/3312-4540

Energy Efficient Forwarding Data and Route Selection (EEFDRS) Algorithm for Wireless Body Area Network

G. Smilarubavathy and S. Ayyasamy

Abstract The major healthcare applications in WBAN is monitoring the human bodies, sensing the data from the human bodies, and collecting the information from it. The body sensors have the restricted energy resource. So its energy exhaustions going to cause a severe network ruin like throughput, latency, and energy efficiency. The various amounts of energy consumption rate occurred due to the various body sensors. Also the topological variations in sensor path are more often render to the routing networks which are unfeasible. This repeated network failure gives the harmful effect on network path selecting ability to upkeep quality of service-driven services. To crack the above-mentioned problem, this paper proposed EEFDRS strategy. This strategy helps to improve the lifetime of the sensor and also balancing the energy consumption and select the best routing path for data transmission. This paper includes the major contributions of (1) to reduce the size of physiological data, compressed sensing technology is used, and the data packets are to be forwarded (2) EEFDRS route selection algorithm is proposed for the best stable, and optimal path for data transmission is selected. In addition, EEFDRS improves reliability and reduce traffic delay.

1 Introduction

In wireless body area network, the biosensors have the abilities to monitor and process the biological data. These sensors are wearable or injected into the human body, which bring together the physiological data such as EEG, blood pressure, temperature, body pressure which is shown in Fig. 1 and Table 1 Wireless body

G. Smilarubavathy (✉) · S. Ayyasamy
Computer Science and Engineering, Coimbatore, India
e-mail: smilarupa@gmail.com

S. Ayyasamy
e-mail: ayyasamyphd@gmail.com

G. Smilarubavathy · S. Ayyasamy
Dr.N.G.P Institute of Technology, Coimbatore, Tamil Nadu, India

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_18

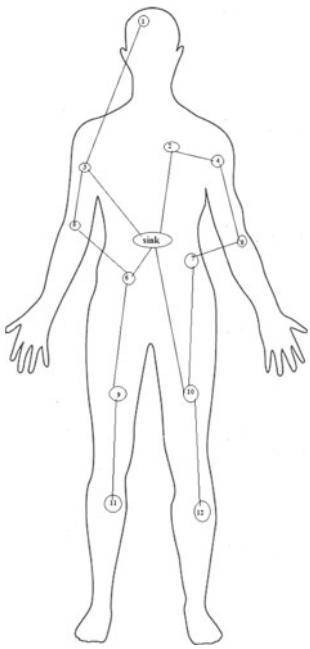


Fig. 1 Topology of WBAN sensors

Table 1 WBAN sensor location and its types

Sensor ID	Location	Type of sensor
1	Head	EEG sensor
2	Heart	ECG sensor
3	Left shoulder	Temperature sensor
4	Right shoulder	Motion sensor
5	Left elbow	Blood pressure sensor
6	Abdomen	Insulin sensor
7	Hip	Urea sensor
8	Right elbow	Pulse rate sensor
9	Left thigh	Blood oxygen level sensor
10	Right Thigh	EMG sensor
11	Right knee	Motion sensor
12	Left knee	Motion sensor

area sensors are mostly depending on the resource and energy [1], moreover, it is very difficult to change the sensor.

Wireless body area networks are usually small, and their intake of energy during the data transmission is primarily determined by the relay sensor's selection and the data transmission size. Data forwarding plays the main part of energy consumption in sensors [2, 3]. Hence, efficiently minimize the data size to be forwarded to maximize the energy efficiency of sensors. In the real-time monitoring healthcare applications using WBAN, the moments of the body cause the shadow effect.

It happens in frequently varying topologies of the network and the status of wireless link. The above mechanism of classical routing for static network topology is unfeasible.

By using the sparsity of sampled signals [4], the original data size should be in a lower dimensional space which can be minimized by using the CS theory. Also, It gets the various diverse sample from the constrict signal to recuperate the original signal with nonlinear algorithms [5]. The data collection from the WBANs sensors is sparse which is specified by the many researchers. In WBAN, using the CS technology can minimize the size of the data. The CS technology's computational complexity at the source sensors is relatively less, which has the controlled computational complexity of the sensors in the WBAN.

Though the above-cited algorithms can minimize the transferred data effectively, the transmission of data in wireless body area network can be attained in a synergistic manner [6] with the cost of energy in relay sensors. The energy utilization of relay sensors is extraordinarily high in the synergistic manner [7]. Due to the body sensors heterogeneity [8], the rate of the energy consumption differs from every node. The initial and remaining energy cannot exactly represent the heterogeneity of the sensor at the outcome [9].

Depending upon the above analysis, the relay sensors have to be selected properly for the data forwarding process of WBAN; hence, the intake of energy is to be balanced to prolong the lifetime of the network. The neighbor sensor's failure detection dynamically varies the topology of the network and changing the relay sensor position gives interrupted communication between the sensors and the loss of data packet [10].

To solve the above problems, this paper proposed an energy efficient forwarding data and route selection algorithm (EEFDRS). When the physiological data are acquired, WBAN sensors employ the binary sparse random measurement matrix [9] and DCT [11] to process data size reduction from the original biological data. Depending on the frequency of sampling data, the intake of energy rate can be decided [12]. While the data transmission path selection process, the rate of energy consumption of biosensors proportion of remaining energy and the foremost degrees of collected data are used to select relay sensor; hence, it achieves the multi-hop transmission of data with balanced energy consumption [1] and to improve the lifetime of the network.

2 System Model

The design of the energy efficient forwarding data and route selection mechanism is very complex in WBAN. To design the WBAN of particular application goals, and to attain the transmission of data in WBAN, the system structure is shown in Fig. 1.

In this specific WBAN, various types of data are collected by the multiple biosensors and transmit the data packets to the other biosensors; finally, the sink receives all the data packets. The data transmission from sensor to sink occurs

within several hops, and the distant biosensors are propoerly selected in the relay for data broadcast.

The architecture of WBAN [10] is shown in Fig. 2. Here the data forwarding path from the sensor node to the server and caregiver are clearly mentioned.

The WBAN sensor's energy consumption mainly consists of three parts as shown in Eq. (1) where E_{tot} indicates the total consumption of energy in biosensors, E_{tra}, E_{rec} and E_{pro} indicate the energy transmitting, processing, and receiving data packets [13]

$$E_{tot} = E_{tra} + E_{rec} + E_{pro} \tag{1}$$

Let the energy consumption for data transmission for each bit of data is E_{ene}, and the amplifying the consumption of energy in circuit is E_{ampl}, the length of data packet is L, the loss coefficient of transmission power is β, and the transmission distance of the data is D. The total energy consumption is calculated by Eq. (2).

$$\begin{aligned} E_{rec} &= L \times E_{ene} \\ E_{tot} &= L \times (E_{ene} + E_{ampl} D^{\beta}) \end{aligned} \tag{2}$$

The compressed sensing technology employing the computational cost, namely the energy consumption in the processing of data, is unavoidable in WBAN; hence, Eq. (3) calculates the intake of energy in transmitting, receiving, and processing of the L data bits is [13].

$$E_{tot} = L \times (E_{ene} + E_{ampl} D^{\beta}) + E_{pro} \tag{3}$$

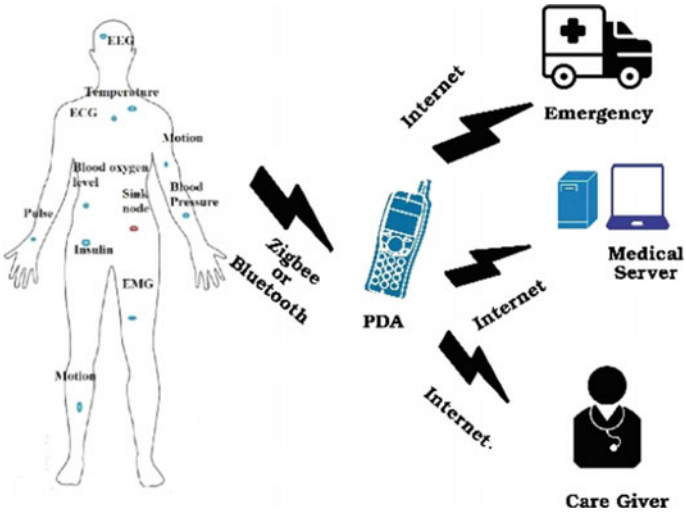


Fig. 2 Architecture of WBAN

3 Data Collection in Compressed Sensing Technology

In WBAN, transmitting all the original physiological data leads to the high consumption of energy and congestions of network. To solve this issue, CS technology is used to reduce the size of the original data at the body sensors. By using the sampling frequency [9, 11], and its value should be lesser than the Nyquist frequency, sparse vectors represent the physiological original data to reduce the transmitted data size and also for consumption of its energy. This sparse vector can get back its original data.

The original data are represented by $z = [z_1, z_2, \ldots\ldots z_n]$ with the support of basis vector $z_{i,}$, (i == 1,2,….,n) is shown in Eq. (4), where the coefficient vector is $u = [u_1, u_2, \ldots\ldots u_n]^T$, and the sparse basis is the γ. And when $\|u\|_0 == p$, where in z, the basis γ, consist the sparsity is p.

$$Z^T_{n\times 1} = \sum_{i=1}^{n} u_i \gamma_i = \gamma_n \times n^u \tag{4}$$

The measurement matrix is chiefly measured by $\psi = [\psi_1, \psi_2, \ldots \psi_n]$ where the lower dimensional space can be reduced by the z, which can get through the n dimensional measurement of data, as shown in Eq. (5), further, it is transferred into the network.

$$y_{N\times 1} = \psi_{N\times n} \times_z T_{n\times 1} = \psi \times \gamma \times u = \Theta \times u \tag{5}$$

The compressed sensing technology is the process to allow the sparsity of original data. In Eq. (5), p consists of $p(p\langle\langle n)$ which has a nonzero value, and the sparsity of p is u.

If matrix Θ fulfills the condition of Eq. (6), N measurements reconstruct the p coefficients accurately. And any p-signal sparsity and $\delta_p \in (0, 1)$ is named as restricted isometry property.

$$1 - \delta_p \le \frac{||\Theta z||_2^2}{||z||_2^2} \le 1 + \delta_p \tag{6}$$

To obtain the original biological data, the original data with the specific sparse random binary measurement matrix was first measured by biosensors, where there are two major challenges; they are to derive the sparsity of the original biological data and the measurement matrix design.

First, ψ the compressed sensing matrix must be independent of γ sparse basis matrix; from that condition, it can able to fulfill the restricted isometry property. The CS technology employs the Gaussian random, Toeplitz, and Gaussian random matrices. Besides, the binary random sparse matrix is employed in this paper which

has to simplify the design of the algorithm and also minimize the difficulty of the implementation.

To minimize the energy intake in WBAN, this active matrix used to convert the matrix addition (MA) from the matrix multiplication (MM). Secondly, the pre-condition is the original biological data sparsity of employing compressed sensing technology. In this paper, DCT is also used as the basis of sparse because of the promising sparse performance representation and fast running speed [9].

4 Selection of Data Forwarding Path

In the initial stage, to allocate the impact factor like α_i, β_i, χ_i, and residual energy threshold E_φ, an initialization message of the network is broadcasted by the sink for every biosensor. From the sink node, every biosensor has the capacity to measure its distance which can be done by the location information [14] inbuilt in the sensor node. Especially, after the deployment process, the sensor position on the human body referred the information about location and can be able to attain manual or camera measurement.

While the data collecting phase is in the progress after it completes the local processing, sensor i send the data forwarding message to the wireless body area network, it contains the sensor location information and ID of the sensor i. After receiving the message, it calculates the residual energy based on Eq. 7, where E_{ini} refers the initial energy. This value compares with E_φ. If the predetermined threshold is more than the sensor j residual energy $E_\varphi \geq E_{res}$, the forwarding request is ignored by sensor j through sensor i, or the forwarding request can accept by sensor j. and transmits the data to sensor i.

$$E_{res} = E_{ini} - E_{tot}^{i} \tag{7}$$

The body sensor's set has the ability to accept the data forward request to sensor i and can be mentioned by $H_i = 1,2,\ldots.n$, and their decision values $Decision_j$ can be measured by these sensors as shown in Eq. (11) based on impact factor on sampling frequency τ_j, impact factor of residual energy σ_j, and the impact factor of importance of degree ϖ_i. This τ_j is the reciprocal of sampling frequency as shown in Eq. (8) where sampling frequency of the sensor i is noted as s_i. This σ_j is calculated in Eq. (9). This importance of degree is measured in Eq. (10).

$$\tau_j = \frac{1/s_i}{\sum_{i=1}^{n} 1/s_i} \tag{8}$$

$$\sigma_j = \frac{E_{resi}}{E_{ini}}$$

$$\sum_{i=1}^{n} \varpi_i = 1 \tag{10}$$

$$Decision_j = \tau_j \times \sigma_j \times \varpi_i \tag{11}$$

The reply message with feedback sent from body sensors within H_i to sensor i, which also consists of $Decision_j$, location information and sensor ID. When the sink, sensor j, and sensor i meet in Eq. 12 called as sensor i which is between the sink and sensor j, the data can be forwarded by the sensor i for sensor j.

$$\begin{aligned} dis_{ji} &< dis_{is} \\ dis_{is} &< dis_{js} \end{aligned} \tag{12}$$

The above equation dis_{js} and dis_{is} determined the distance travel to sink from sensor i and j, respectively. From the relay candidates sets $H_i^{'}$. The sensor selects the maximum decision value in $H_i^{'} = 1^{'}, 2^{'}, \ldots . n^{'}$ which acts like a relay sensor. The sink can straightly transmit the data when $H_i^{'}$ is empty.

4.1 EEFDRS Algorithm

The algorithm is used for energy efficient transmission path selection. In this route selection, artificial bee colony algorithm is also used. The summary of the entire algorithm process is given below.

```
Input: Number of Sensor nodes
Output: Energy Efficient path selection RS
Process:
        Collect the physiological data z from sensors
        Calculate the basis vector γ, coefficient vector u
        Measurement matrix ψ compressed to lower
dimensional space n
        Restricted Isometry property has to be applied
        Data compressed by Compressed Sensing Technology
        Construct the Routing Table R
        for(i=1;i<R;i++)
           Route_i =R.nextHop
        POP_i = Route_i
        For each node n in POP_i
        Do
           Find degree of node and its signal strength
           Bufferlength.B
        fit= node.ComputeFitness(B,S,D)
        Aggregate Fit to path as FitPathValue
RS = ArtificialBeeColony.Optimize POP_1,
Return RS
```

5 Energy Efficiency Analysis

The energy efficiency of the WBAN sensor can be analyzed based on the parameters of energy consumption per bit E_{con}, E_{ene}, E_{ampl}, and β. The two types of posture are chosen; they are walking and sitting, and the two different sensor ID is chosen; they are ECG sensor and blood oxygen level sensor which are shown in Figs. 3 and 5. In this analysis, the energy consumption rate of sensor 2 and 9 are totally different, because of the variation of sampling frequencies. In this postures, at the initial stage, sensor 9 serves as relay; hence, the energy consumption is higher than the sensor 2. Along with its own data, sensor 9 forwards the data of neighboring node. When the neighboring node reaches the predetermined threshold value, it stops transmitting the data packets.

Similarly, posture standing and posture lying are analyzed which are shown in Figs. 4 and 6. From the above analysis based on the various postures, the consumption of energy is prolonged by using EEFDRS algorithm and the energy is also preserved. With the help of EEFDRS, number of retransmission of data is also reduced. Data packet loss and delay are also reduced.

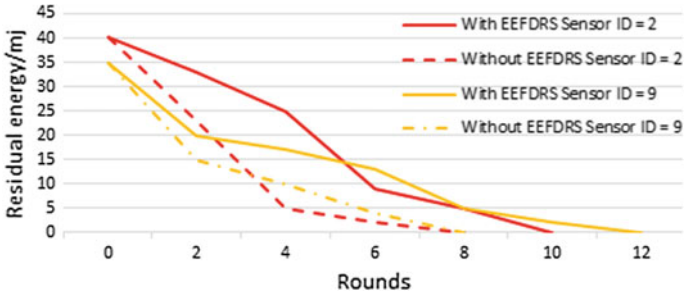


Fig. 3 Posture—walking

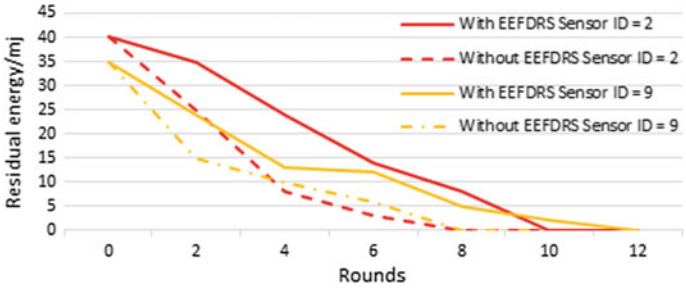


Fig. 4 Posture—standing

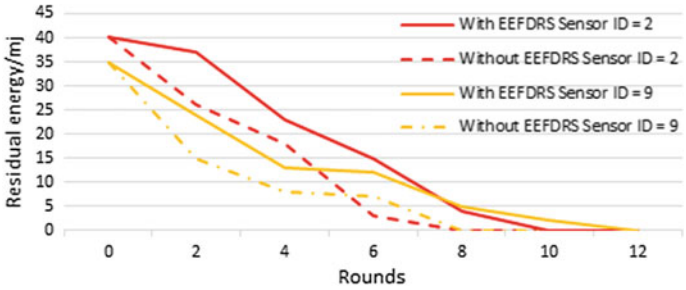


Fig. 5 Posture—sitting

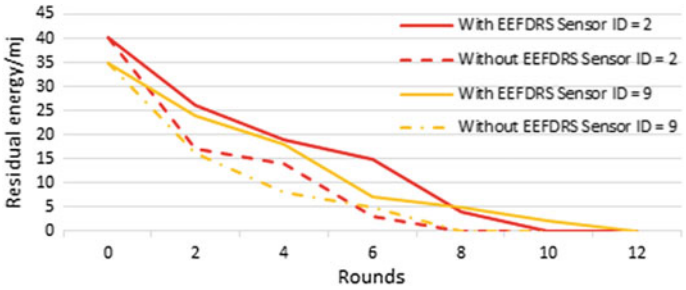


Fig. 6 Posture—lying

6 Conclusion

The energy resource in WBAN is very limited; thus, energy depletion is the major problem. So the EEFDRS was proposed to improve the energy efficiency, deals with the topological changes in network, and extends the lifetime of the sensors. The proposed work also shows the optimal and reliable path for data packet forwarding from sensor to sink node. Our future studies are to make the cost-effective energy-efficient data transmission

References

1. Ge, Y., Liang, L., Ni, W., Wai, A.A.P., Feng, G.: A measurement study and implication for architecture design in wireless body area networks. In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PERCOM) Workshops, pp. 799–804 (2012)
2. Mamaghanian, H., Khaled, N., Atienza, D., Vandergheynst, P.: Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes. IEEE Trans. Biomed. Eng. **58**(9), 2456–2466 (2011)

3. Wu, D., He, J., Wang, H., Wang, R.: A hierarchical packet forwarding mechanism for energy harvesting wireless sensor networks. In: IEEE Comput. Mag. **53**(8), 92–98 (2015)
4. Xie, S., Wang, Y.: Construction of tree network with limited delivery latency in homogeneous wireless sensor networks. Wirel. Pers. Commun. **78**, 231–246 (2014)
5. Zhang, X., Wu, H., Ma, Y.: A new auto-focus measure based on medium frequency discrete cosine transform filtering. Appl. Comput. Harmon. Anal. **40**, pp. 430–437 (2016)
6. Nidhya, R., Arunachalamand V.P., Karthik, S.: A study on requirements, challenges and applications of wireless body area network. Asian J. Electr. Sci. **6**(2), pp. 30–36 (2017)
7. Liang et al., X.: Exploiting prediction to enable secure and reliable routing in wireless body area networks. In: Proceedings of the IEEE INFOCOM, pp. 388–396 (2012)
8. Liang, L., Ge, Y., Wai, A.A.P.: A low overhead treebased energy-effcient routing scheme for multi-hop wireless body area networks. Comput. Netw. **70**(18), pp. 45–58 (2014)
9. Candès, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal Process. Mag. **25**(2), 21–30 (2008)
10. Movassaghi, S., Abolhasan, M., Jamalipour, A.: Wireless body area networks: a survey. IEEE Comput. Surv. Tuts. **16**(3), 1658–1686 (2014)
11. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
12. Nidhya, R., Arunachalam, V.P. Dr., Karthik, S. Dr.: Multitier security architecture with energy requirement for medical wireless sensor network. Int. J. Print. Packag. Allied Sci. **4**(5), 3885– 3897 (2016)
13. Dapeng, W.U., et al.: An energy-efficient data forwarding strategy for heterogeneous WBANs. IEEE ACCESS **4**, 7251–7261 (2016)
14. Lin, C.S., Chuang, P.J.: Energy-efficient two-hop extension protocol for wireless body area networks. IET Wirel. Sens. Syst. **3**(1), 37–56 (2013)

Skin Melanoma Assessment Using Kapur's Entropy and Level Set—A Study with Bat Algorithm

V. Rajinikanth, Suresh Chandra Satapathy, Nilanjan Dey, Steven Lawrence Fernandes and K. Suresh Manic

Abstract Skin melanoma is considered as a deadliest form of skin malformation originates in human community. Due to its increasing incidence rates, it is necessary to build an accompanying procedure to assist the clinical detection and diagnosis process. Visual examination and the digital dermoscopy are the two common procedures widely adopted by the doctors to detect and verify skin melanoma. This paper proposes a soft-computing assisted tool to investigate the skin melanoma images. In this work, bat algorithm-assisted Kapur's multithresholding is considered to preprocess the image, and the level set-based segmentation is adopted in the postprocessing stage to mine the skin melanoma section. The experimental investigation is implemented using the benchmark DERMIS dataset. The effectiveness of proposed technique is confirmed by measuring the familiar image similarity measures through a relative study among extracted skin melanoma with the ground truth. The experimental result verifies that the proposed technique is

V. Rajinikanth (✉)
Department of Electronics and Instrumentation, St. Joseph's College of Engineering, Chennai 600119, Tamil Nadu, India
e-mail: rajinikanthv@stjosephs.ac.in

S. C. Satapathy
School of Computer Engineering, Kalinga Institute of Industrial Technology (Deemed to be University), Bhubaneswar 751024, Odisha, India

N. Dey
Department of Information Technology, Techno India College of Technology, Kolkata 700156, India

S. L. Fernandes
Department of Electronics & Communication Engineering, Sahyadri College of Engineering & Management, Mangalore 575007, India

K. S. Manic
Department of Electrical & Computer Engineering, Caledonian College of Engineering, University College, Sultanate of Oman, Muscat, Oman

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_19

easy to implement and offers superior values of Jaccard (0.8805), Dice (0.9138), sensitivity (0.9927), specificity (0.9177), and accuracy (0.9628).

1 Introduction

Skin melanoma is one of the dangerous cancers commonly affects the human community, irrespective of the age, gender, and race. Premature detection of the skin cancer province will support to decrease the possibility of death. The skin melanoma can arise in any part of the skin, and it can be identified with the well-planned screening procedures [1]. At the time of screening execution, the suspicious skin layers are visually scrutinized by an experienced dermatologist. The initial screening procedure considers the famous rule called the ABCD rule [2], in which the asymmetry, border, color, and diameter of the suspicious skin section are visually examined. If any doubtful skin abnormality is recognized during the visual examination procedure, a digital dermatoscopy is then considered to record the image in order to confirm its nature and the severity [3]. The recorded dermoscopy images are subsequently examined by means of a dedicated computer-assisted detection (CAD) system to investigate its severity and the range in order to plan for the further treatment process. The treatment involves (i). therapeutic action to eliminate the skin abnormality using dedicated drugs and (ii). employing surgical procedure to remove abnormal section.

Due to its importance, the considerable numbers of evaluation and localization procedures for the skin melanoma are extensively implemented by the researchers and scientists [3, 4]. The skin cancer literature also confirms that a substantial number of measures are proposed and implemented for the autonomous investigation of the cancerous skin layers [5].

In this paper, a CAD scheme is proposed by incorporating the preprocessing based on chaotic bat algorithm (CBA)-based Kapur's entropy and the postprocessing based on distance regularized level set (DRLS). All the experimental procedures are implemented using the MATLAB software. In this study, a well-known benchmark dermoscopy database known as DERMIS [6] is considered to examine the performance of the proposed CAD system. The DERMIS database contains 70 numbers of RGB images (44 melanoma and 26 nonmelanoma) recorded with a variety of indiscretion such as improper lighting, irregular dimension, and poor visibility. During the experimentation process, the performance of the preprocessing section is measured using the image quality measures, and the performance of postprocessing section is assessed with the image similarity measures using a comparative evaluation among the extracted melanoma region and ground truth (GT) offered by the expert member. The investigational outcome confirms that implemented CAD system tenders superior values of the image similarity measures, like Jaccard, Dice, sensitivity, specificity, and accuracy, for all the images of considered database [7, 8].

2 Methodology

This section provides the details of procedures considered in this work to examine the benchmark DERMIS database. Figure 1 depicts the processing phase involved in the proposed CAD system. In order to have a fair assessment, initially all the RGB test images and its related GT are resized to a 256 x 256 pixel image. Later, the CBA and Kapur's entropy-based thresholding procedure is implemented to improve the infected section in dermoscopy image. During the postprocessing operation, DRLS-based segmentation is executed to mine the infected skin section. Finally, a comparative analysis is carried out between extracted melanoma section and GT, and the similarity measures are then recorded.

2.1 Kapur's Multithresholding

This section presents a capable entropy function known as the Kapur's approach proposed in 1985 [9]. Originally, Kapur's approach was proposed to split gray level case based on its histogram. This approach discovers the optimal T by maximizing the overall entropy of the image. Let $T = [T_1, T_2, \ldots, T_{k-1}]$, the vector for image thresholds. Then, Kapur's function will be as in Eq. (1);

$$J_{\max} = f_{Kapur}(T) = \sum_{j=1}^{k} H_j^C \; for \; C\{1, 2, 3\} \tag{1}$$

Entropy for $_{RGB}$ is calculated for red, green, and blue segments based on the assigned T value. For multithreshold case, this process can be expressed as in Eq. (2);

$$\begin{aligned} H_1^C &= \sum_{j=1}^{T_1} \frac{Ph_j^C}{\omega_0^C} ln\left(\frac{Ph_j^C}{\omega_0^C}\right), \\ H_2^C &= \sum_{j=T_1+1}^{T_2} \frac{Ph_j^C}{\omega_1^C} ln\left(\frac{Ph_j^C}{\omega_1^C}\right), \\ &\vdots \\ H_k^C &\sum_{j=T_k+1}^{L} \frac{Ph_j^C}{\omega_{k-1}^C} ln\left(\frac{Ph_j^C}{\omega_{K-1}^C}\right) \end{aligned} \tag{2}$$

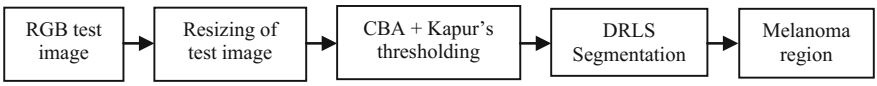


Fig. 1 Implementation of proposed skin melanoma evaluation procedure

where Ph_j^C is the probability allocation of intensity regions and $\omega_0^C, \omega_1^C, \ldots \omega_{k-1}^C$ are probability incidence for k stages. Complete description of Kapur's approach is available in [10–12].

2.2 Heuristic Algorithm

In recent years, due to its merits and the computation simplicity, heuristic algorithm-based approaches are widely proposed to solve multithresholding problem of the gray scale and RGB images [13–15]. In the proposed work, the chaotic bat algorithm (CBA) based on Ikeda map discussed in [15] is adopted to find the optimal threshold for the skin melanoma images. In this paper, a tri-level thresholding procedure is adopted to group the image pixels into background, skin region, and the melanoma section as discussed in the recent research work. The proposed thresholding approach is implemented on the RGB and gray scale versions of the test images. Finally, the performance of the CBA is then validated using the traditional bat algorithm (BA) and firefly algorithm (FA) existing in the literature [12–15]. Additional details regarding the heuristic approaches considered in this paper can be found in [16]. In order to have a fair comparison, the algorithm parameters are chosen as follows: number of agents (N) is 25, number of iteration is 2500, and the stopping criterion is J_{max}.

2.3 DRLS Segmentation

In image segmentation applications, level set approaches have been largely considered to mine vital supporting data from test picture. This paper implements the current edition of level set approach known as the distance regularized level set (DRLS) technique to extract abnormal region from the DERMIS dataset [17].

This can be represented as shown in Eq. (3);

$$\text{Considering the curve growth is } = \frac{\partial C(s,t)}{\partial t} = F\text{N} \tag{3}$$

where C—the curve vector consisting the spatial (s) and temporal (t) parameters, F—the haste variable, and N—the deepest standard vector to the curve C. Additional information regarding the DRLS can be found in [18]. Other related approaches can be found in [19–22].

2.4 Performance Evaluation

A comparative examination involving the mined tumor and ground truth is carried out to evaluate the superiority of proposed CAD. The similarity values, like Jaccard, Dice, sensitivity, specificity, and accuracy [7, 8], are computed.

These performance measures are presented in Eqs. (4)–(8):

$$Jaccard(I_{GT}, I_S) = I_{GT} \cap I_S / I_{GT} \cup I_S \tag{4}$$

$$Dice(I_{GT}, I_S) = 2(I_{GT} \cap I_S)/|I_{GT}| \cup |I_S| \tag{5}$$

$$Sensitvity = T_P/(T_P + F_N) \tag{6}$$

$$Specificity = T_N/(T_N + F_P) \tag{7}$$

$$Accuracy = (T_P + T_N)/(T_P + T_N + F_P + F_N) \tag{8}$$

where I_{GT} is ground truth image, I_S represents segmented image, T_P, T_N, F_P, and F_N denote true positive, true negative, false positive, and false negative, respectively.

3 Results and Discussion

The experimental analysis of the implemented CAD is discussed in this section. This study considers 256 × 256 sized RGB dermoscopy test images of DERMIS dataset. Primarily, the CBA and Kapur's-based tri-level thresholding and DRLS segmentation technique are experimented using melanoma test image with pseudo name SSM21. Later, similar approach is repeated with the BA and FA discussed in Sect. 2, and the convergence of the optimization search and its corresponding J_{max} (17.7826) is depicted in Fig. 2. From this image, it can be noted that the search convergence of the CBA (327th iteration) is better compared with the BA (811th iteration) and FA (375th iteration). Finally, all the heuristic approaches will offer the similar J_{max}.

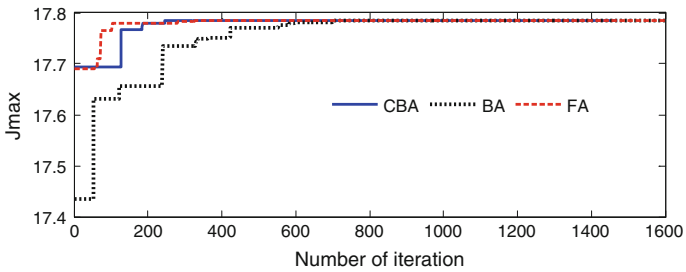


Fig. 2 Convergence of optimization search (for RGB SSM21 image

Figure 3 depicts the chosen test image (SSM21) and its corresponding outcomes. The proposed tri-level thresholding is implemented in both the RGB and the gray scale versions of the image, and finally, the extracted melanoma region is compared against the GT. Figure 3a, b, e, and g represents the test image, thresholded RGB image, DRLS-segmented RGB image, and the mined melanoma region, respectively. Similarly, Fig. 3c, d, f, g, and h represent the thresholded gray scale image, DRLS bounding box, segmented gray image, mined melanoma region, and GT, respectively. This study confirms that both the RGB and the gray scale approaches are efficient in providing better result. In which the CPU time taken by the RGB image is considerably large compared to gray scale case. Hence, for further investigation, only the gray scale versions of the test images are considered.

Figure 4 shows the results obtained for the chosen test image cases. In which the pseudo names SSM4, SSM7, and SSM32 are from the melanoma cases and remaining images, like 5, 10, and 29 are from the non-melanoma cases of DERMIS. After extracting the melanoma section from the chosen test image, a comparative analysis is carried out with the corresponding GT in order to evaluate the performance of the proposed CAD system. During this procedure, the image similarity measures discussed in [7, 8] are computed, and their values are clearly presented in Tables 1 and 2.

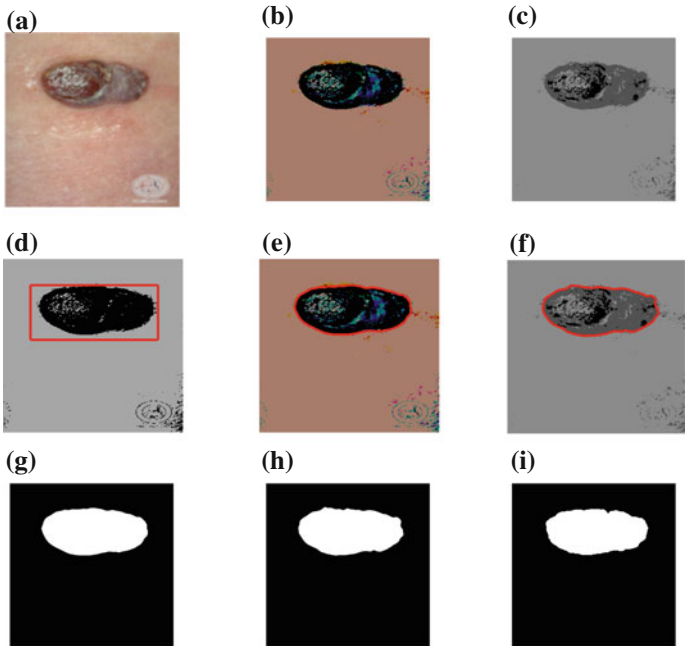


Fig. 3 Experimental results for SSM21 image. **a** test image, **b**. thresholding outcome for RGB image, **c**. thresholding outcome for gray scale image, **d**. initialization of the box in DRLS, **e**. segmentation results for RGB image, **f**. segmentation results for RGB image, **g**. extracted region with RGB image, **h**. extracted region with gray image, **i**. ground truth

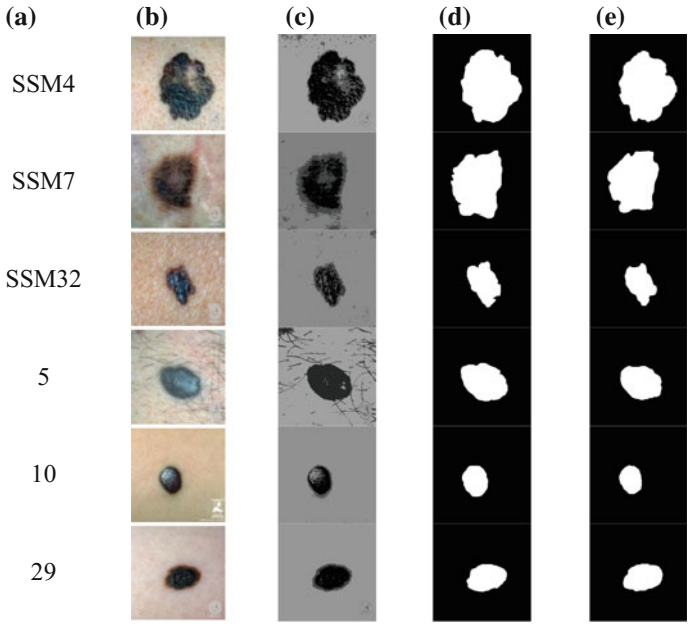


Fig. 4 Experimental results. **a** pseudo name, **b** test image, **c** threshold gray scale image, **d** extracted skin section, **e** ground truth

Table 1 Image similarity index obtained for the chosen skin melanoma images

Pseudo name	Jaccard	Dice	Sensitivity	Specificity	Accuracy
SSM21 (RGB)	0.9354	0.9666	0.98541	1.0000	0.9927
SSM21 (Gray)	0.9449	0.9717	0.9887	0.9978	0.9932
SSM4	0.9035	0.9493	0.9557	0.9996	0.9774
SSM7	0.8403	0.9132	0.9497	0.9713	0.9604
SSM32	0.8374	0.9115	0.9829	0.9890	0.9860
5	0.8287	0.9063	0.9730	0.9997	0.9863
10	0.7906	0.8830	0.9859	0.9947	0.9903
29	0.9327	0.9652	0.9967	0.9649	0.9807
Average	**0.8767**	**0.9333**	**0.9772**	**0.9896**	**0.9834**

Table 2 Average similarity index obtained for DERMIS dataset

Test images	Jaccard	Dice	Sensitivity	Specificity	Accuracy
Melanoma (44 images)	0.8805	0.9138	0.9927	0.9177	0.9628
nonmelanoma (26 images)	0.8616	0.9062	0.9889	0.9055	0.9553

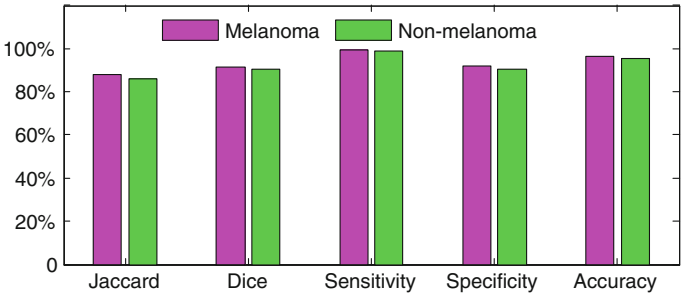


Fig. 5 Graphical representation of the average similarity index obtained for melanoma and nonmelanoma images

Table 1 shows the Jaccard, Dice, sensitivity, specificity, and accuracy values of the chosen images of DERMIS dataset. For SSM21 image, the gray scale version offered better result compared with the RGB. Hence, for other images, the gray scale version is adopted. The average value of Table 1 confirms that proposed approach is efficient in extracting the melanoma section from the DERMIS database.

Table 2 and Fig. 5 present the overall similarity obtained for the melanoma (44 images) and nonmelanoma (26 images) cases of the DERMIS database. From this table and the figure, it can be observed that proposed CAD system offers better result for the image similarity values.

4 Conclusion

This paper employs a CAD by uniting Kapur's thresholding and DRLS technique to enhance the outcome of skin melanoma scrutiny. In this study, benchmark DERMIS dataset is adopted to experiment the performance of the constructed CAD. The RGB digital dermoscopy images are primarily investigated with the tri-level thresholding procedure using the CBA, BA, and FA. The results substantiate that the CPU time taken by the CBA is superior compared to the BA and FA. After the preprocessing approach, DRLS-based segmentation is considered to postprocess the skin cancer dataset. A relative examination between the segmented skin region and the ground truth image for the melanoma case offered better values of Jaccard (88.05%), Dice (91.38%), sensitivity (99.27%), specificity (91.77%), and accuracy (96.28%). Approximately, similar results are attained with the nonmelanoma images. Hence, the proposed CAD system works well, and in the future, it can be used to diagnose real-time clinical images.

References

1. Xu, L., Jackowski, M., Goshtasby, A., Roseman, D., Bines, S., Yu, C., Dhawan, A., Huntley, A.: Segmentation of skin cancer images. Image Vis. Comput. **17**, 65–74 (1999)
2. Nachbar, F., Stolz, W., Merckle, T., et al.: The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. J. Am. Acad. Dermatol. **30**, 551–559 (1994)
3. Amelard, R., Glaister, J., Wong, A., Clausi, D.A.: Melanoma decision support using lighting-corrected intuitive feature models. In: Computer Vision Techniques for the Diagnosis of Skin Cancer, Series in BioEngineering, pp. 193–219 (2013)
4. Glaister, J., Wong, A. and Clausi, D.A: Segmentation of skin lesions from digital images using joint statistical texture distinctiveness. IEEE Trans. Biomed. Eng. **61**(4), 1220–1230 (2014)
5. Amelard, R., Glaister, J.: Wong, A. and Clausi, D.A: High-level intuitive features (HLIFs) for intuitive skin lesion description. IEEE Trans. Biomed. Eng. **62**(3), 820–831 (2015)
6. http://vip.uwaterloo.ca/demos/skin-cancer-detection
7. Chaddad, A., Tanougast, C.: Quantitative evaluation of robust skull stripping and tumor detection applied to axial MR images, Brain Inf. **3**(1), 53–61(2016). https://doi.org/10.1007/s40708-016-0033-7
8. Rajinikanth, V., Satapathy, S.C., Fernandes, S.L., Nachiappan, S.: Entropy based segmentation of tumor from brain MR images–a study with teaching learning based optimization. Pattern Recogn. Lett. **94**, 87–94 (2016). https://doi.org/10.1016/j.patrec.2017.05.028
9. Kapur, J.N., Sahoo, P.K., Wong, A.K.C.: A new method for gray-level picture thresholding using the entropy of the histogram. Comput. Vis. Graph. Image Process. **29**, 273–285 (1985)
10. Bhandari, A.K., Kumar, A., Singh, G.K.: Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using Kapur's, Otsu and Tsallis functions. Expert Syst. Appl. **42**, 1573–1601 (2015)
11. Lakshmi, V.S., Tebby, S.G., Shriranjani, D., Rajinikanth, V.: Chaotic cuckoo search and Kapur/Tsallis approach in segmentation of t.cruzi from blood smear images. Int. J. Comput. Sci. Inf. Secur. (IJCSIS) **14**(CIC 2016), 51–56 (2016)
12. Manic, K.S., Priya, R.K., Rajinikanth, V.: Image multithresholding based on Kapur/Tsallis entropy and firefly algorithm. Indian J. Sci. Technol. **9**(12), 89949 (2016)
13. Rajinikanth, V., Aashiha, J.P., Atchaya, A.: Gray-level histogram based multilevel threshold selection with bat algorithm. Int. J. Comput. Appl. **93**(16), 1–8 (2014)
14. Rajinikanth, V., Couceiro, M.S.: Optimal multilevel image threshold selection using a novel objective function. Adv. Intell. Syst. Comput. **340**, 177–186 (2015)
15. Satapathy, S.C., Raja, N.S.M., Rajinikanth, V., Ashour, A.S.: Dey, N: Multi-level image thresholding using Otsu and chaotic bat algorithm. Neural Comput. Appl. (2016). https://doi.org/10.1007/s00521-016-2645-5
16. Yang, X.S.: Nature-Inspired Metaheuristic Algorithms, 2nd edn. Luniver Press, Frome (2011)
17. Li, C., Xu, C., Gui, C., Fox, M.D.: Distance regularized level set evolution and its application to image segmentation. IEEE Trans. Image Process. **19**(12), 3243–3254 (2010)
18. Vaishnavi, G.K., Jeevananthan, K., Begum, S.R., Kamalanand, K.: Geometrical analysis of schistosome egg images using distance regularized level set method for automated species identification. J. Bioinf. Intell. Control **3**(2), 147–152 (2014)
19. Kumar, R., Talukdar, F.A., Dey, N., Ashour, A.S., Santhi, V., Balas, V.E., Shi, F.: Histogram thresholding in image segmentation: A joint level set method and lattice boltzmann method based approach. Adv. Intell. Syst. Comput. **455**, 529–539 (2016). https://doi.org/10.1007/978-3-319-38771-0_52

20. Kumar, R., Rajan, A., Talukdar, F.A., et al.: Neural Comput. Appl. (2016). https://doi.org/10.1007/s00521-016-2267-y
21. Li, Z., Dey, N., Ashour, A.S., Cao, L., Wang, Y., Wang, D., McCauley, P., Balas, V.E., Kai Shi, K., Shi, F.: Convolutional neural network based clustering and manifold learning method for diabetic plantar pressure imaging dataset. J. Med. Imaging Health Inf. **7**(3), 639–652 (2017)
22. Ngan, T.T., Tuan, T.M., Minh, N.H., Dey, N.: Decision making based on fuzzy aggregation operators for medical diagnosis from dental x-ray images. J. Med. Syst. **40**(12), 280 (2016)

Estimation of Distance of a Target Speech Source by Involving Monaural Features and Statistical Properties

R. Venkatesan and A. Balaji Ganesh

Abstract The paper discusses a novel system for the estimation of distance of a target speaker by involving statistical properties in a reverberant condition. The system involves the extraction of statistical features from both cepstral and envelope coefficients of a speaker at different distances. Further, different spectral or monaural features are analysed at distinct distances for different room environments. The distance-dependent statistical properties are considered for the feature extraction process. A set of statistical parameters are used to learn GMM-EM pattern recognizer for effective classification. The results observed that the system performance is very much dependent on the reverberation time and also robustness of the monaural features. The results of the proposed system show the significant improvement in signal-to-noise ratio of 0 dB (babble noise) under reverberation time 0.48 s over other existing methods.

1 Introduction

The human auditory model has the capability of determining the distance of sound source, even when there is no visual information is available [1–3]. Full-source localization includes measuring azimuth and distances of sound sources which plays the crucial role for both human and artificial listeners [2]. In many applications such as intelligent hearing-aid devices, speaker recognition system, auditory scene analyser and audio surveillance system are required to track the moving speakers [1–3]. The distance estimation module consists of two stages, namely

R. Venkatesan (✉)
Department of Electronics and Communication Engineering,
Velammal Engineering College, Surapet, Chennai 600066, India
e-mail: venky88an@gmail.com

A. B. Ganesh
Department of Electronics and Electrical Engineering,
Velammal Engineering College, Surapet, Chennai 600066, India
e-mail: abganesh@velammal.edu.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_20

feature extraction and pattern recognition, which are responsible for estimating the distance between the source and destination. Research findings are available based on factors affecting auditory distance perception model, such as room reverberation, room volume, directivity and external disturbances [4–6]. Hioka et al. [5] suggested a novel technique by involving spatial correlation model for direct-to-reverberant ratio (DRR) estimation. This method is applicable only for detecting sounds in limited range of distances, and it is adopted for applications where microphone arrays are employed explicitly. Lately, research works are found based on training the machine learning algorithms for distance estimation [1, 2, 7]. Vesa et al. [8] have introduced the distance detection based on the short-time magnitude-squared coherence between two ears and also the system performs classification using Gaussian maximum-likelihood scheme. Later, Georganti et al. [1] have proposed the improved technique based on extraction of statistical properties from the single-channel microphone which are incorporated with classification-based Gaussian mixture model. Several research findings are available based on utilizing cepstral and envelope coefficients for automatic speaker and speech recognition applications [9–11]. Sadjadi et al. [9] introduced mean Hilbert envelope coefficients which is an alternative to MFCCs and produces promising results for robust speaker identification under noisy reverberant environment. Recently, some research works are available to obtain the statistical features from the cepstral coefficients for classification framework. Nandhini et al. [10] proposed a lung sound analysis and its classification based on statistical properties of cepstral features. The author utilized these features for training the classifier.

2 System Architecture

The proposed method analyses input speech source sounds from same target speakers but at different distances. The study is based on the statistical properties of fused coefficients of envelope and cepstral features. The system has proposed distance estimation of speech source by exploring monaural features as it shows robustness against noisy conditions. The statistical parameters, such as envelope skewness, envelope kurtosis, linear prediction residual ratio, along with the other acoustic parameters, are computed from fused monaural features. The extracted features from the monaural cues are promoted into Gaussian mixture model-expectation maximization (GMM-EM) for learning process.

The developed system involves the comparison of results based on statistical properties of monaural features and existing system at different distances. The results from the pattern recognizer are utilized for selection of better ear and additionally to obtain higher signal-to-noise ratio for automatic speaker recognition module. Figure 1 shows the system architecture of proposed auditory distance perception module. The speech signal of the speaker from TIMIT dataset and its corresponding spectrograms for different distances at reverberation time of 1.15 s is shown in Fig. 2.

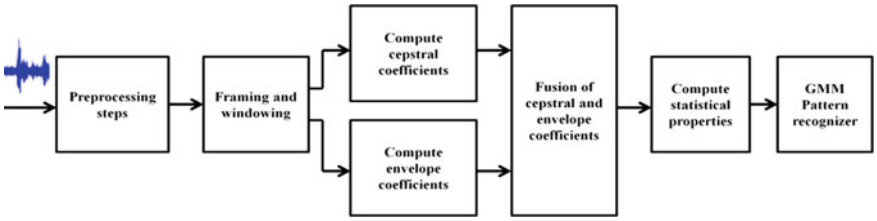


Fig. 1 Schematic structure of proposed auditory distance perception module for distant speech source

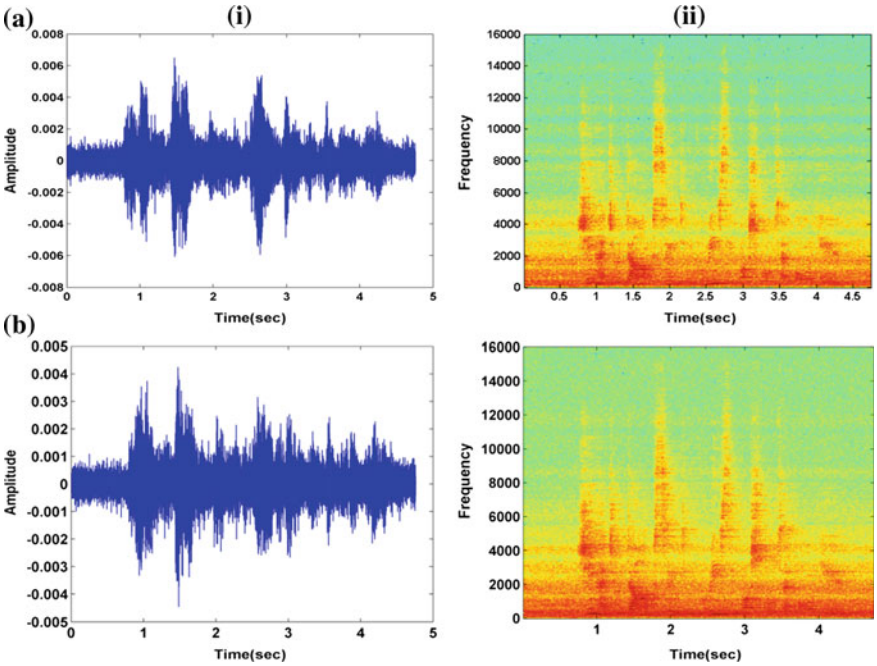


Fig. 2 Speech signals and their corresponding spectrograms for different distances (1 m, 2 m) in stairway at reverberation time of 1.15 s

It is learnt that the cepstral and envelope features are predominantly used in speech processing applications, including speaker and speech recognition systems. The study analyses the features of both cepstral and envelope for the development of distance perception model. Also, the variation in the histograms of cepstral and envelope features at different distances is studied.

The monaural features that are considered for the distance estimation are MHEC and MFCC. The cepstral and envelope analyses are inspired mainly through various speaker recognition systems. The detailed descriptions on the extraction of cepstral and envelope coefficients are given in the following sections.

2.1 Mean Hilbert Envelope Coefficient (MHEC)

A 24-channel Gammatone filter banks are utilized for constructing MHEC feature from speech samples. The speech signal is decomposed into 24 bands by executing above process. The smoothed Hilbert envelopes are responsible to obtain useful acoustic speaker-sensitive information. Eventually, 36-dimensional MHEC features have been obtained by considering first and second derivatives.

2.2 Mel-Frequency Cepstral Coefficient (MFCC)

The Mel coefficients are determined by means of segmenting input signal into 20 ms frames with 10 ms frame shift. The power spectrum is obtained through hamming windowing of each frame.

Then, the derived power spectrum is warped into Mel scale, and the further process is accomplished by log compression and discrete cosine transform (DCT) to obtain 39-dimensional MFCC coefficients.

Figure 3 shows the normal distribution followed by MFCC and MHEC features. The cepstral and envelope features have large variations in frequencies of histograms in different distance. These variations in frequency and curve of normal distribution of histograms have dependency over reverberation time of rooms.

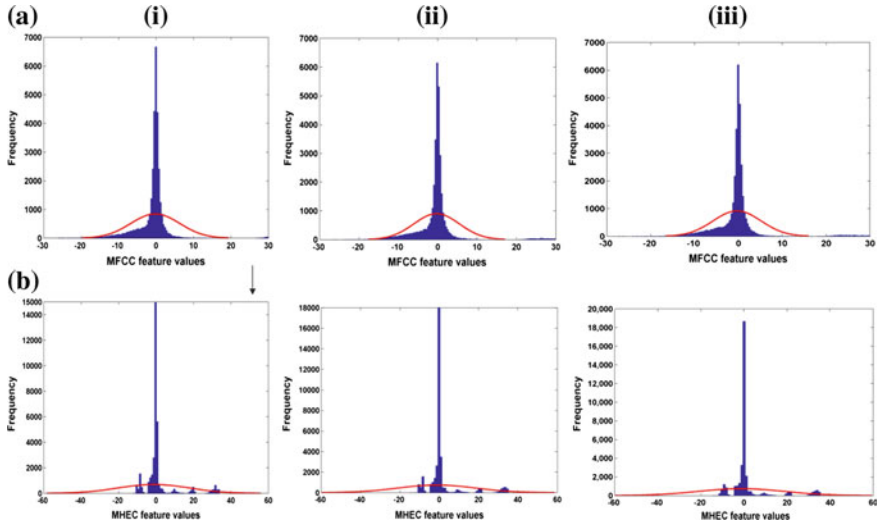


Fig. 3 Histograms for cepstral (MFCC) and envelope (MHEC) features of speech source for different distances (1 m, 2 m, 3 m) in stairway at reverberation time of 1.15 s

3 Statistical Properties for Classification

Hilbert envelope $E_{i,t}$ of the cepstral and envelope coefficients are computed independently. Additionally, the following statistical parameters, skewness, kurtosis, standard deviation, mean and variance, are considered for the above-discussed coefficients. The above-designed extraction process yields features by following our previous work [2].

Linear prediction (LP) analysis: The linear prediction analysis [1] is referred by predicting the each sample of speech signal as a linear combination of past samples. The individual frame of the speech samples is processed by linear prediction analysis of order, p [36]. It is found that linear prediction coefficients have dependency over change in distances of receiver with respect to speech source (or vice versa). The predicted sample [1], s_p is obtained by

$$s_p = -\sum_{j=1}^{p} a_j s(n - j) \tag{1}$$

where s(n) is the nth instant speech sample and a_j are the LP coefficients. Linear prediction ratio [1] is computed by the ratio of percentile (0.9), P_{Fr}, to the rms, R_{Fr}, that is,

$$\text{LPrat} = \frac{P_{Fr}}{R_{Fr}} \tag{2}$$

The estimation of linear predictive ratio for the speech sources is carried out at the various distances, and it is found that the above-discussed feature is dependent on different distance of source with reference to receiver.

Linear predictive residual kurtosis and linear predictive residual skewness: Linear predictive residual amplitude values of each frame are highly preferred for computing the statistical parameters, namely linear predictive kurtosis, LPkurts [1], linear predictive skewness, LPskew [1]. The above-discussed parameters are described as

$$\text{LPkurts} = \frac{\frac{1}{N}\sum_{t=1}^{N}(K_{s,t} - \mu_{ks})^4}{\left[\frac{1}{N}\sum_{t=1}^{N}(K_{s,t} - \mu_{ks})^2\right]^2} \tag{3}$$

where μ_{ks} is the mean of $K_{s,t}$ and $K_{s,t}$ specifies the linear predictive amplitude values

$$\text{LPskew} = \frac{\frac{1}{N}\sum_{t=1}^{N}(W_{s,t} - \mu_{sk})^3}{\left[\frac{1}{N}\sum_{t=1}^{N}(W_{s,t} - \mu_{sk})^2\right]^{\frac{3}{2}}} \tag{4}$$

where μ_{sk} is the mean of $W_{s,t}$ and $W_{s,t}$ denotes linear predictive amplitude values. Both these features have dependency over the distance amongst source/receiver.

A total set of ten distant-dependent statistical features are calculated from the 74-D of feature level fused coefficients that are integrated into GMM-EM pattern recognizer for distance estimation.

4 GMM-EM for Distance Estimation

GMM-EM is initialized by k-means clustering, and also it is a statistical model with weighted sum of N multivariate Gaussian components [7] which is obtained by

$$P(x|\lambda) = \sum_{i=1}^{N} w_i p_i(x) \tag{5}$$

where x represents D-dimensional feature vector, whose probability is being estimated, and w_i is the weight of the *i*th probability density, that is described as:

$$P_i(x) = \frac{1}{\sqrt{(2\pi)^D |\sigma_i|}} \exp\left(\frac{-1}{2}(x - \mu_i)' \sigma_i (x - \mu_i)\right) \tag{7}$$

where μ is the mean, σ is the covariance of Gaussian distribution, and $(')$ represents the vector transpose.

The GMM-EM comprises likelihood function to make decision for classifying different distances. The GMM-EM has dependency over the number of Gaussian components. The EM algorithm is employed to measure the GMM parameters with the maximum value of 300 iteration, and it is set to $\in = 1e^{-5}$.

5 Result and Discussions

The speech samples at different distances are generated by considering the original speech sources (from TIMIT database) which are convolved with room impulse responses (RIR) created from Aachen Impulse Response (AIR) database [2] under different room properties and other impulse responses from [12]. Nearly, 100 speech samples for each class (0 m, 1 m, 2 m and 3 m) are utilized for experiments. The confusion matrix for fused monaural features is analysed for different distances at reverberation time (0.48 s) with SNR of 0 dB (babble noise), and the results are tabulated. Tables 1 and 2 show the confusion matrix obtained for existing method [1] and proposed method consequently when the speaker is unknown. The estimated mean performance based on classification using existing method, fused coefficients are 75.4, and 79.2% respectively. The proposed distance estimation

Table 1 Confusion matrix related to performance of the existing method based on statistical and source-specific features without involving monaural features

A/P	0 m	1 m	2 m	3 m	Accuracy	Precision	Sensitivity	Specificity
0 m	99	1	0	0	99	96.12	99	99
1 m	0	74	14	12	88.6	70.5	74	86.4
2 m	0	25	32	43	80.4	51,6	32	92.5
3 m	2	7	16	75	81.25	57.3	75	83.1

Table 2 Confusion matrix representing the performance of proposed method by incorporating source-specific statistical features of fused monaural coefficients

A/P	0 m	1 m	2 m	3 m	Accuracy	Precision	Sensitivity	Specificity
0 m	99	1	0	0	99.6	95.2	99	98.75
1 m	0	82	8	10	91.6	77.35	82	94
2 m	1	20	36	43	84.2	70.5	36	96.25
3 m	1	9	8	82	85.2	59.5	81	86.25

method produces an average improvement of more than 3% over reverberation room, 0.48 s compared to other conventional method [1].

The system performs well as it is not dependent on signal gain, and the level of microphone as well as the performance is not more affected by the output of individual speaker. The developed system is robust towards the overlapping of the feature space values across different classes. Figure 4 explains the performance of proposed method by incorporating monaural features with the pattern recognizer where the speaker is known for different distances, and it compared to conventional algorithm [1].

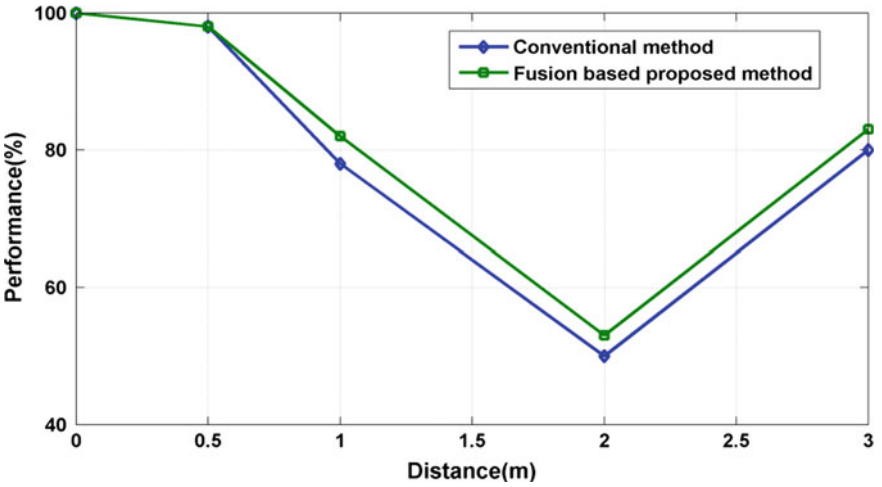


Fig. 4 Performance (in %) of proposed method when the speaker is known for different distances at reverberation time of 0.48 s and made comparison with baseline method

6 Conclusion

The paper discusses the merits of exploring machine learning technique for the estimation of distance based on statistical properties of fused coefficients of cepstral and envelope features. The proposed system is tested in a noisy environment (SNR of 0 dB) under reverberation time of 0.48 s. The mean performance of proposed systems is computed and compared with conventional technique. The estimated distance is found to be very useful in the construction of full localization system for estimating ideal binary mask and also for the selection of better ear. The estimated ideal binary mask provides discriminant information by improving signal-to-noise ratio in the speaker recognition module. The developed system has a broad range of cognitive speech processing-related applications, such as audio surveillance, hands-free communication systems.

References

1. Georganti, E., May, T., van de Par, S., Harma, A., Mourjopoulos, J.: Speaker distance detection using a single microphone. IEEE Trans. Audio Speech Lang. Process. **19**(7), 1949–1961 (2011)
2. Venkatesan, R., Balaji Ganesh, A.: Full sound source localization of binaural signals. In: International Conference on Wireless Communication, Signal Processing and Networking (2017). (Accepted)
3. Lu, Y.C., Cooke, M.: Motion strategies for binaural localisation of speech sources in azimuth and distance by artificial listeners. Speech Commun. **53**(5), 622–642 (2011)
4. Kuster, M.: Estimating the direct-to-reverberant energy ratio from the coherence between coincident pressure and particle velocity. J. Acoust. Soc. Am. **130**(6), 3781–3787 (2011)
5. Hioka, Y., Niwa, K., Sakauchi, S., Furuya, K., Haneda, Y.: Estimating direct-to-reverberant energy ratio using D/R spatial correlation matrix model. IEEE Trans. Audio Speech Lang. Process. **19**(8), 2374–2384 (2011)
6. Lu, Y.C., Cooke, M.: Binaural estimation of sound source distance via the direct reverberant energy ratio for static and moving sources. IEEE Trans. Audio Speech Lang. Process. **18**(7), 1793–1805 (2010)
7. Georganti, E., May, T., Van de Par, S., Mourjopoulos, J.: Sound source distance estimation in rooms based on statistical properties of binaural signals. IEEE Trans. Audio Speech Lang. Process. **21**(8), 1727–1741 (2013)
8. Vesa, S.: Sound source distance learning based on binaural signals. In: Proceedings of the 2007 Workshop on Application of Signal Processing Audio, Acoustic, ,pp 271–274 (2007)
9. Sadjadi, S.O., Hansen, J.H.L.: Mean Hilbert envelope coefficients (MHEC) for robust speaker and language identification. Speech Commun. **72**, 138–148 (2015)
10. Sengupta, N., Sahidullah, M., Saha, G.: Lung sound classification using cepstral-based statistical features. Comput. Biol. Med. **75**, 118–129 (2016)
11. Venkatesan, R., Balaji Ganesh, A.: Unsupervised auditory saliency enabled binaural scene analyzer for speaker localization and recognition. In: Advances in Signal Processing and Intelligent Recognition Systems, vol. 674. Springer, Berlin (2018)
12. Vesa, S.: Binaural Source distance learning in rooms. IEEE Trans. Audio Speech Lang. Process. **17**(8), 1498–1507 (2009)

Termite-Motivated Simulation of Cooperative Behavior

Dhruv Chamania, Amit Adate and Parveen Sultana

Abstract One of the most instrumental factors for successful collaboration in a group is its capacity to perform an assortment of tasks. The capacity to perform said tasks is termed as collective intelligence. In this paper, we have analyzed the collective biological behavior observed in termite colonies. Further, we have simulated an environment in which the mechanisms inspired from termite colonies are applied for the task of two-dimensional structural buildup. We have elaborated by implementing learning mechanisms which are yielding progressive results with both conditional and quantifiable benchmarks.

1 Introduction

Over the past few years, there has been a tremendous amount of growth in the integration of artificial intelligence in our daily lives. Many innovations are being done that are focused on easing the overall load of mankind. Researchers tend to take motivations from biological aspect of the nature to solve given problems like ant colony [1] or particle swarm [2]. Many of these innovations are focused on performing tedious work by using multiple independent units to work together intelligently to solve problems like the building of structures or solving assembly line problems. One such idea that makes use of the above parameters is the intelligence that termites show while construction of huge structures called as termite mounds. The ability of a colony of insects to create structures several meters in length and diameter despite each insect having a small size is staggering. One has to look into this from a mathematical and technical perspective to understand how the collective behavior gets defined in order to optimize algorithms based on ant colonies or particle swarms.

D. Chamania (✉) · A. Adate · P. Sultana
SCOPE, VIT University, Vellore 632014, Tamil Nadu, India
e-mail: dhruvajay.chamania2014@vit.ac.in

A. Adate
e-mail: adateamit.sanjay2014@vit.ac.in

P. Sultana
e-mail: hparveensultana@vit.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_21

Understanding this collective biological behavior can enable us to create physical structures from relatively small individual intelligent units. The structures to be built can be as complex as buildings or houses which require loads of natural resources. Directly building structures like these using robots without any analysis or simulation can lead to wasting resources. Hence, there is a need to study collective intelligence in a systematic way [3]. Not performing adequate analysis can also cause impediments in improving space-time complexities.

In this paper, we present a tool design that is able to deal with the parameters related to termite-inspired structures in a mathematical way. The software architecture of the tool has been developed with a view to help resolve issues that arise with associated complexity. The software tool design will enable working on different aspects of the basic units that emulate termites and using them in a way to help implement forming of different structures. The methodology that we make use of, its basic layout will help us save time, resources and will help improve the overall time complexity related to this problem. I hope, our analysis will serve as a guide to people working on this problem in future.

Rest of our work is planned in the following way. Section 2 gives account of some literature in relation with termite colony's biological and technological aspects. In Sect. 3, we extensively deal with the methodology that we are going to adapt. Section 4 provides us with results that we have obtained by applying the architecture defined in Sect. 3. And we end with describing the possible applications of our research in future studies.

2 Related Work

The background on meta-heuristics is to be found in the artificial intelligence research communities. Collective intelligence heuristics have become powerful and popular in computational intelligence and its many applications. There are some important developments in recent years. We would like to provide a timely review of such developments, namely collective resolution of creative problems catered to various notions of conceptual designs. In recent trends, it is identified that higher forms of collective intelligence play a salient role in meta-heuristic computing and advances in collaborative work strategies. According to Flores et al. [4], collective efforts are considered as the metric to be improved rather than individual methods. The paper describes the importance of the cohesion between individuals to create and complete the set experimental tasks and how it tends to improve overall results and not fully rely on the ability of one single individual. The paper follows on with stating the recent applications of collective intelligence in the industry. It also describes the development tools used for building aspect-specific frameworks for implementing real-world problems. Although our final output of the structure and the parts of the structure are fixed making our problem non-heuristic, one still needs to look into how people deal with space-time complexities in collective intelligence.

We also investigated literature dealing with simulations resulting from collective intelligence observed in termites. Researchers have approached in different ways to implement the industrious termites' methodology to solve various problems. A biological research of this can be seen in work Tuner [5]. The author has provided us with a detailed analysis of the groundworks behind the making of macro-termite structures. He has given us a comprehensive view of the working, maintenance, and the swarm behavior that the termites have and how they progress over a period of time. This covers almost all the literature that one needs for implementing it in different hemispheres like 2D to 3D reconstruction, which the author has suggested in his conclusion.

Looking at it from a technological viewpoint, Radhika Nagpal and her team have made some tremendous progress in emulating termite behavior, which they have demonstrated in robots. A subsidiary of her teams' work, Wergel et al. [6], is our base paper. It has a detailed description behind the motivation, methodology, and application of termite behavior idea in robots. The author has provided us with a mathematical look into the collective mind of termites, which keep only the final structure in their minds, while still helping in building parts of the huge structures. The paper elaborates their procedure in designing of the robots and detailing the structures which have been made using termite behavior emulation. The authors' detailed description of the building process along with the different approximations that were made for the robots to function properly gives us a clearer perspective on how we can approach the problem in a simulated environment for us to measure different efficiency-related parameters. The final aim is to train robots to build real-life structures but as shown that the author has concluded with a more pictorial representation of the working of robot with just some basic ramps. With the help of this base paper, we are going to formulate our methodology and implementation.

Other works and implementations, different from our goals but on the same concept are Vimal et al. [7, 8]. More recent work, Nagpal et al. [9] show further progress that has been made by her team on the robot construction by termite project.

3 Methodology

Our proposed method to develop termite-inspired cooperative behavior is focused on three parameters, viz. the environment and memory in which we are going to make our termites work, the definition of each of our termite units, and finally the way in which we are going make our termites learn.

3.1 Space and Memory of our Termite Colony

We decided to go for two-dimensional structural buildup as the working space in which we want our termite colony behavior to operate in. In the developed environ-

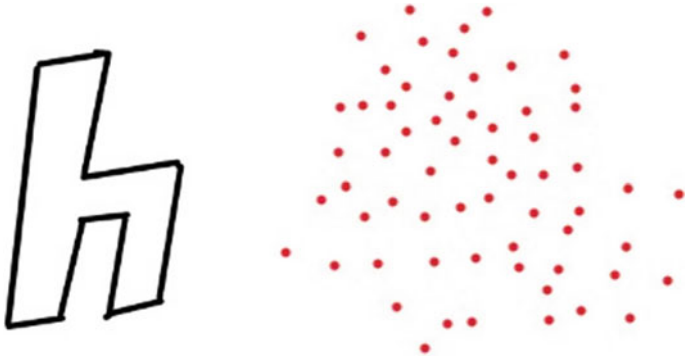


Fig. 1 Fixed collective structure along with the randomly generated termites

ment, the positional metrics are represented in a grid format and the cost metric is represented as the ink required to build the structure. As seen in Sect. 2, all the termites have a fixed collective final structure in their mind. For this, we built a custom designed structure as shown in Fig. 1 that is stored in each termite's functional memory. Each individual unit approaches the environment with respect to contributing toward generating the stored structure in their collective memory.

3.2 Termite Unit

On the basis of the environment that we formed, we gave each termite a fixed capacity of ink. The ink is representing of the block that the termite makes use if to build a given part. Once the termite exhausts its reserve, then we bring in the next termite into the environment to further complete the task. Every incoming termite analyzes the current structure, learns upon it, and then evaluates its contribution to be performed. While contributing, it can perform two operations, firstly to spatially choose which of points are not available to fill and secondly to fill them with ink. Upon exhaustion of the ink reserve, the process gets rolled back and a new termite gets rolled in. We randomly generate a swarm of termites with each having the above-mentioned properties as depicted in Fig. 1. To build the pre-defined structure stored in the memory, the allotment for the structural block is done spatially, in extension to their x-coordinate.

3.3 Intelligence in Our Termites

For incorporating intelligence in our termites, we decided to implement a support vector machine (SVM). Its main purpose is to segregate upon the two operations. We

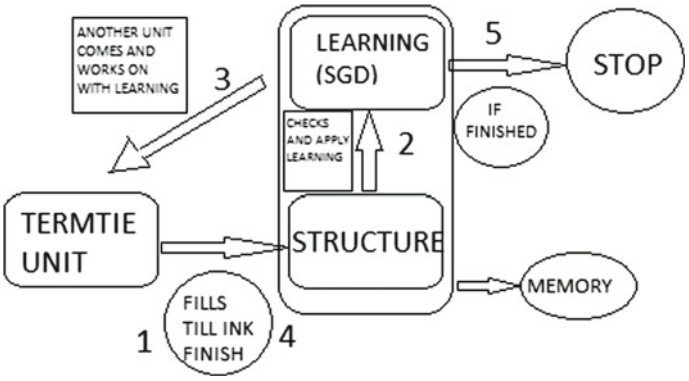


Fig. 2 Flow diagram of our implementation

used supervised learning along with stochastic gradient descent (SGD) loss function. An update of SGD hyper-parameters is performed with every termite roll-in. The equations used are based on Andrew et al. [10]. The extent to which each termite predicts the metrics of its contribution decides the learning rate for the next roll-in [10]. Hence, the learning rate starts to gradually increase. The model gets trained sequentially with the learning rate being picked up after each termite roll-out.

3.4 Implementation

Figure 2 represents our implementation workflow. The unit starts working on the structure, and SGD is applied to the associated structure. This process continues and we keep checking if it is complete. If not it will continue working on the structure, applying learning from what it had learned during the previous iterations. This will continue till completion.

4 Results

In our simulated environment, the colony of generated termites shares the same functional memory, but each termite has access to a data frame that updates as the structure is built up. Our randomly generated function yielded a colony of 68 termites which are depicted by yellow dots. As seen in Fig. 3, the termites iteratively start working on the structure stored in the aforementioned collective mind in the two-dimensional space. After a single termite completes it capacity, the path that the next termite takes has to be learned upon the data frame to generate the two-dimensional structure. The SVM acts as a memory along with incorporating learning and predicts

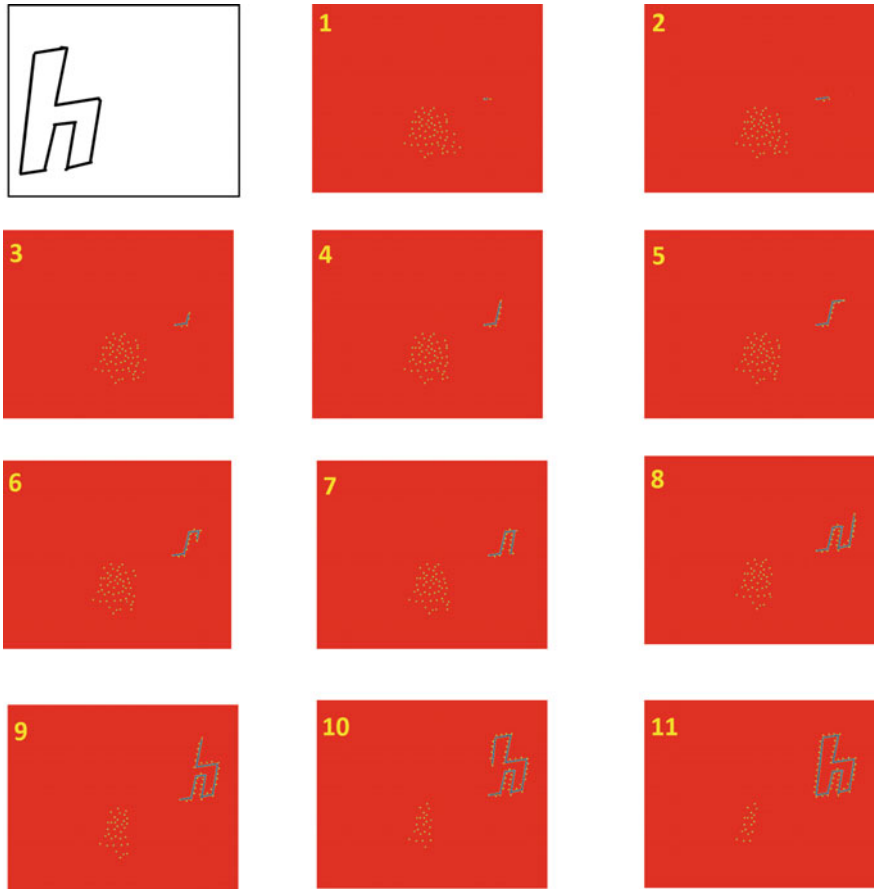


Fig. 3 Heat map showing formation of our 2-D termite mound over different iteration

the gradient of its successor on the precepts learned about from its predecessor. This is incorporated in the SVM that was designed.

The path is indicated by the blue line that extends over the iterations. Our termite unit makes its choice on the basis of the gradient decent and the learned output in the most efficient way possible to continue. This step of intelligent thinking can be observed when we look into the frames 2–3 or in frame 7–8 of Fig. 3 where according to the base guidelines of our termite buildup, they needed to continue on the x-axis. But, the descent makes it change in axis to proceed in making of the structure. This intelligence can help in providing with better space-time complexity. For the transformation of the structure, the two-dimensional buildup was observed with the roll-call of 42 termites out of the 68 termites. For testing of any other such structures, one can simply create a pixelated image and providing it to our tool. The termite units in our tool start completing the structure intelligently. That together

with the collective memory help us simulate the termite cooperative behavior in the same way as the termites in nature work.

To view our code and the images generated: https://github.com/amitadate/termite-sim.

5 Conclusion

For our experiment, we were able to deal with the probabilistic part of our termite colony. We were able to make our termite units intelligently select the most efficient path possible and also make it behave in a more collective way. Others can make use of our tool to try out various structures and extend our tool implementation.

Next steps for our research shall deal with three-dimensional structures which will allow creating structures as mentioned in Sect. 1. Another important parameter that we did not consider is the time complexity of the solution. The efficient path taking is useful but that together with the time aspect of the termite simulating mound will be the most accurate depiction of the termite colonies.

References

1. Dorigo, M., Birattari, M., Stutzle, T.: Ant colony optimization. IEEE Comput. Intell. Mag. **1**(4), 28–39 (2006)
2. Kennedy, J.: Particle Swarm Optimization. pp. 760–766. Springer, Boston (2010)
3. Chmait, N.: Understanding and measuring collective intelligence across different cognitive systems: an information-theoretic approach. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 5171–5172 (2017)
4. Flores, R.L., Negny, S., Belaud, J.P., Le Lann, J-M.: Collective intelligence to solve creative problems in conceptual design phase. In: Procedia Engineering. TRIZ and Knowledge-Based Innovation in Science and Industry, vol. 131, pp. 850 – 860 (2015)
5. Scott Turner, J.: Termites as models of swarm cognition. Swarm Intell. **5**(1), 19–43 (2011)
6. Werfel, J., Petersen, K., Nagpal, R.: Designing collective behavior in a termite-inspired robot construction team. Science **343**(6172), 754–758 (2014)
7. Dhokia, V., Essink, W.P., Flynn, J.M.: A generative multi-agent design methodology for additively manufactured parts inspired by termite nest building. CIRP Ann. **66**(1), 153–156 (2017)
8. Saffre, F., Hildmann, H., Deneubourg, J.-L.: Can individual heterogeneity influence self-organised patterns in the termite nest construction model? Swarm Intell. (2017)
9. Kirstin, P., Radhika, N.: Complex design by simple robots: a collective embodied intelligence approach to construction. Archit. Des. **87**(4), 44–49 (2017)
10. Le, Q.V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., Ng, A.Y.: On optimization methods for deep learning

Combined Feature Extraction for Multi-view Gender Recognition

A. Annie Micheal and P. Geetha

Abstract Automatic gender classification is accepting much research enthusiasm because of its numerous applications. Perceiving the sex of a person is a challenging task because of huge variations in posture, illumination, occlusion, scaling, and facial expression. In this manuscript, we propose a novel approach for multi-view gender recognition by considering the facial shape and texture features. Dominant rotated local binary pattern (DRLBP) and rotation invariant local phase quantization (RILPQ) descriptors are utilized for extracting texture features. Pyramid histogram of oriented gradient (PHOG) descriptors is utilized for extracting shape feature. RBF kernel SVM is used to classify the gender classes, namely male and female. Experimental results indicate that the fusion of DRLBP, RILPQ, and PHOG utilizing an SVM with RBF kernel outperforms state of the art on the LFW, Adience, and FEI face database.

1 Introduction

Gender recognition is to perceive the sex of a human in view of the characteristics differentiating between manliness and womanliness. It is treated as a binary class problem (male vs. female), wherein the given face picture is assigned to any of the classes. Gender recognition has numerous applications, for example, surveillance system, biometric identification, security control. Gender recognition techniques generally done either based on appearance-based approach and feature-based approach. In the appearance-based approach, the complete face picture is utilized as the feature for classification. In feature-based approach, from the facial image,

A. Annie Micheal (✉) · P. Geetha
Department of Information Science and Technology,
College of Engineering Guindy, Anna University, Chennai, India
e-mail: annymick@gmail.com

P. Geetha
e-mail: geethap@annauniv.edu

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_22

features like Haar, Gabor, or local binary pattern (LBP) are extracted. It is a challenging task to perceive the sex of a person due to different factors, for instance, varying poses, facial expression, occlusion, illumination changes, and image background. Among these factors, multi-view gender classification is a most difficult issue. In our work, an efficient multi-view gender classification method considering the shape and texture features is proposed. DRLBP and RILPQ descriptors are utilized for extracting texture features. PHOG descriptor is utilized for extracting shape feature. Two predefined classes, namely male and female, are classified by adopting RBF kernel SVM classifier.

The remaining of this manuscript is structured as follows. In Sect. 2, some related works are reviewed. In Sect. 3, the system architecture for multi-view gender recognition is presented. Section 4 presents the proposed approach. Section 5 describes the experimental results and analysis. Section 6 includes the conclusion of the paper.

2 Related Work

Multi-view gender recognition is becoming the hottest research topic. Shakhnarovich et al. extracted Haar features from 3500 images under 30° frontal orientation and accomplished the performance of 79.0% using AdaBoost [1]. For multi-view system, H.C Lian et al. [2], local binary pattern (LBP) features, were extracted and support vector machine (SVM) was utilized for classification and obtained a classification rate of 96.75%. In [3], Block-based LBP was used by Mäkinen, E et al. and SVM classifier was used to recognize the gender of a person. Two databases, namely FERET and faces, gathered from www with and without hair were utilized for the experiment. An accuracy of 92 and 84% was accomplished for FERET database and www images. LBP features' dimension was reduced in [4] and the neighbors were partitioned into the cross and diagonal neighbors which improved the gender classification system. Ninety-two percent of accuracy was attained for FERET database. Using front face images, the success of LBP was established in [5] for gender recognition. The LBP is not robust to viewpoints changes and rotations. In [6], Gabor filters and histogram of oriented gradients (HOG) were used for multi-pose gender classification. HOG descriptor extracts the shape feature, which was primarily designed for detecting the pedestrian. In [7], Ren used a fusion of scale-invariant feature transform (SIFT), HOG features, and Gabor wavelet for gender recognition and attained an accuracy of 98%. HOG computational cost is less but not robust to the rotation. H.-T. Nguyen [8] combined two local features, namely elliptical local binary patterns (ELBPs) and a local phase quantization (LPQ). For a given input, these two features were applied which produces two output images. The histogram is obtained by dividing each output image into disjoint rectangular sub-regions. All the histograms were concatenated to produce a

combined local patterns (CLP) vector. This vector comprises both ELBP and LPQ patterns. The sex of a person is recognized from these test images by utilizing binary SVM classifier. A gender classification algorithm should be robust to variations in pose, which is still a challenging issue regarding accuracy.

3 The Proposed Methodology

The proposed framework for multi-view gender recognition is displayed in Fig. 1.

3.1 Face Detection

The normalized pixel difference (NPD) feature [9] can be computed using Eq. (1)

$$f(a,b) = (a-b)/(a+b). \tag{1}$$

where a, b $\geq$ 0 are the two pixels values in an image. Three kinds of equations can be studied for the NPD feature

$$-1 \leq \frac{a-b}{a+b} \leq 0. \tag{2}$$

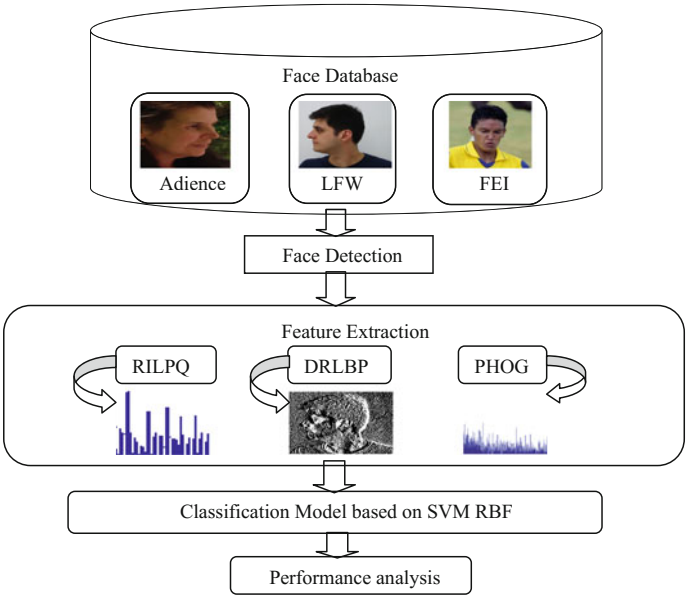


Fig. 1 System architecture for multi-view gender recognition

$$0 \leq \frac{a-b}{a+b} \leq 1. \tag{3}$$

$$\theta_1 \leq \frac{a-b}{a+b} \leq \theta_2. \tag{4}$$

where $\theta_1 < 0$ and $\theta_2 > 0$. Equation (2) is applicable if the pixel *a's* value is lesser than the pixel *b's* value, and (3) is applicable if the pixel *a's* value is greater than the pixel *b's* value. There will be a prominent edge between pixels *a* and *b*, if (4) does not hold. But the background pixel can be darker or brighter than the face. This can be studied by a quadratic splitting. Deep quadratic (DQ) tree divides the whole face manifold into several sub-fold manifolds. Four NPD features are required to construct the tree, so the maximum depth of the tree is four. The DQ tree is learned as the weak classifier in each iteration. Discrete AdaBoost algorithm is utilized to construct strong classifiers. The AdaBoost classifier learns a threshold for rejecting non-faces. The detected face is cropped and converted to the grayscale image. The feature extraction techniques are applied to this grayscale image.

3.2 Dominant Rotated Local Binary Pattern

The texture descriptor proposed in this manuscript depends on the LBP [2].The LBP value of the notable center pixel in the neighborhood is obtained by converting the binary code into a decimal number as shown in Fig. 2.

In DRLBP, the order of weights is not fixed but adaptive based on the dominant direction. In *dominant direction,* the difference between the central pixel and its neighbors is maximum as shown in Eq. (5):

$$D = \arg\max_{i \in (0,1 \ldots N-1)} |p_n - p_c|. \tag{5}$$

The weights are rotated with regard to the *dominant direction*, so it is termed as rotated local binary pattern (RLBP). The RLBP operator is described in Eq. (6)

$$RLBP_{R,N} = \sum_{i=0}^{N-1} a(p_n - p_c) \cdot 2^{\mathrm{mod}(i-D,N)}. \tag{6}$$

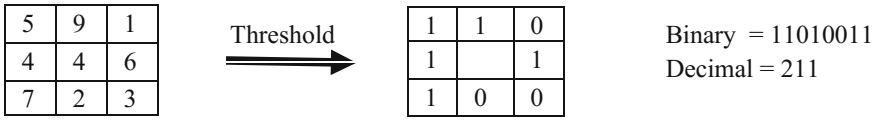


Fig. 2 Calculating LBP value

where *mod* is the modulus operator. 2*mod* (i − *D*, N) indicates the weight which relies upon *D*. The *mod* operator circularly moves the weights with regard to the dominant direction, but the sequence of the weights remains the same.

3.3 Pyramid of Histogram of Gradient

PHOG descriptor depends on the idea of HOG. Several HOGs are combined to form the PHOG descriptor. The image is segmented into regions which are denoted as pyramid levels. Canny edge detector is utilized to produce edges for the entire image. A $3 * 3$ Sobel mask is utilized to obtain the image gradients, *Ga* and *Gb* in horizontal and vertical directions. The gradient orientation θ and magnitude G_C are calculated in every edge pixels using Eqs. (7) and (8):

$$\theta = \tan^{-1}(Gb/Ga). \tag{7}$$

$$\mathrm{Gc} = \sqrt{Ga^2 + Gb^2}. \tag{8}$$

The orientations are mapped to a specific range using Eq. (9), depending on the number of bins utilized for the histogram.

$$O_B = (\theta/(D/(e - bins))). \tag{9}$$

where *D* is the predefined angle for mapping the edge orientation histograms. The weights of gradient magnitude are summed at a dense spatial grid and the weighted votes are joined together into cells of orientation bins. The resulting gradient histograms and edge orientation are concatenated into a feature descriptor at each spatial pyramid.

3.4 Rotation Invariant Local Phase Quantization

RILPQ comprises of two phases: calculating the characteristic orientation and extraction of directed descriptor. For every pixel position *a* in an image *f (a)*, the local D-by-D neighborhood N_a is rotated at each location to the characteristic orientation. Let M_θ be the rotation matrix with regard to angle θ. On a circle of radius r, the local Fourier coefficients are calculated using Eq. (10) at frequencies $S_i = r[\cos(\phi_i)\ \sin(\phi_i)]^V$, where $\phi_i = 2\pi i/D$ and i = 0,..,D–1

$$F(b, a) = \sum_{c \in N_a} f(c) w_M(c - a) e^{-j2\pi b^V c}. \tag{10}$$

where b is the frequency, and w (a) is the circular Gaussian window function defined as

$w(a) = 1/(2\pi\sigma^2 \exp(-aa^v/2\pi\sigma^2)$, if a $\leq N_G/2$ and 0 otherwise In case of rotation, for every pixel position a this results in a vector

$S(a) = \lfloor F(S_0, a'), \ldots, F(S_{D-1}, a') \rfloor$. The signs of the imaginary part of S(a) are only considered to quantify and calculate the characteristic orientation. From the quantized coefficient, the characteristic coefficient is estimated using complex moments as $o(a) = \sum_{i-0}^{D-1} k_i e^{j\phi_i}$ where k_i is the ith quantized component of K(a). The binary descriptors of the RILPQ are extracted next. The neighborhood at each location is rotated by utilizing the Eq. (11) with regard to the direction of the characteristic orientation,

$$F_\xi(b, a) = \sum_c f(c)' w_M(M_{\xi(a)}^{-1}(c-a)) e^{-j2\pi a^{vM_{\xi(a)}^{-1}, c}}. \tag{11}$$

A vector results for each pixel position as shown in Eq. (12)

$$F_\xi(a) = \lfloor F_\xi(b_1, a), F_\xi(b_2, a), F_\xi(b_3, a), F_\xi(b_4, a) \rfloor \tag{12}$$

The phase information in the Fourier coefficients is noted by watching the signs of the real and imaginary parts of each component in $F_\xi(a)$. This is accomplished by utilizing a simple scalar quantization $u_i = 1$, if $x_i \geq 0$, else 0, where x_j is the jth component of the vector U(a) = [Re{F(a), Im{F(a)}]. The eight binary coefficients u_j(a) are resulted utilizing simple binary coding $f_{RI-LPQ}(a) = \sum_{j=1}^{8} u_j 2^{j-1}$ which is in the range of 0–255. The histogram values are calculated for all the positions and it is utilized as a 256 feature vector. Finally, we fusion all the three descriptors to form a fused feature vector which is given as in input to SVM-based RBF kernel. SVM-based RBF classifies the features into their respective classes.

4 Experimental Results and Analysis

4.1 Implementation Details and Analysis

Labeled faces in the wild (LFW), FEI, and Adience are the benchmark datasets for gender recognition and it consists of images captured in the wild. The LFW database totally comprises 13,233 images with 5749 images of individuals (4263 male, 1486 female). 1680 of the people pictured have two or more distinct photos in the dataset. Adience database has 26,580 images (of 2284 individuals). The sources

of the images included in this dataset are Flickr albums, assembled by automatic upload from iPhone5 (or later) smart-phone devices. FEI face database consists of 2800 images with 200 individuals. All images are colorful and taken against a white homogenous background in an upright frontal position with profile rotation of up to about 180°. MATLAB 2015a version is utilized for implementing the proposed approach. The training phase comprises of 2755 female and 3280 male faces from LFW database, 1176 male and 924 female faces from FEI and 2139 male, and 2108 female faces from Adience database are used. The face is detected using NDP features and discrete AdaBoost from the given facial image. The detected face is cropped and resized to 256 × 256 pixels. We experimented with various image sizes, but the accuracy of the system was not up to the satisfaction because increasing or decreasing the image size decreases the systems' accuracy. The resized image is converted to grayscale image. The DRLBP, PHOG, and RILPQ feature extraction techniques are applied to this gray scale image. For the DRLBP, two radius values (R = 2, 3) and two neighbors (P = 8, 16) are considered. For the PHOG, the pyramid level L, binning Histogram H and angle θ are set as L = 2, 3, H = 8, 16 and θ = 360. For RILPQ, the window size (M) = 2, 3, 4, 5 is considered. In order to perform classification, 75% samples are utilized for training and the others are utilized for validation. In addition, SVM with three different kernels are experimented on each feature descriptor, and the results are conferred in Tables 1, 2, and 3. The SVM-RBF kernel is found more suitable for transforming the data to high-dimensional space, as it to improve the performance of the system when compared to other approaches as shown in Table 4. Figure 3 represents the accuracy compared with three SVM kernel functions for the combination of three features on LFW, FEI, and Adience dataset.

Table 1 Gender recognition results for PHOG descriptor with kernel SVM on FEI, LFW, and Adience database

PHOG parameter	Recognition rate (%)								
	FEI			LFW			Adience		
	SVM linear	SVM poly	SVM RBF	SVM linear	SVM poly	SVM RBF	SVM linear	SVM poly	SVM RBF
L = 2, H = 8	80.4	75.7	82.9	79.4	82.5	83.6	65.1	67.2	69.3
L = 2, H = 16	83.1	78.2	85.3	77	82.6	84.2	68.3	70.2	75.4
L = 3, H = 8	87.2	81.8	88.4	82.7	85.6	86.8	71.9	73.4	77.5
L = 3, H = 16	89.5	85.4	**91.0**	81.9	88.1	**88.5**	73.5	75.6	**79.5**

Table 2 Gender recognition results for DRLBP descriptor with kernel SVM on FEI, LFW, and Adience database

DRLBP parameter	Recognition rate (%)								
	FEI			LFW			Adience		
	SVM linear	SVM poly	SVM RBF	SVM linear	SVM poly	SVM RBF	SVM linear	SVM poly	SVM RBF
R = 2, H = 8	70.2	73.6	75.5	71.6	65.2	60.7	66.3	58.2	59.4
R = 2, H = 16	75.3	77.4	79.1	76.9	69.2	64.5	70.1	61.6	61.3
R = 3, H = 8	78.3	80.1	82.7	79.1	71.4	67.1	73.2	63.7	65.0
R = 3, H = 16	80.8	82.1	**84.2**	81.9	74.0	**68.9**	75.1	65.8	**67.4**

Table 3 Gender recognition results for RILPQ descriptor with kernel SVM on FEI, LFW, and Adience database

RILPQ window size	Recognition rate (%)								
	FEI			LFW			Adience		
	SVM linear	SVM poly	SVM RBF	SVM linear	SVM poly	SVM RBF	SVM linear	*SVM poly*	*SVM RBF*
2 × 2	79.2	78.2	80.5	68.3	63.2	69.1	57.1	75.9	61.9
3 × 3	80.4	80.2	83.5	71.7	68.0	72.7	60.4	77.0	65.0
4 × 4	85.3	82.4	86.2	73.5	70.6	75.3	63.2	60.6	68.2
5 × 5	88.0	84.8	**88.4**	75.1	73.8	**77.4**	65.5	63.2	**70.4**

Table 4 Gender recognition results for three features extraction on FEI, LFW, and Adience database

Features	Parameter	Recognition rate (%)								
		FEI			LFW			Adience		
		SVM linear	SVM poly	SVM RBF	SVM linear	SVM poly	SVM RBF	SVM linear	*SVM poly*	*SVM RBF*
DRLBP+ PHOG+ RILPQ	R = 3, H = 16 L = 3, H = 16 5 × 5	80.9	76.8	**95.3**	85.2	80.6	**98.7**	84.8	81.4	**96.0**

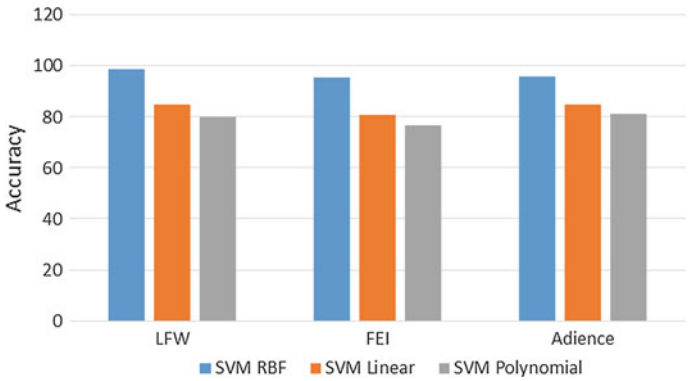


Fig. 3 Accuracy compared with three SVM kernel functions for the combination of three features on LFW, FEI, and Adience dataset

5 Conclusion

We have proposed an approach for gender recognition from unconstrained face images using facial shape and texture description and kernel-based SVM. Our method adopts DRLBP and RILPQ texture descriptor and PHOG shape descriptor with support vector machine for classification. The main objective of this paper was to determine the most suitable parameter combinations for performing gender recognition on unconstrained face image. The experimental results indicated that PHOG descriptor with pyramid level $L = 3$ and binning histogram $H = 16$, DRLBP descriptor with radius $= 3$ and neighbors $= 16$ and RILPQ descriptor with window size 5 combined with RBF kernel SVM has attained an optimum performance of 98.7, 95.3, and 96% for LFW, FEI, and Adience dataset. The results obtained demonstrate that the proposed descriptor outperforms the state-of-the-art approaches.

References

1. Shakhnarovich, G., Viola, P.A., Moghaddam, B.: A unified learning framework for real-time face detection and classification. In: IEEE International Conference on Automatic Face and Gesture Recognition (2002)
2. Lian, H.C., Lu, B.L.: Multi-view gender classification using local binary patterns and support vector machines. In: Advances in Neural Networks-ISNN 2006, LNCS, vol. 3972, pp. 202–209 (2006)
3. Mäkinen, E., Raisamo, R.: An experimental comparison of gender classification methods. Pattern Recogn. Lett. **29**(10), 544– 1556 (2008)
4. Fang, Y., Wang, Z.: Improving LBP features for gender classification. In: International Conference on Wavelet Analysis and Pattern Recognition, pp. 373 − 377. Hong Kong (2008)

5. Yang, Z., Ai, H.: Demographic classification with local binary patterns. In: Lee, S.-W., Li, S. (eds.) Advances in Biometrics, pp. 464–473. Springer, Berlin (2007)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 886–893 (2005)
7. Ren H., Li, Z.-N.: Gender recognition using complexity-aware local features. In: 22nd International Conference on Pattern Recognition (ICPR), pp. 2389–2394 (2014)
8. Nguyen, H.-T.: Combining local features for gender classification. In: 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (2015)
9. Annie Micheal, A., Geetha, P.: Multi-view face detection using normalized pixel difference feature. In: Proceedings of the Sixth International Conference on Communication and Signal Processing. Adhiparasakthi Engineering College, Tamil Nadu (2017)

Resource Allocation Using Modified Banker's Algorithm for Next-Generation Wireless Networks

Patil Harshali and Purohit Seema

Abstract Active wireless users are increasing enormously every day. Past two decades active users' statistics reveal the need for more radio resources. Radio resource is a limited resource, and hence, the utilization of it in an effective manner is a key point in radio resource management (RRM). This paper discusses the approach of resource allocation to requesting users using a modified banker's algorithm (MBA). As an output, the algorithm provides an allocation sequence to various users without causing any interference to the existing user communication. The suggested method uses the parameters such as allocation of resources to existing users, utmost requirement of the users for seamless communication, number of available resources specific to the given environment, requests arriving from various users, and the priority for communication channel.

1 Introduction

With the extensive use of smartphones and tablets running bandwidth-greedy applications, the data access is expected 100–1000 times in the near future. Subsequently, the next-generation networks also need to advance with this growing demand. The technologies used to meet these challenges are ultra-wideband and WiGig. In ultra-wideband (UWB), the signals can be spread across the bandwidth around 3 GHz and 10 GHz.

In wireless local area networks in 4G, 60 GHz spectrum has been considered for development. The unlicensed spectrum (Wi-Fi bands) 2.4 GHz and 5 GHz bands are

P. Harshali (✉)
MET Institute of Computer Science, Bandra (W), Mumbai, India
e-mail: harshalip_ics@met.edu

P. Harshali
University Department of Computer Science, University of Mumbai, Mumbai, India

P. Seema
Mathematics Department, Kirti College, Dadar (W), Mumbai, India
e-mail: supurohit@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_23

available. The 60 GHz band is available in WLAN and fast session transfer is possible between 60GHz and 2.5 GHz/5 GHz. IEEE 802.11ac is known as Wi-Fi, which offers high throughput wireless local area networks (WLANs) on the 5 GHz band. IEEE 802.11ad is known as WiGig. The IEEE 802.11ad provides data rate more than 10 times of IEEE 802. 11 standard, i.e.; 802.11ad provides up to 7Gbps. The tri-band networking, wireless docking, increased data transfer rate, and uncompressed video streaming are the basis of the required change in IEEE 802.11 communication standards. As per the Shannon theory, the maximum transmission rate in bps = B $\log_2(1 + S/N)$ where S/N is the signal-to-noise ratio. B is the channel bandwidth. If the S/N rate is fixed, and if the bandwidth B is increased, then the transmission rate is also increased. Alternatively, for a relatively large bandwidth B, the S/N can be reduced to reduce or maintain the transmission rate. The bandwidth of the signal can be increased by "spreading" it over wider frequency block.

In next-generation wireless broadcast system, different types of broadcast signals are transmitted over a carrier frequency using time division multiplexing technique. The advancement in next-generation wireless network includes frequency hopping transmissions over multiple carrier frequencies, which achieves coverage gain and capacity gain [1, 2].

2 Background

Real-world applications of routing algorithms are very complex. In such applications, heterogeneous, highly dynamic, and potentially noncooperative environments are captured and processed. The multiagent systems are an emerging choice for modeling such system. In [3], the authors describe an ejection pool principle-based algorithm. The improvement in algorithm uses local search strategy for the ejections [4]. For effective resource allocation, consider an agent-based algorithm for vehicle routing problem with time windows. It is based on the concepts of shipping company and underlying shipping company truck. The planning is based on contract net protocol (CNP) escorted by a "simulated trading" enhancement strategy based on finding the optimal customer exchanges by solving a maximal pairing problem on a graph representing the proposed exchanges [5].

In [6], the author presented a generic model for routing agents through an infrastructure graph. A resource graph is generated whose nodes correspond to the nodes and edges of the infrastructure graph to calculate the route. The edges of the resource graph can be read as a successor relation.

In [7], the paper presents a modeling approach that applies the time window routing method to shuttle systems. The suggested model for the system is a mixed graph and presents a concept for moving vehicles safely and efficiently through the storage system.

Deadlock-free routes for nodes requesting resources are modeled by a graph using the time window routing method. For each node, the algorithm maintains a list of free time windows, i.e., spectrum free slots through which secondary nodes

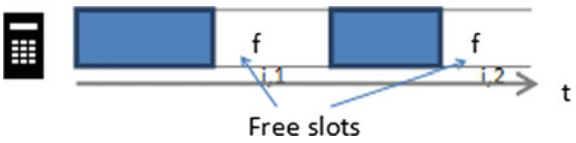


Fig. 1 Free time window for frequency channel f_i

resource allocation requirements can be fulfilled. Each free slot/hole is defined by the end of the preceding reserved time window for the primary user and the beginning of the subsequent reserved time window of the allotted node either primary or secondary as depicted in Fig. 1. In the following, figures $f_{i,1}$ and $f_{i,2}$ represent free slots of the given channel, whereas filled portion is reserved time of the given channel.

3 System Model

The simplified model for radio resource allocation is discussed below. The resource requirement of secondary users is modeled as spectrum band as a resource. Different secondary users try to occupy the vacant slots/bands in the system.

Banker's algorithm is used in operating system for deadlock avoidance. It identifies whether system will be in safe state or not. Banker's algorithm requires prior knowledge of resource requirements by each process to identify whether deadlock situation will occur or not. In wireless domain, the similar concept is used for fixed channel allocation model. Number channels available in spectrum are prior known, and hence, the maximum availability of channel at any instance of time is known by environment administrator/monitor. The users requesting for channel for communication are treated as processes which request for resources. The numbers associated with user indicates the number of processes and the requirement of each user for free channel for communication. The radio spectrum is underutilized, and hence, to utilize it effectively, the free channel can be allotted to the requesting users. Banker's algorithm has limitation that process requirement must be prior known, which is not available in all cases. Instead of applying the banker's algorithm, slight modification is applied for resource allocation and verifying whether the system works fine or not in the given scenario.

The modified banker's algorithm requires following structural components.

Table 1 describes the components of banker's and modified banker's algorithm. The user needs certain number of channels for communication; these are defined by the need.

Table 1 Banker's algorithm components

Component	Representation
Process	User
Allocation matrix	Current status
Maximum	Max channel requirement
Need	Resource requirement

3.1 The Description of Banker's Algorithm

In banker's algorithm, a set of resources that are available with the system are represented as R, a sequence of processes {Pi} ∀i = 1, 3…., n, and a set of all resources currently held by processes P_1, P_2,…, P_{i-1} are represented by allocation of resources to processes. To remain safe, each process P_i must request the resources from available resource classes/type [8, 9].

The necessary and sufficient condition for a system to remain in a safe state is that each sequence $\{P_i\}\forall$ i = 1, 3…., n is safe. Safe sequence is an indication that no cycle exists, and hence, there will not be a deadlock in system. Banker's algorithm data structure can be described as follows:

For n processes P_1, P_2,…., P_n, m resource classes or types R_l, R_2,…, R_m, available resources are V_I, V_2,…, V_m, and resource requests Q_l, Q_2,…, Q_m. A temporary variable work holds the availability of resource class/type instances. Finish vector keeps track of visiting the requested resource once during the process of resource allocation.

Finish[i] = false, (i = l, 2,…, n.) // For all processes, set the process execution is not completed by setting finish flag as false.

$$\text{Max} = \begin{pmatrix} R_{11} & R_{12} & \dots & R_{1m} \\ R_{21} & R_{22} & \dots & R_{2m} \\ R_{31} & R_{32} & \dots & R_{3m} \\ \dots & \dots & \dots & \dots \\ R_{n1} & R_{n2} & \dots & R_{nm} \end{pmatrix}$$

$$\text{Allocation} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ A_{31} & A_{32} & \dots & A_{3m} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nm} \end{pmatrix}$$

$$\text{Need} = \begin{pmatrix} N_{11} & N_{12} & \dots & N_{1m} \\ N_{21} & N_{22} & \dots & N_{2m} \\ N_{31} & N_{32} & \dots & N_{3m} \\ \dots & \dots & \dots & \dots \\ N_{n1} & N_{n2} & \dots & N_{nm} \end{pmatrix}$$

Resource class/type vector represents the number of different types' resource instances in the system. It is a maximum vector of resource instances.

Available resources vector indicates the number of available resources in the system at present.

Request vector is of process/user Pi. Max matrix represents a maximum request matrix which describes each process's/user's maximum request for resources.

Allocation matrix defines the current allocation state. This vector represents the number of resource instances assigned to each process currently.

Need matrix represents the requirement of number of resource instances for each process. Therefore, the need matrix calculate as follows

Need[i, j] = Max[i, j] - allocation[i, j].

3.2 Algorithm: Banker's Algorithm

Input: Processes P1, P2,…Pn
Request[i], a resource request by process Pi
Output: Safe sequence of Process execution. All process is able to complete the execution.
Temporary variable: Work[i]
Flag variable: finish[i]
Procedure variable: allocation[i]
Computational variable: need[i] (process Pi's need)

Procedure Banker's Algorithm

1) if Request[i] > Need[i]
// Resource request from Pi exceed its maximum demand
Display error message and exit
2) if Request[i] > Available
//Reject process P_i's request as adequate number of resources are not available for //this request in the system
Display Error message and exit
3) If above conditions are not satisfied then assign requested resources, modify the corresponding data structures and vectors/arrays as per the request entry.
Available -= Request[i]
Allocation[i] += Request[i]
Need[i] -= Request[i]
4) Check for safety
Assign the available resource instances to work vector
Set the process execution flag information for process Pi as False.
When ((Finish[i] = = false) and (Need[i]≤Work))
5) Find Pi which can complete execution or finish
Work += Allocation[i]
// Pi free all assigned resources hold by it
Set the process execution flag information for process Pi as True.
Repeat step number 5.
6) if (Finish[i] = = true) for all then
Assign resources
else
Reject request of P_i.

In this paper, we consider the banker's algorithm using time resource request. The resource allocation issues of multiple broadcast signals over multiple carrier frequencies in next-generation signal communications are discussed here. Fixed broadcasting with single antenna, fixed broadcasting with multiple antennas, mobile broadcasting with single antenna, and mobile broadcasting with multiple antennas are the different types of broadcasting signals. Fixed broadcasting usually has fewer requirements on frequency hopping, whereas mobile broadcasting strives for improving the transmission robustness. The demand for multiple carrier frequencies is more in mobile broadcasting as compare to fixed broadcasting.

In communication, different types of signals are used with different resource requirements such as time and number of carrier frequencies. The resources are limited hence, effective utilization is necessary. Banker's algorithm is used for deadlock avoidance, and for that it checks whether the resource allocation is leading the system in safe or unsafe state. If the resource allocation causes the system to be unsafe state, then that request will be postponed by banker's algorithm. As the resource requirement are satisfied, the processes return the resources which will not be possible in the next-generation wireless network resources, so the algorithm needs to adjust the required resources instead of postponing the requirement. To fulfill this need, we need a modified banker's algorithm. The main challenge is to assign the requested time resources, with required number of carrier frequencies.

In modified banker's algorithm, the safety judgment is specified by two flags flag_sT1 and flag_sT2. Constraints applied to the algorithm are the time resource requirement must be equal to or greater than minimum frame length, i.e., flag_sT1 = 1, and the other constraint is after allocation of time resource to certain carrier frequency, the remaining total time resource must be equal to or greater than minimum frame length, i.e., flag_sT2 = 1. If the constraint is not satisfied, then set the corresponding flag to 0, i.e., for constraint 1, flag_sT1 = 0 and for constraint 2, flag_sT2 = 0 [2].

3.3 Algorithm: Resource Allocation

Consider K requests are provided for time T.

(1) Arrange the time resource requirements in descending order of all next-generation broadcast network signal.
(2) Select the least frequently used carrier frequencies for the broadcast signal communication. Thus, one by one the broadcast signals will be communicated.
(3) If no other signal for communication, then assign all time resources to the corresponding carrier frequency of current broadcast.
(4) For the remaining carrier frequencies of the broadcast signal, distribute the available time resources.
(5) Safety criteria: check whether the current broadcast signal allocation leads to safe state or not. In other words, if both the flags falg_sT1 and flag_sT2 is set to 1,

then the current resource allocation is successful. If all signal resource requirements are satisfied, i.e., K is achieved, then the process ends otherwise go to next step. If both flags are set 0, then resource allocation fails and exits the process otherwise go to step (7).

(6) Allocate resources (free/nonallotted time of carrier frequencies) to the unhandled broadcast signal.

(7) For any of carrier frequency of broadcast signal with either flag_sT1 = 0 or flag_sT2 = 0, the allotted resources are adjusted. If flag_sT1 = 0, then allot minimum frame length time resource to current signal request otherwise if flag_sT2 = 0, then allot the remaining time resource of nonallocated carrier frequencies. Then perform step (4)

The modified banker algorithm is actually applied in step (5).

Multiple carriers are used, and requirements of each user are identified; using this information, allocations of resources can be obtained from above-discussed algorithm.

Let us assume a sample case; let us consider four next-generation network broadcasting types, and they are sharing five carrier frequencies. The time resource distribution cycle is 250 ms. For the broadcast signals, we need time resources in descending order as R1 = 400 ms, R2 = 350 ms, R3 = 300 ms, and R4 = 200 ms. The demand for the number of carrier frequencies by the signals is need 1 = 4, need 2 = 5, need 3 = 3, and need 4 = 2, respectively. For example, for the fixed broadcast signals, the default minimum frame length is set as 40 ms and for mobile broadcast signal is 30 ms. Hence, we need to allot the slot such that they satisfy the constraints. As per MBA, the allocation of resources is depicted in Fig. 2.

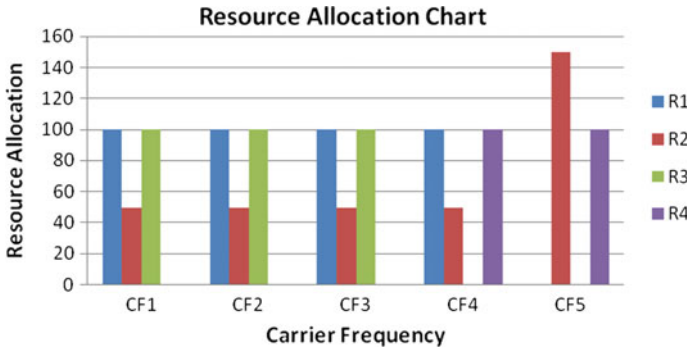


Fig. 2 Allocation of resources using MBA

4 Conclusion

As per modified banker's algorithm, the safety check needs to be done strictly if there is a change in parameter values and resources. If unstable state is caused by parameter value change, then adjustment in resource allocation needs to be done. If the minimum frame length value and the need of next-generation network signal are relatively small, then there are fewer chances of unsafe state occurrences in the system. If numbers of carrier frequencies are increased, then the possibility of successful resource allocation using the existing algorithm becomes a complex job. The modified banker's algorithm will work fine with the constraints of minimum frame length and relatively small number of carrier frequencies. Seamless communication can be possible on different channels for various users using the modified algorithm.

References

1. Digital Video Broadcasting (DVB); Transmission System for Hand held Terminals (DVB-H), ETSI EN 302 304 (2004)
2. Wang, H., Tian, L., Li, M., Mu, W., Gao, X.C.: Banker's algorithm based resource allocation in next generation broadcasting wireless systems. In: 10th International Conference on Communications and Networking in China (China Com) (2015)
3. Lim, A., Zhang, X.: A two-stage heuristic with ejection pools and generalized ejection chains for the vehicle routing problem with time windows. INFORMS J. Comput. **19**(3), 443–457 (2007)
4. Nagata, Y., Br¨aysy, O.: A powerful route minimization heuristic for the vehicle routing problem with time windows. Oper. Res. Lett. **37**(5), 333–338 (2009)
5. Kalina, P., Vok˘rínek, J.: Algorithm for Vehicle Routing Problem with Time Windows Based on Agent Negotiation
6. Ter Mors, A.W.: The world according to MARP. Dissertation. Technische Universiteit Delft (2010)
7. Lienert, T., Fottner, J.: No more deadlocks – applying the time window routing method to shuttle systems. In: Proceedings 31st European Conference on Modelling and Simulation. ISBN: 978-0-9932440-4-9
8. Dijkstra, E.: Cooperating sequential processes. Technical Report EWD123, Technological University, Eindhoven, The Netherlands (1965)
9. Ma, X., Yan, J.: An improved parallel Banker's algorithm based on petri net. In: International Conference on Electronic and Mechanical Engineering and Information Technology, 2011. IEEE, New York. 978-1-61284-088-8/11

Synthesis Methods of Baugh-Wooley Multiplier and Non-restoring Divider to Enhance Primitive's Results of QCA Circuits

Bandan Kumar Bhoi, Neeraj Kumar Misra, Manoranjan Pradhan and Rashmishree Rout

Abstract As a semiconductor industry continues growing toward miniaturization and high speed, it is challenged by the rising uncertainties in the scaling for further devices shrink in the nanometer scale. Scaling leads to quantum effect at the nanoscale. Quantum dot cellular automata (QCA) is the alternative approach to synthesize the digital logic circuits with high density and high computation speed. In this paper, an accurate approach to synthesize and optimize the Baugh-Wooley multiplier and non-restoring divider in the presence of QCA technology has been proposed. The proposed designs are robust and utilize a wire-crossing type of single layer, with minimal clock phasing. The synthesis approach and optimization are perfectly scalable across layout construction of designs and can find better primitive's results of QCA circuit performance.

1 Introduction

There has been unprecedented growth in the era of nanoelectronics because of more research and development in this era. QCA attempt a nanoscale design and aims a new technique of logic computation and information flow in imminent technology. QCA technology has nominal energy dissipation because its circuits are assembled by the quantum cells. In this aspect, QCA is an emerging nanotechnology that can reduce the transistor size to nanometers [1]. QCA involves low complexity and robust designs for various arithmetic operations because the key element of this technology is a cell instead of a transistor.

B. K. Bhoi (✉) · M. Pradhan · R. Rout
Department of Electronics & Telecommunication, Veer Surendra Sai University of Technology, Burla 768018, India
e-mail: bkbhoi_etc@vssut.ac.in

N. K. Misra
Department of Electronics Engineering, Institute of Engineering and Technology, Lucknow 226021, India

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_24

Different combinations of these cells can accomplish any Boolean logic function [2, 3, 4]. Because of these advantages, many works have been accomplished by researchers in the last decade. In this paper, an efficient multiplier such as the Baugh-Wooley multiplier is proposed that multiplies two 2's complement numbers. In addition, a non-restoring divider has been proposed in which restoration is not required. The multiplier comprises a number of full adders, whereas the non-restoring (NR) divider comprises of NR cells, and each NR cell is composed of a full adder and an Ex-OR gate. In both the designs, the basic functional unit is the full adder. In each case, one-bit full adders have been used [5, 6, 7, 8, 9]. The full adder used in these proposed circuits is very efficient, as it requires only two majority gates and few numbers of cells; hence, occupies very less area. The Ex-OR gate used along with the full adder in the NR cell is implemented using only 12 cells. It is a three-input majority gate, whose one input is set to zero, i.e., having the polarization −1.00. Both the designs are verified using QCA designer simulator.

The rest of this paper is organized as follows. Section 2 introduces the basics of QCA. Section 3 presents the synthesis model of multiplier and divider circuits based on QCA technology. Comparison results are presented in Sect. 4. Finally, the concludes in Sect. 5.

2 Basics of QCA

Each individual cell has four quantum dots located at the corners of the square-shaped cell. These dots are coupled by tunnel barriers in such a manner that the electrons can tunnel between the dots but cannot leave the cell, making conduction possible between cells as no physical transport of charge takes place. So there is no leakage current issue in QCA. The cell works on coulombic interaction. When a small amount of negative potential is applied to a lead near the quantum dot containing an electron, the next cell changes its ground states to decrease the coulombic repulsion. There are basically two states of polarization of a cell, namely logic "0" and logic "1".

The logic "0" has the polarization of −1.00, and logic '1" has the polarization of 1.00 (Fig. 1a). The cells interact only with their nearest neighbors. One position change in electrons in quantum dots leads to transfer of information from one end to the other. There are two fundamental gates in QCA, i.e., (a) majority gate (b) inverter. With the help of these two gates, almost every Boolean function can be implemented. However, this requires proper clocking of cells in an array. The logic function implemented by the three-input majority gate is $Y = AB + BC + AC$ where Y is the output of the majority gate, and A, B, C are the three inputs (Fig. 1b). If any of its inputs is set to logic "0", then the majority gate will behave as an AND gate, i.e., $M(A, B, 0) = AB$, and if is set to logic "1", then it will work as

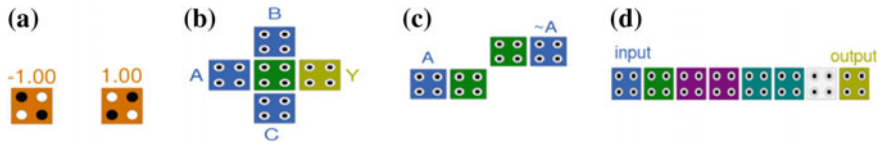


Fig. 1 QCA basics **a** QCA cell **b** majority gate **c** inverter **d** binary wire

an OR gate, i.e., M(A, B, 1) = A+B [3, 4]. When two cells are placed diagonally, they perform the NOT operation (Fig. 1c). The binary wire helps to connect different gates in various manners to implement various Boolean functions (Fig. 1d). Clocking decides the particular time when the input will reach the required cell. There are four types of clock known as clock zones. These zones are clock 0, clock 1, clock 2, and clock 3. The cells in clock 0 are represented in green, clock 1 in magenta, clock 2 in blue, and clock 3 in white color. The clocking of cells in an array should be done in an orderly manner, i.e., first clock 0, then clock 1, clock 2, and clock 3, respectively.

3 Proposed Multiplier and Divider

3.1 Proposed Layout of Multiplier

In this section, a new QCA layout of the Baugh-Wooley multiplier is proposed. This multiplier is designed for the multiplication of signed numbers. Here a 4 × 4 size Baugh-Wooley multiplier is implemented that uses only a few sets of full adders and an inverter. The full adder is the key element of the multiplier, which is realized by only two majority gates in QCA. The QCA layout of this full adder is mentioned in Fig. 2a. This feature makes the proposed multiplier more efficient than the existing one [5].

The full adder has three inputs, i.e., A, B, Ci and produces output, i.e., sum and cout. The expressions for the outputs of the full adder are SUM = A ⊕ B ⊕ Ci = M(M(A, B, Ci), M(M(A, B, Ci), B, Ci), A) and Cout = AB + BCi + ACi = M(A, B, Ci). The block diagram of the four-bit Baugh-Wooley multiplier circuit is shown in Fig. 2b. There are three layers present in this multiplier [10, 11, 12, 13, 14], where each layer comprises four numbers of full adders. The initial carry into each layer is 0. The final carry-out of each stage is fed to the next stage, and the carry-out of the last stage is inverted to produce the MSB bit of the product. Here in the proposed QCA layout of 4-bit multiplier, 32 numbers of majority gates and seven numbers of inverters are used which is given in Fig. 3.

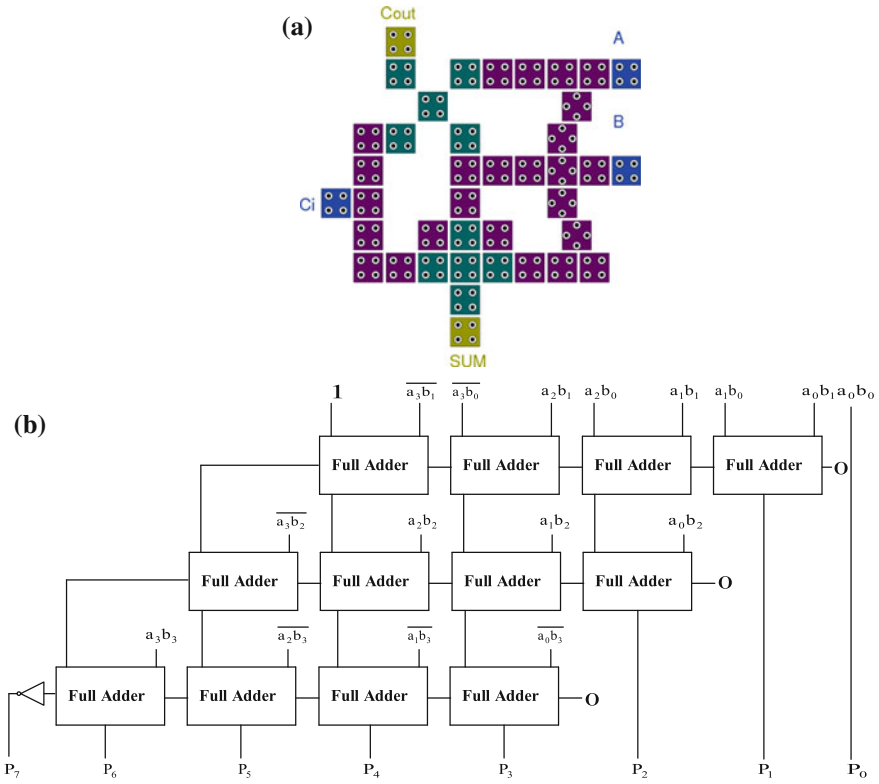


Fig. 2 (a) QCA layout of full adder (b) four-bit Baugh-Wooley multiplier circuit

3.2 Proposed Layout of Non-restoring Divider

The non-restoring divider consists of a various number of non-restoring cells [15, 16] where each cell is composed of a one-bit full adder and a two-input XOR gate [17, 18]. The full adder used here composed of only two majority gates which takes three inputs and produces two outputs is shown in Fig. 2a. Among these three inputs, one is connected to the output of the Ex-OR gate and the other two are A which is the dividend bit and Cin, i.e., the initial carry. The output produced by the full adder is the SUM which serves as the remainder bit, and Cout serves as the quotient bit in the divider. The Ex-OR gate used in the proposed one is composed of only one majority gate which takes only two inputs, i.e., T and D, and the other input is fixed at zero. The input T to the Ex-OR gate is the control input, i.e.; the output depends upon T as when T = 0, the output of Ex-OR will be D, and when T = 1, the output will be the complement of D input. The output of Ex-OR is fed to

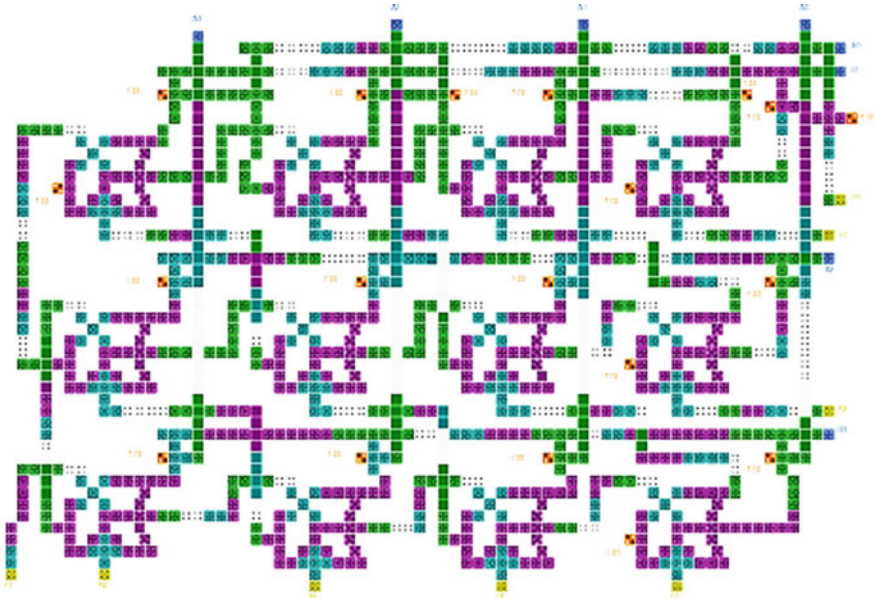


Fig. 3 QCA layout of Baugh-Wooley multiplier

one input of the full adder, thus controlling the full adder. Here the Ex-OR gate is the physical arrangement of 13 cells, and the full adder comprises 40 cells which are in different clock zones. The proposed 4 × 4 non-restoring array divider is constructed using 16 non-restoring cells as it requires a n^2 number of cells for implementing a n × n divider array [17]. Here the divider circuit consists of four layers, out of which the first layer is the main layer where all the inputs are provided, and the input signal T controls the operation of the NR(non-restoring) cells in the first array [18]. The divisor is represented as D0 D1 D2 D3, the dividend as A0 A1 A2 A3, the quotient as q0 q1 q2 q3, and the remainder as r3 r4 r5 r6. The carry-out bit q0 from the first layer behaves as the control input for the second layer, and similarly, q1 for the third layer, and q2 for the fourth layer. The QCA layouts of the proposed NR cell and four-bit divider are given in Figs. 4 and 5, respectively. All QCA circuits are designed using QCADesigner tool [19].

4 Comparison Result

The proposed multiplier and divider uses a full adder [18] which requires only 40 QCA cells to realize. This makes the proposed circuits more area efficient as its layout design occupies very less area and few numbers of cells. The Baugh-Wooley

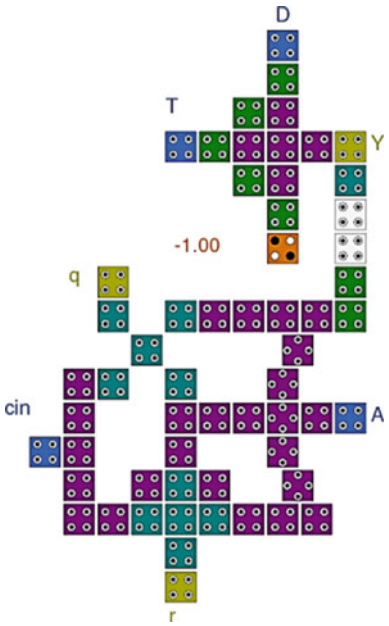


Fig. 4 QCA layout of NR Cell

multiplier proposed here contains 1330 numbers of cells and occupies an area of 1.57 μm^2. Therefore, it is clear that the multiplier proposed here is 33% and 13% more efficient in cell count and area than the existing one. The comparison is shown in Table 1 below. The non-restoring divider comprises a number of non-restoring cells that contains an Ex-OR gate and a full adder. Table 2 shows the comparison between the basic building block called NR cells of various existing designs of divider circuits.

The Ex-OR gate used here is designed only with 13 cells. So the non-restoring divider proposed here uses only 2056 number of cells and occupies 2.26 μm^2 area. This implies an improvement of 60% and 77% in cell count and area, respectively than the previous one [15]. The comparison between the proposed one and the existing one can be seen from Table 3 below which clearly indicates the improved features such as area occupation, number of the cell employed, and latency of the proposed 4 $\times$ 4 non-restoring divider.

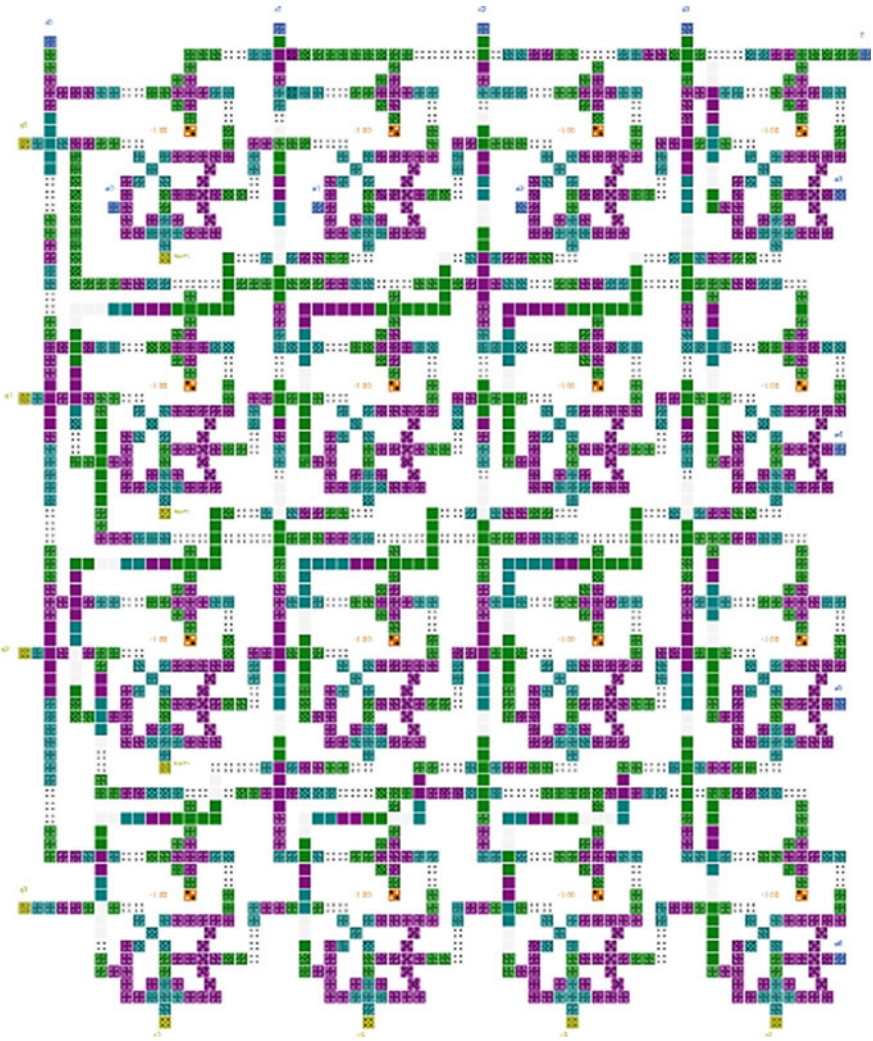


Fig. 5 QCA layout of 4 × 4 non-restoring array divider

Table 1 Comparison between multipliers

Multipliers	No. of majority gates used	No. of inverters used	Total no. of cells	Area (μm^2)
4 × 4 wallace [14]	24	13	3295	7.39
4 × 4 Dadda [14]	18	9	3384	7.51
Baugh-Wooley [14]	36	19	1982	1.8
Proposed	32	7	1330	1.57

Table 2 Comparison between NR Cells

Nr cell	Cell count	Area (In μm^2)
[15]	235	0.35
[17]	147	0.27
[18]	60	0.08
Proposed	58	0.11

Table 3 Comparison between dividers

Non-restoring dividers	No. of majority gates used	No. of inverters used	Total no. of cells	Area (μm^2)
4 × 4 NRD [15]	96	16	5124	9.99
4 × 4 NRD [17]	112	48	6865	10.95
4 × 4 NRD [18]	48	–	3180	6.5
Proposed	48	–	2054	2.26

5 Conclusion

In this paper, a Baugh-Wooley multiplier and a non-restoring divider have been proposed. The non-restoring divider is designed using a number of NR cells. Both the designs are implemented in QCADesigner. These designs can further be extended to a larger data size. The proposed multipliers can reach a cost efficient and perform well with respect to existing state-of-the-art design. The physical implementation of the design was adopted to ensure the proposed design feasibility. Moreover, a robust design based on cell coupling can be effectively realized under coplanar technique. For future work, these designs can be useful in implementing ALU with greater feature than the existing one and will also be useful in the field of cryptography.

References

1. Lent, C.S., Tougaw, P.D., Porod, W., Bernstein, G.H.: Quantum cellular automata. Nanotechnology. **4**, 49–57 (1993)
2. Misra, N.K., Sen, B., Wairya, S.: Designing conservative reversible n-bit binary comparator for emerging quantum-dot cellular automata nano circuits. J. Nanoeng. Nanomanufact. **6**(3), 201–216 (2016)
3. Tougaw, P.D., Lent, C.S.: Logical devices implemented using quantum cellular automata. J. Appl. Phys. **75**(3), 1818–1825 (1994)
4. Misra, N.K., Sen, B., Wairya, S., Bhoi, B.: Testable novel parity-preserving reversible gate and low-cost quantum decoder design in 1D molecular-QCA. J. Circuits Syst. Comput. **26**(09), 1750145 (2017)

5. Hashemi, S., Tehrani, M., Navi, K.: An efficient quantum-dot cellular automata full-adder. Sci. Res. Essays **7**(2), 177–189 (2012)
6. Kumar, D., Mitra, D.: Design of a practical fault-tolerant adder in QCA. Microelectron. J. **53**, 90–104 (2016)
7. Abedi, D., Jaberipur, G., Sangsefidi, M.: Coplanar full adder in quantum-dot cellular automata via clock-zone-based crossover. IEEE Trans. Nanotechnol. **14**(3), 497–504 (2015)
8. Perri, S., Corsonello, P., Cocorullo, G.: Area-delay efficient binary adders in QCA. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **22**(5), 1174–1179 (2014)
9. Mohammadi, M., Mohammadi, M., Gorgin, S.: An efficient design of full adder in quantum-dot cellular automata (QCA) technology. Microelectron. J. **50**, 35–43 (2016)
10. Cho, H., Swartzlander, E.: Adder and multiplier design in quantum-dot cellular automata. IEEE Trans. Comput. **58**(6), 721–727 (2009)
11. Kim, S.W., Swartzlander, E.: Parallel multipliers for quantum-dot cellular automata. In Proceedings of IEEE Nanotechnology Materials and Devices Conference, pp. 68–72 (2009)
12. Cocorullo, G., Corsonello, P., Frustaci, F., Perri, S.: Design of efficient QCA multiplexers. Int. J. Circuit Theory Appl. **44**(3), 602–615 (2016)
13. Walus, K., Jullien, G., Dimitrov, V.: Computer arithmetic structures for quantum cellular automata. In Record of Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, pp. 1435–1439 (2003)
14. Sridharan, K., Pudi, V.: Design of a multiplier in QCA. In: Design of Arithmetic Circuits in Quantum Dot Cellular Automata Nanotechnology, pp. 73–78. Springer International Publishing, Berlin (2015)
15. Sayedsalehi, S., Azghadi, M.R., Angizi, S., Navi, K.: Restoring and non-restoring array divider designs in quantum-dot cellular automata. Inf. Sci. **311**(1), 86–101 (2015)
16. Mohammadi, M., Gorgin, S., Mohammadi, M.: Design of non-restoring divider in quantum-dot cellular automata technology. IET Circuits Dev. Syst. **11**(2), 135–141 (2017)
17. Cui, H., Cai, L., Yang, X., Feng, C., Qin, T.: Design of non-restoring binary array divider in quantum-dot cellular automata. Micro & Nano Lett. **9**(7), 464–467 (2014)
18. Sasamal, T.N., Singh, A.K., Ghanekar, U.: Design of non-restoring binary array divider in majority logic-based QCA. Electron. Lett. **52**(24), 2001–2003 (2016)
19. Walus, K., Dysart, T.J., Jullien, G., Budiman, A.R.: QCADesigner: A rapid design and simulation tool for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **3**(1), 26–31 (2004)

Elegant Energy Competent Lighting in Green Buildings Based on Energetic Power Control Using IoT Design

B. Tarakeswara Rao, B. V. V. S. Prasad and Subba Rao Peram

Abstract Green building is a critical research zone in IoT. The energy productivity in green structures is a key for worldwide supportability. However, many elements influence energy utilization by the gadget and a large portion of the green structures are not by any stretch of the imagination green because of static energy supply and concentrated control on gadgets. Here, we propose a plan utilizing LEDs to influence utilization of dynamic energy to supply to build energy proficiency and circulated bunching of gadgets for shrewd programmed control.

1 Introduction

Green building is an important research area in IoT. The energy efficiency in green buildings is vital for global sustainability. According to general survey lighting in smart building is responsible for more carbon dioxide emission than other equipment, most of the currently available green building uses location-based control over the system to attain power efficiency but the test bed fails under multiple subjects [1–3]. In some green buildings, centralized control of devices is used to implement the dynamic power policy [4, 5]. Many green buildings also use elimination of DC-to-AC and AC-to-DC conversions where lot of power is lost over conversion [6]. This work focusses on green building lighting to minimize the carbon dioxide emission and increase the power efficiency by using a design based on LEDs. The key aspects of the work are as follows:

B. T. Rao
Kallam Harnadhareddy Institute of Technology, Guntur, India
e-mail: tarak7199@gmail.com

B. V. V. S. Prasad (✉)
DRK College of Engineering and Technology, Hyderabad, India
e-mail: prasad_bvvs2004@yahoo.co.in

S. R. Peram
Vignan's University, Guntur, India
e-mail: subbarao.peram@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_25

(1) Dynamic power control: Through microcontroller's power to the devices are controlled depending on the surrounding conditions measured with the help of sensors [7].
(2) Data gathering: Through serial monitoring [8], data from microcontrollers are stored in MySQL database
(3) Future prediction: The data stored [9] in database are processed to attain the power consumption pattern for gathering information such as future device failure, quality of the product, and rate of product production.

The result shows that due to changes in system design using LEDs and dynamic power control rather than florescent [10] devices and static power control, the carbon dioxide emission is decreased and increases power efficiency effectively.

2 Dynamic Power Control

High-power LEDs are used to produce bright white light [11] instead of commercial CFL bulbs. As LEDs are sensitive to small voltage changes, a driver circuit is used to regulate voltage. Room lighting intensity is measured using light-dependent resister (LDR) to adjust the light emission diode (LED) brightness automatically. Pulse width modulation (PWM) signal is used to control the LED brightness by dynamic changing the power supply. Sensors [12] are used here to measure device temperature and LED intensity.

Here, Fig. 1 shows that the variation of power supply of the device is inversely proportional to the change in surrounding light intensity. As the environmental light intensity increases, the power supply decreases dynamically in time.

Figure 2 shows the working principle of the device as the device continuously monitor the surrounding environmental [13, 14] changes then adjusts its power supply depend on the light intensity of the surrounding dynamically. Here using LDR, the room brightness is measured continuously based on measured brightness

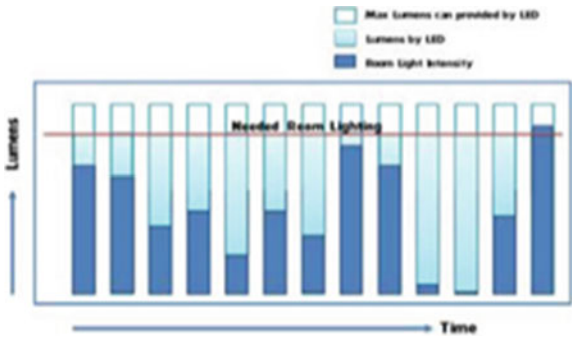


Fig. 1 Dynamic control

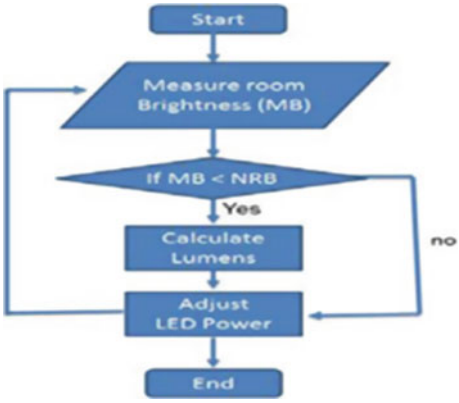


Fig. 2 Flowchart for dynamic power control

(MB) and the power is dynamically [15, 16] adjusted depending on the condition if MB is found to be less than needed room brightness (NRB) needed lumens are calculated to adjust the power supply to the device.

Figure 3 below illustrates the correlation between device power supply and lumens in variation of time (Fig. 4).

This system [17] design using dynamic power control of LEDs provides better lighting compared to static power control and it improves power efficiency by 35%. Power needed is controlled depending on the surrounding conditions such as climatic change and structural design of the building [18]. The more the building is designed to allow sunlight inside the building less power will be consumed which in turn will increase the efficiency [19, 20]

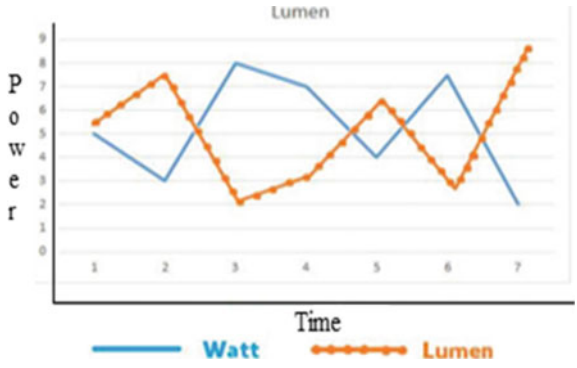


Fig. 3 Correlation among power versus lumens over time

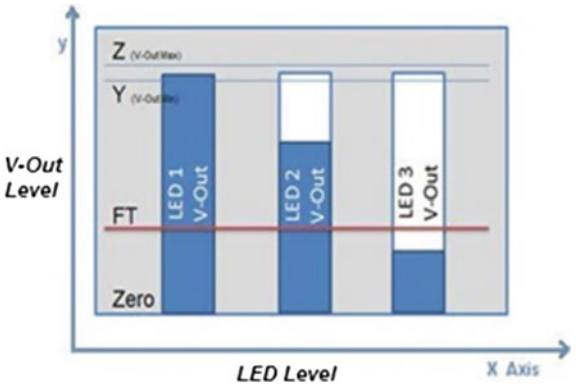


Fig. 4 Model chart

3 Data Gathering

For gathering data, a test bed is designed to provide the same environment of green building which allows sunlight pass inside the building to increase the power efficiency [21]. Here, a set of LDR and temperature sensors are connected with the microcontroller to observe the surrounding environment. Data obtained through sensors [22] are stored in database for future processing. In Table 1, it shows that the value gathered from sensors like LDR and temperature sensor as V-Out and temperature.

The data are gathered for different seasons [23] to predict the effect of temperature and humidity over the newly designed system. This will help to improve the system design depending on the surrounding conditions. It is observed that each season and also small changes in surrounding [24] environment greatly affect the system power policy, so it is not possible to provide a static or location-based power policy in this design. The collected data are stored in the cloud. Administrator [25] can track the device status regularly through the stored data. Manufacturer can make use of these data to predict LED failures before it occurs and increase product production depending on device failure rate.

Table 1 Model database

Device ID	Watt	V-Out	Temperature	Status	Lifetime (h)	Date	Place
1	3	900	26	Alive	1600	17–2–16	Delhi
2	4	1200	29	Alive	1800	18–2–16	Andhra
3	5	200	25	Dead	2100	19–2–16	Tamil Nadu
1	3	600	25	Warning	1700	17–2–16	Delhi
2	4	1200	27	Alive	1900	18–2–16	Andhra
3	5	0	20	Dead	2120	19–2–16	Tamil Nadu
1	3	150	24	Dead	1800	17–2–16	Delhi

4 Future Prediction

Here in future prediction module, the stored data are mined to attain details about the probability of failure of LEDs. This helps to measure product quality manufactured on certain factory at certain date. Objectives of this module are as follows:

1. To identify future device failure before it occurs.
2. To calculate product quality of certain batch of devices.
3. To predict increase in product production rate early
4. To identify need of change in product design depending on environment

To identify future device failure:

As LEDs cannot consume all the power supplied to it and some power will flow through the LED to the other end of the circuit until the circuit is closed, we can observe the status of the LED as alive or failure. Let the power flow to the other end be V-Out; then for given power supply X, the V-Out will be Y to Z

If V-Out is greater than zero and smaller than Y

V-Out threshold:

The V-Out threshold is used to identify the status of the LED such that it is dead or alive. This threshold will change dynamically with respect to supplied voltage as the voltage increases the threshold value also increases.

Calculation:

Let V be the V-Out obtained for X voltage threshold (T) = X * C

Here, C is the V-Out by the LED for 1 W.

If V < T & > 0, the device is assumed to be in failure risk.

If V <=1, the device is assumed to be failure.

Figure 5 shows that dynamic monitoring of the devices to predict the failure and estimate the failure status of the devices based on V-Out threshold. It shows the amount of lumen emitted by that device and power supplied to the device in watt. The V-Out of the device is used to know device status such as active, warning, and dead accurately.

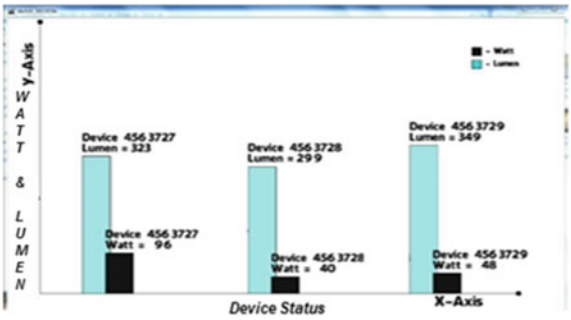


Fig. 5 Dynamic power supply monitoring

In the above Fig. 6 status of three devices are plotted such indicates active, failed and device which is going to faildepend on the V-Out Threshold. To improve product quality and service by manufacturer:

By tracking device, ID manufacturer is able to gather possible failure detail about the product made on particular date and place. Then devices can be monitored dynamically for possible failure detection and to predict lifetime of the devices. Even the data gathered will be used to change the device design based on the surrounding environment as the environmental conditions have a great impact on the device performance. To predict increase in product production rate early: Here, processing tool is used to fetch and plot the information in the sensor data stored in database. The below chart shows the percentage of devices failed percentage of devices alive percentage of devices about to fail. It is plotted for all the devices from a single manufacturer totally and also plotted location wise. This helps to predict the quality of the device and take remedial action accordingly.

Average power consumption and average lifetime of all installed devices so that mass monitoring of devices is achieved.

Then carbon emission over the devices was also monitored to minimize the CO_2 emission by decreasing the power consumption.

Algorithm for calculating room brightness
Calculation for measure room Brightness (MB) ()
{
Case-1

1. For planes opposite to the heading of candlepower (inverse square law)
2. Foot unending (fc) = I/D2

I = Candlepower in candelas (cd)
D = Direct separate between the light and the point where light level is figured
Case 2

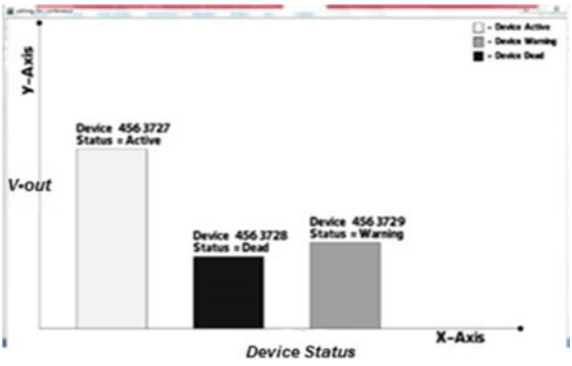


Fig. 6 Dynamic device status monitoring

Many work planes are not opposite to the bearing of light force, which why figuring light level at a point is valuable for such applications. In this cases, we frequently should decide light levels on work planes that are not flat and opposite but rather tilted or even vertical. For title even or vertical planes:

1. Horizontal foot-candles (fch) = I/D2 * H
2. Vertical foot-candles (fcv) = I/D2 * L

I = Candlepower in candelas (album)
D = Direct separation between the light and the point where light level is ascertained
H = Distance between the light and the point coordinate beneath on the work plane
L = Distance between the point and the point where light level is being ascertained
D = SquareRoot of (H2 + L2) or D2 = H2 + L2
} Calculation Dynamic power control ()
{ On the off chance that (MB < NBR)
{
Ascertain Lumens { Case-1 {

Required Light Output installation (lumens) = (Maintained Illumination in Foot candles x Area in square Feet)/(Number of Fixtures x Coefficient of Utilization x Ballast factor x Light Loss Factor)

}
Case-2
{

Light Loss Factor (LLF) = Ballast Factor x Fixture Ambient Temperature factor x supply Voltage variety Factor x light Position Factor x ideal factor x Fixture Surface Depreciation Factor x light Burnouts Factor x Lamp Lumen Depreciation Factor x Fixture Dirt Depreciation Factor x Room surface Dirt Depreciation Factor

Light Burnout Factor = 1 – Percentage of lights Allowed to Fail Without being Replaced

}
Else
{
Modify drove control
Goto measure room Brightness (MB) ()
} } } }

The proposed algorithm illustrated and is used for calculating the room brightness so that only desired amount of energy is provided to the lightening system for energy efficiency. To identify Need of change in product design depend on environment By analyzing the surrounding conditions such as temperature of the device that failed, we are able to identify the changes that need to be done for the particular

5 Result and Comparison

Here, all the devices are also dynamically monitored to predict the overall device failure and environment or area in which the device is installed to increase the device lifetime. The dynamic power control algorithm is used for calculating the only required watts of power needed at that moment and is supplied to the unit for lightening which is very effective in utilizing the power.

Figure 3 illustrates the correlation between device power supply and lumens in variation of time. Through microcontroller's power to the devices are controlled depending on the surrounding conditions measured with the help of sensors. In figure compares the variations of lumens and watts with respect to time and desired power.

Figure 4 illustrates that the V-Out threshold is used to identify the status of the LED such that it is dead or alive. This threshold will change dynamically with respect to supplied voltage as the voltage increases the threshold value also increases.

Figure 5 shows that dynamic monitoring of the devices to predict the failure and estimate the failure status of the devices based on V-Out threshold. It shows the amount of lumen emitted by that device and power supplied to the device in watt. The V-Out of the device is used to know device status such as active, warning, and dead accurately.

Figure 6 status of three devices is plotted such indicates active, failed, and device which is going to fail depending on the V-Out threshold. To improve product quality and service by manufacturer

Figure 7 shows the percentage of devices failed percentage of devices alive percentage of devices about to fail. It is plotted for all the devices from a single manufacturer totally and also plotted location wise. This helps to predict the quality of the device and take remedial action accordingly.

Figure 8 describes the then carbon emission over the devices were also monitored to minimize the CO_2 emission by decreasing the power consumption. During

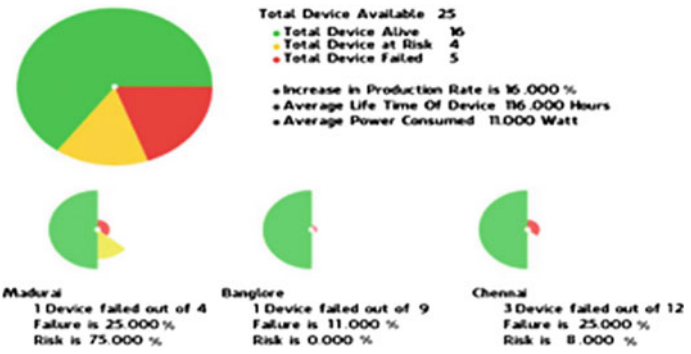


Fig. 7 Predict increase in product production

Fig. 8 Power consumption and CO_2 emission

daytime, we do not require lights so these automatically decrease the CO_2 emission than night; overall by using these smart lighting, we can also save the environment as shown in Fig. 9. Comparison with existing system with proposed method performs better with respect to my simulation results.

The proposed environment-based control system consumes only 18.33% of power compared to what the existing system (location based) consumes on daytime. But there is no change in power consumption on nighttime compared to existing system. Another inference that could be derived from the results is that, in the existing system with location-based power control policy, the performance (power consumption) decreases with increase in number of objects, whereas in our proposed system (environment based) the performance remains the same.

Table 2 shows manufacturer database which contains device Id, manufacture data, and location of manufacturing.

Fig. 9 Comparison with existing system

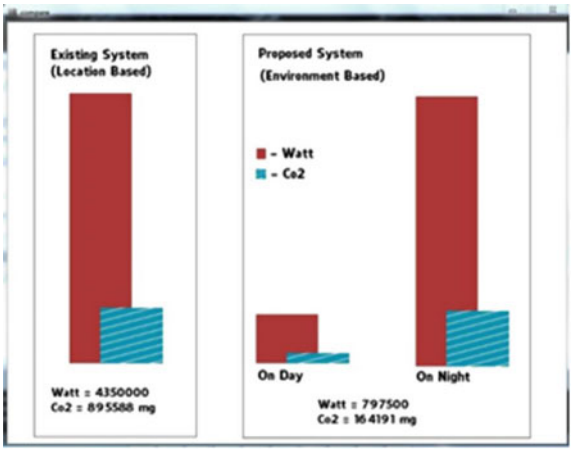


Table 2 Manufacturer database

Device ID	Manufacturing date	Place	…
1	17-2-I6	Delhi	…
2	18-2-16	Andhra	…
3	19-2-16	Tamil Nadu	…
4	17-2-16	Delhi	…
5	18-2-16	Andhra	…
6	19-2-16	Tamil Nadu	…

6 Future Work

Data will be mined to obtain patterns about effect of temperature and climatic changes in power consumption over the system. Details about effect of time over the LEDs can also be gathered. Following extensions could be done.

1. By observing effect of temperature over the system, power policy could be created for different seasoning.
2. The patterns are used to improve the system performance by altering the physical components.
3. Lifetime of the device is calculated depending on various factors such as power consumption, time, and device temperature.

7 Conclusion

In this paper, a new design is proposed and implemented to increase the power efficiency of a green building based on the environmental factors. Thus, carbon emission is minimized. In the proposed approach, an algorithm is used for calculating the room temperature for lighting system and then dynamic power control algorithm is used for only giving sufficient power with out any wastage of power for efficient handling of power.

References

1. Pan, J., Jain, R., Pan, S., Vu, T., Saifullah, A., Sha, M.: An IoT framework for smart energy in buildings: design, prototype and experiment. IoT J. **2**(6), (2015)
2. Dogo, E.M., et al.: Development of Feedback Mechanism for Microcontroller Based SMS Electronic Strolling Message Display Board (2014)
3. Jagan Mohan Reddy, N., Venkareshwarlu, G., et al.: Wireless electronic display board using GSM technology. Int. J. Electr. Electron. Data Commun. ISSN: **1**(10), 2320–2084 (2013)
4. Victoria Moreno, M., Skarmeta, A.F., Jara A.J.: Exploiting IoT-based sensed data in smart building to model its energy consumption. In: IEEE ICC 2015 SAC Internet of Things

5. Understanding the Internet of Things (IoT) (2014)
6. Stankovic, J.: Research directions for the internet of things. Int. Things J. IEEE **1**(1), 3–9 (2014)
7. Gubbi, J., et al.: Internet of Things (IoT): a vision, architectural elements, and future directions. Futur. Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
8. Yashiro, T., et al.: An internet of things (IoT) architecture for embedded appliances. In: Humanitarian Technology Conference (R10 - HTC), 2013 IEEE Region, vol. 10. IEEE, New York (2013)
9. Vermesan, O., Friess, P. (eds.): Internet of Things—From Research and Innovation to Market Deployment. River Publishers (2014)
10. Jain, R.: The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling. Wiley, Hoboken (1991)
11. Internet of Things. Wikipedia (2016)
12. Graham, M., Haarstad, H.: (2011) Transparency and development: ethical consumption through Web 2.0 and the internet of things. Res. Artic. **7** (2011)
13. Jayavardhana, G., Rajkumar, B., Marusic, S. and Palaniswami, M.: Internet of things: a vision, architectural elements, and future directions. Future Gener. (2013)
14. Grieco, A., Occhipinti, E., Colombini, D.: Work postures and Musculo—Skeletal disorder in VDT operators Bollettino de Oculistica. Supplement **7**, 99–111 (1989)
15. Gigli, M., Koo, S.: Internet of things, services and applications categorization. Adv. Internet Things **1**, 27–31 (2011). https://doi.org/10.4236/ait.2011.12004
16. Pahlavan, K., Krishnamurthy, P., Hatami, A., Ylianttila, M., Makela, J.P., Pichna, R., Vallstron, J.: Handoff in hybrid mobile data networks. Mob. Wirel. Commun. Summit **7**, 43–47 (2007)
17. ITU Internet Reports, International Telecommunication Union. The Internet of Things: 7th edn. (2005) www.itu.int/internetofthings/on
18. Want, R.: An introduction to RFID technology. IEEE Pervasive Comput. **5**, 25–33 (2006)
19. Li, B.A., Yu, J.J.: Research and application on the smart home based on component technologies and internet of things. Proc. Eng. **15**, 2087–2092 (2011). https://doi.org/10.1016/j.proeng.2011.08.390
20. Chen, X.-Y., Jin, Z.-G.: Research on key technology and applications for the internet of things. Phys. Proc. **33**, 561–566 (2012).http://dx.doi.org/10.1016/j.phpro.2012.05.104
21. Razzak, F.: Spamming the internet of things: a possibility and its probable solution. Proc. Comput. Sci. **10**, 658–665 (2012). https://doi.org/10.1016/j.procs.2012.06.084
22. Shao, W., Li, L.: Analysis of the development route of IoT in China. Perking: China Sci. Technol. Inf. **24**, 330–331 (2009)
23. Sun, C.: Application of RFID Technology for Logistics on Internet of Things (2012)
24. Moeinfar, D., Shamsi, H., Nafar, F.: Design and Implementation of a Low—Power Active RFID for Container Tracking @ 2.4 GHz Frequency: Scientific Research (2012)
25. Bicknell, IPv6 Internet Broken, Verizon Route Prefix Length Policy (2009)

Energy-Efficient Routing Protocols for WSNs: A Comparative Analysis

B. A. Mohan and H. Sarojadevi

Abstract Cluster-based routing protocols are used to reduce the energy consumption of sensor nodes, which involves them in short-distance communication to cluster head than to sink which is located far away. Cluster head collects and aggregates data to decrease number of transmissions to sink. As cluster head performs multitask, energy is extensively used, which in turn leads to premature death. To avoid this, cluster head role is shifted among the member nodes in the cluster. A slight modification is to introduce mobile nodes to relay information from cluster heads to sink to reduce the energy consumption of the cluster heads. In this paper, comparative analysis of reactive protocols (LEACH and LEACH-C), single mobile node protocol (NDCM), and multiple mobile node-based hybrid protocol (HADMMN) is carried out. Use of multiple mobile nodes significantly improves the performance of the network.

1 Introduction

Wireless sensor network (WSN) is an interconnection of small sensor nodes that are capable of communicating with other nodes through radio waves and are deployed randomly to sense, do surveillance, and understand the physical environment. Network of sensor nodes and systems is seen by observers as an important technology that will concur major development in the future for a number of applications. Apart from sensing, this often also widens to a subject further which will develop interest in control and activation.

An approach to conserve node energy is to use efficient routing protocol. Routing is a process of discovering the path to the destination through intermediate

B. A. Mohan (✉) · H. Sarojadevi
Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology, Bengaluru 560064, India
e-mail: mohan.ba@nmit.ac.in

H. Sarojadevi
e-mail: sarojadevi.n@nmit.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_26

nodes. Transmitting data using intermediate node saves the energy of nodes placed far apart. Data communication consumes energy up to 80% of sensor node's energy. Developing efficient routing protocols to extend the network lifetime is therefore challenging.

Wireless sensor networks (WSNs) present special requirements while designing a new routing protocol that should be efficient. The essential constraint for any efficient routing protocol is to conserve energy of the nodes in the network.

2 Literature Review

2.1 Hierarchical Routing Protocols

Benefit of clustering protocols is mainly reducing the energy consumption of nodes, which involves them in short-distance communication to cluster head than to sink which is located far away. Cluster head collects and aggregates data to decrease the number of transmissions to the sink making cluster-based protocol energy efficient.

Sensor nodes deployed in the network are divided into clusters, and each cluster has information about its member nodes and a path to reach other cluster or sink node. Clusters are capable of doing different complex tasks by distributing workloads across all connected nodes. Clustering improves the system availability to the users by aggregate resources for accomplishing task with tolerance to faults and failures. The first routing protocol based on clustering approach is introduced by Wendi B. Heinzelman called "Low Energy Adaptive Clustering Hierarchy" (LEACH) in 2002 [1, 2, 3].

2.2 Cluster-Based Hybrid Routing Protocols

Hybrid protocol combines features from two or more protocols. Many protocols have been proposed and implemented. Some example protocols are "Adaptive Periodic Threshold-Sensitive Energy-Efficient Sensor Network Protocol" (APTEEN), "Hybrid Energy-Efficient Distributed Clustering" (HEED), "Partition-Based Hybrid Clustering Routing Protocol for WSNs" (PHCR), etc.

APTEEN [4]—This is the expansion over "Threshold-Sensitive Energy-Efficient Sensor Network Protocol" (TEEN) [5]. Its main focus is to capture data at regular interval as in LEACH and time critical event-based data as in TEEN. Results show that proposed hybrid protocol's performance lies in between "LEACH" and "TEEN".

HEED [6]—This protocol extends LEACH by adding node degree as an additional parameter for cluster head selection. The primary goal of this protocol is to extend network lifetime by ending the clustering process in a fixed number of iterations, minimizing control overhead and producing well-distributed cluster heads.

PHCR [7]—This protocol addresses the problem of unequal distribution of clusters. Nodes in the network are grouped into multiple sectors using partitioning algorithm. Initially, sink node selects the cluster heads which is in the center of densely populated area and other nodes in each sector joins the cluster head to form a sector. The second closest sensor node to the center of sector is elected as cluster heads for the second round; from third round onward, the cluster heads are chosen by previous round's cluster heads. The simulation shows that the proposed protocol improves network lifetime greatly.

Rolling Grey Model (GM) for Energy Optimization in WSNs—The approach discussed in this paper involves rolling GM (1, 1) as prediction model to optimize routing scheme in terms of energy minimization. Energy-efficient cluster-based routing protocol with residual energy and distance-based cluster head selection method are used to reduce overheads during cluster formation. Obtained results are better in comparison with GM (1, 1) model [19].

Cluster-Based Time Series Prediction for Data Reduction in WSNs—The paper discusses energy-efficient cluster-based routing protocol with residual energy and distance-based cluster head selection method. The cluster formation in the proposed approaches is different from other routing protocols as its clusters will be fixed throughout the network lifetime. This reduces the overhead in cluster formation for every round. Further, inter-cluster and intra-cluster transmissions are also minimized which results in an overall reduction in energy consumption of WSNs [20].

2.3 Hybrid Routing Protocol Based on Mobile Nodes

Efficient energy utilization is the challenging issue in sensor nodes. The amount of energy spent to transmit signal between nodes depends on the distance between them. Therefore, the use of single hop communication mode to reach sink is costly in WSN; on the other hand, multi-hop communication mode causes the nodes nearer to sink deplete energy considerably faster than other nodes in the network, which partitions the network into parts. To address this problem, mobile data collection strategies are introduced in sensor networks. The mobile element takes the responsibility of relaying data from sensor nodes to sink by using long-distance direct communication channel to sink or must move to some point in network to establish a connection to sink. Furthermore, mobile data collector helps to extract data from disconnected network parts also.

Data collection is a central problem in WSN. Current investigation proves that the use of mobile data collectors is advantageous as compared to static multi-hop routing [8, 9, 10, 11].

Following are some of the protocols based on mobile nodes.

Multiple Mobile Node-Based Hybrid Protocols [12]—"A Hybrid Approach for Data collection using Multiple Mobile Nodes" (HADMMN) [13] To conserve the energy of WSN using hierarchical/clustering routing protocol, the job of cluster head is taken shift between the member nodes in every round to uniformly

distribute the load and to avoid isolation of the network. Mobile nodes are introduced to further reduce the energy consumption in the cluster heads by relaying data from cluster heads to sink. According to this protocol, base station divides the network into sectors in the first round; each sector consists of a few clusters with one mobile node, and from second round onward, decentralized cluster head election algorithm is used. Mobile nodes are intended to relay data from cluster heads to sink, conserving the energy of cluster heads.

Mobile Node Data Collection Protocols [14, 15, 16]—"Node Density-based Clustering and Mobile Collection" (NDCM) associates the feature of cluster-based routing and mobile element-based data collection. First, cluster head gathers data from its cluster members and then the mobile sink visits these cluster heads for collecting data. The node that has maximum connectivity or node degree is most likely to become cluster head, which improves the efficiency of the network.

Optimal Mobile Sink Selection Scheme—Current studies have shown the benefits of using mobile nodes in gathering data from sensor nodes. However, node movement and direction cause connectivity disruptions, leading to delay in data delivery. To address this issue, a new approach is introduced, based on the connection expiry time [17]. Nodes nearer to mobile sink data collector will join the cluster, assuming that closeness to the sink offers the longest connectivity time and low communication cost. This protocol reduces energy consumption, improving packet delivery ratio.

New Partitioning-Based Data Gathering Algorithm for Path Constrained Cases—The idea of involving multiple mobile sinks for data collection with assistance of area splitting algorithm (ASA) [18] has shown significant improvement in the performance of network in terms of power consumption and network latency. However, network with time and energy constraints demands more number of mobile sinks and a method for load balancing. A novel technique called "network partitioning algorithm" (NPA) for data collection was introduced. This protocol groups the nodes into two partitions based on centroid of nodes locations. This process is repeated until the prohibited number of mobile elements is reached or path constraint for mobile elements is satisfied. The results show that the proposed algorithm performs better than ASA with respect to minimizing the mobile elements needed and data collection latency.

3 Performance Evaluation

To perform comparative analysis of energy-efficient routing protocols, LEACH and LEACH-C are chosen from proactive protocols category, NDCM is a single mobile sink-based protocol, and HADMMN is a multiple mobile node-based protocol. To compare these protocols, parameters like control overhead, throughput, energy consumption, and network lifetime are considered.

The performance of HADMMN for 50, 75, and 100 nodes is shown in Table 1. Results show that there is improvement in the throughput when number of nodes is

Table 1 Result analysis of HADMMN for 50, 75, and 100 nodes

	For 50 nodes	For 75 nodes	For 100 nodes
Throughput	2.4e-05 Mb/s	4.8e-05 Mb/s	4.8e-05 Mb/s
Control overhead	354	510	642
Average energy	0.371 J	0.608 J	0.877 J
Network lifetime	1415048 ms	1294836 ms	1196636 ms

increased to 75 nodes and remains constant till 100 nodes, which establishes that sustained throughput is ensured with large number of nodes. The control overhead and average energy consumption increase as nodes in network are scaled up from 50 to 75 and from 75 to 100 nodes, respectively. Overall network lifetime decreases as nodes in the network are increased from 50 to 75 and from 75 to 100 nodes.

The results are as expected, since there are more nodes actively participated in the network, consuming resources and energy.

HADMMN protocol is compared with LEACH which uses distributed clustering algorithm and LEACH-C which uses centralized clustering algorithm. The results for these protocols are shown in Fig. 1. The simulation is done for 100 nodes, and HADMMN outperforms LEACH and LEACH-C as shown in Fig. 1. The lifetime of the network for HADMMN scheme is greater than LEACH and LEACH-C.

Table 2 shows the performance results of HADMMN vs NDCM which is a single mobile sink-based protocol. HADMMN's throughput is a little more for 50 and 60 nodes; when nodes are increased to 70, its throughput drastically increases. HADMMN's network lifetime is more compared to NDCM and starts decreasing suddenly when nodes are increased beyond 60.

In sum, hierarchical network configuration with mobile nodes improves the performance and is highly energy efficient. The hybrid protocol proposed in research work is an approach for large-size WSN deployment. This is due to the fact that HADMMN has very low energy consumption and low overhead, while guaranteeing high and sustained throughput.

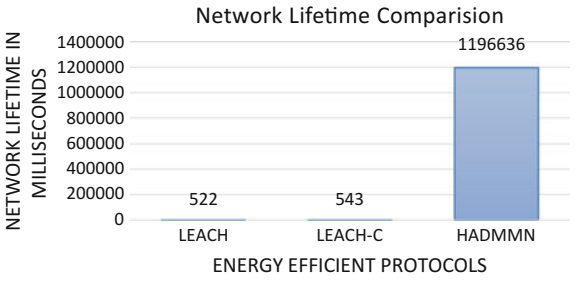


Fig. 1 Network lifetime of LEACH, LEACH-C, and HADMMN

Table 2 Result analysis of HADMMN and NDCM for 50, 60, and 70 nodes

	Protocols	50 nodes	60 nodes	70 nodes
Throughput in Mb/s	HADMMN	2.4e-05	2.4e-05	4.8e-05
	NDCM	1.368e-05	1.368e-05	1.368e-05
Control overhead in no. of packets	HADMMN	354	432	498
	NDCM	40886	51958	58625
Avg. energy in joules	HADMMN	0.3710	0.4549	0.5602
	NDCM	248.92	366.31	459.09
Network lifetime in ms	HADMMN	1415048.12	1384783.25	1311902.80
	NDCM	2008.65	1637.95	1524.72

4 Conclusions and Future Work

The results show employing multiple mobile nodes for data collection in network performs better than single mobile node and single mobile data collector is better compared to static data collection as in LEACH and LEACH-C. This paper concludes that deploying multiple mobile nodes in a network improves the performance. Many challenges posed by mobile node-based WSN are determining optimal number of mobile nodes, link brakeage, speed at which mobile nodes must move, and trajectory for mobile nodes. Future work involves addressing these challenges.

References

1. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wirel. Commun. **1**(4), 660–670 (2002)
2. Singh, J., Mishra, A.K.: Clustering algorithms for wireless sensor networks: a review. In: 2nd International Conference on Computing for Sustainable Global Development, vol. 2, pp. 637–642 (2015)
3. Gnanambigai, J., Rengarajan, N., Navaladi, N.: A clustering based hybrid routing protocol for enhancing network lifetime of wireless sensor network. In: 2nd International Conference on Computing for Sustainable Global Development, pp. 1–4 (2014)
4. Manjeshwar, A., Agrawal, D. P.: APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In: International Parallel and Distributed Processing Symposium, (2002)
5. Manjeshwar, A., Agrawal, D. P.: TEEN: A routing protocol for enhanced efficiency in wireless sensor networks. In: 15th International Parallel and Distributed Processing Symposium IPDPS, pp. 2009–2015 (2001)
6. Younis, O., Fahmy, S.: HEED: A hybrid energy-efficient distributed clustering approach for Ad Hoc sensor. Networks. **3**(4), 366–379 (2004)
7. Xinhua, W., Jianbing, C.: A Partition-based hybrid clustering routing protocol for WSN, In: International Conference on Internet Technology and Applications, pp. 1–4 (2011)

8. Wang, C., Ma, H.: Data collection with multiple controlled mobile nodes in wireless sensor networks. In: IEEE 17th International Conference on Parallel and Distributed Systems, pp. 489–496 (2011)
9. Pan, L, He., Xu, J.: Analysis on data collection with multiple mobile elements in wireless sensor networks. IEEE Globecom 2011 Proceedings, (2011)
10. Fotue, D., Labiod, H., Engel, T.: Controlled data collection of mini-sinks for maximizing packet delivery ratio and throughput using multiple paths in wireless sensor networks. In: IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications, pp. 758–764 (2012)
11. Yin, F., Li, Z., Wang, H.: Energy-efficient data collection in multiple mobile gateways WSN-MCN convergence system. In: IEEE 10th Consumer Communications and Networking conference CCNC 2013, pp. 271–276 (2013)
12. Francesco, M.Di., Das, Crewman, S.K.: Data collection in wireless sensor networks with mobile elements: a survey. J. ACM Trans. Sens. Netw. (TOSN). (1), vol. 8, (2011)
13. Mohan, BA., Sarojadevi, H.: A hybrid approach for data collection using multiple mobile nodes in WSN (HADMMN). In: IEEE International Conference on Emerging Trends in Engineering and Technology, pp. 736–739 (2016)
14. Ruonan, Zhang., Jianping, Pan., Jiajia Liu., Di, Xie.: A hybrid approach using mobile element and hierarchical clustering for data collection in WSNs. In: 2015 IEEE Wireless Communications and Networking Conference (WCNC). pp. 1566–1571 (2015)
15. Wang, J., Yin, Y., Kim, J., Lee, S., Lai, C.: An mobile-sink based energy-efficient clustering algorithm for wireless sensor networks. In: IEEE 12th International Conference on Computer and Information Technology, pp. 678–683 (2012)
16. Ruonan, Zhang., Jianping, Pan., Jiajia, Liu., Di, Xie.: A hybrid approach using mobile element and hierarchical clustering for data collection in WSNs. In: IEEE Wireless Communications and Networking Conference, pp. 1566–1571 (2015)
17. Abuarqoub, A.: Optimal mobile sink selection scheme for multiple sink mobile wireless sensor networks. In: European Intelligence and Security Informatics Conference. pp. 169–172 (2015)
18. Alhasanat, A., Al-Khassaweneh, M., Twaisi, A.: A new partitioning-based data gathering algorithm for path-constrained mobile elements in wireless sensor networks. In: IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies AEECT, pp. 1–5 (2013)
19. Singh, Dhirendra Pratap, et al.: Energy optimization in WSNs employing rolling grey model. In: International Conference on Signal Processing and Integrated Networks (SPIN), (2014)
20. Singh, D.P. et al.: An efficient cluster-based routing protocol for WSNs using time series prediction-based data reduction scheme, Int. J. Measur. Technol. Instrum.Eng.(IJMTIE), pp. 18-34 (2013)

Analysis of Cancer Data Set with Statistical and Unsupervised Machine Learning Methods

T. Panduranga Vital, K. Dileep Kumar, H. V. Bhagya Sri and M. Murali Krishna

Abstract Research on cancer is very important where cancer is one of the leading diseases which causes more deaths worldwide. Data mining is very useful for analysis of medical data, especially in cancer. Statistical analysis results are very useful for assumptions, preventions and diagnosis of cancer. The main aim of this paper is analysing the cancer data set collected form zone 1 (Srikakulam, Vizianagaram, Visakhapatnam) of Andhra Pradesh for estimations or diagnosis and preventions of cancer disease. The analysis is very useful with its good results. In this, we use statistical and clustering methods like k-means and hierarchical and multidimensional scaling (MDS). As per the statistical reports, most of the women are affected by cancer than men in the zone 1 of AP. Most of the women cancer patients suffered from breast cancer, and most of the men cancer patients are affected by lung cancer. The analysis also gives interesting results about living styles and habits linked with cancer. Unsupervised machine learning algorithms also give the good results for predicting cancer. The hierarchical cluster study projections clearly describe the cause of occurring cancer in zone 1 districts of Andhra Pradesh that the main combination factors are smoke, drink, gutkha (chewing tobacco) and related job.

T. Panduranga Vital (✉) · K. Dileep Kumar · H. V. Bhagya Sri · M. Murali Krishna
Department of Computer Science & Engineering,
Sri Sivani College of Engineering, Srikakulam, Andhra Pradesh, India
e-mail: vital2927@gmail.com

K. Dileep Kumar
e-mail: kadamati.dileep@gmail.com

H. V. Bhagya Sri
e-mail: bhagya.hanumanthu@gmail.com

M. Murali Krishna
e-mail: maadugula@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_27

1 Introduction

Data mining (DM) is an effective and efficient method for knowledge extraction from raw data. Raw data faces different difficulties that make traditional or conventional strategy improper for knowledge extraction [1, 2]. DM should have the capacity to handle different data types in all configuration formats. Medical or healthcare data mining is multifaceted field with commitment of medicine and data mining [3, 4]. The important medical characteristics that specified are: estimate the health costs, diagnosis and visualization; extract the hidden values from biomedicine information and find a relationship amongst diseases and drugs.

The development of information storage technology has prompted to produce a large amount of raw data that considers perspectives [5, 6]. These perspectives are mounting of modern storage equipment and algorithm advancement. Useful and effective knowledge can be gained by the normal raw data. Knowledge discovery is the insignificant extraction of verifiable, already obscure and helpful prospective information from the data [7]. Attractive components for extracted knowledge are sensible time multifaceted nature, understandability, precision and valuable result. The extracted knowledge gives the new information for further utilization discoveries. DM was initially considered as equivalent word of KDD [8, 9].

De Falco extracted knowledge by differential evolution in the form of if … then rules to predict the results of diseases. In this approach, results of DM were compared with a skilled oncologist's research work [10]. Survivability of breast cancer disease can be predicted by k-means. Results demonstrate that evaluation, stage of cancer, number of primaries and radiation are the most prognostic factors in [11].

Lung cancer is one of the frequently occurred and vulnerable cancers mostly affected by air pollution, cigarette smoking and cardiopulmonary syndrome identified by statistically significant and rich association techniques [12, 13]. Due to industrialization and urban development have increased human revelation to plentiful cyan genetic substances that was raised about their relationship to the aetiology of chronic diseases [14]. The medical scientists have long-utilized maps to track the spread of disease with using powerful novel tools [15] in geographic information system technology that assists to reveal far more than simply the 'where' and 'when' of an epidemic.

2 Methodologies

The data set has been analysed by DM techniques. The frequency of patients has been analysed for the collected data from zone 1 (Srikakulam, Vizainagaram and Visakhapatnam) districts of Andhra Pradesh, India. The questionnaire has been framed based on the present personal profile, living and food habits, travelling mode, previous history, internal and external factors, etc.

The data was collected with 1100 instances (550 cancer instances and 550 non-cancer) and 46 attributes and one class (place, age, cancer type, family history, drinking, smoking, tea, coffee, milk, job, morning-eat, lunch-eat, dinner-eat, travel, living status, fruits, vegetables, sleeping hours, tension, cooldrinks and ice cream, study, height, weight, BP, pains, hair loss, gutkha (chewing tobacco), marital status, blood group, treatment type, bathing, oils, fast food, other disease, morning walk, use mobile, drugs or using tablets, treatment mode, diagnosis mode, meditation, mosquito repellents, injuries, speak level, watching TV, think levels and class status of cancer (0 (no) for non-cancer and 1 (yes) for cancer).

K-Means (KM) Algorithm

The KM algorithm is dividing the set of observation points into k clusters [16]. Let R is the real number set and Rd said to be the d-dimensional object space. Given a finite set X that is determined by Eq. (1).

$$X = \{x_1, x_2 \ldots x_n\} \tag{1}$$

where 'n' indicates the object's number.

The KM algorithm partitions the set into subset S, whose subsets are mentioned in Eq. (2)

$$S = \{S_1, S_2, \ldots\ldots\ldots, S_k\} \tag{2}$$

where k is a predefined number.

Each cluster C is represented by an object. It is mentioned in Eq. (3)

$$C = \{c_1, \ldots., c_k\} \tag{3}$$

where C is the centre set in the object area.

The Euclidean distance estimation measures using the distance amongst items and cluster centres. Equation (4) represents the objective function that it should be minimized.

$$f = \sum_{i=1}^{k} \sum_{i=1}^{n} \left(||x_i(j) - C_j||\right)^2 \tag{4}$$

where

$x_i(j)$	is a particular ith data point at jth cluster
C_j	is centroid or cluster centre
n	is the number of instances or data points
$\|\|x_i(j) - C_j\|\|$	is the Euclidean distance(ED) between $x_i(j)$ and C_j

The cluster centres are derived by Eq. (5)

$$C_j = \left(\frac{1}{n}\right) \sum_{x \in s}^{n} X_j \tag{5}$$

where C_j is the count of data objects.

Each cluster is characterized by its centre point called the centroid. Generally, the distance measures used in clusters do not represent the spatial distance values. In general, the solution is to find the global minimum which is the exhaustive choice of starting points. But the usage of many replicates with random initial point leads to a solution called global solution. A centroid is a point whose coordinates are obtained by calculating the mean of each coordinate of the points of samples classified into the clusters. The KM algorithm input is the number of clusters k and n objects and the output is set of k clusters which minimize the squared error function.

3 Results and Discussion

The statistical analysis provided the relationship of the collected data sets. As per analysis, there are more number of cancer instances in Visakhapatnam district compared to other two districts like Vizianagaram and Srikakulam. There are more number of female patients (64.3%) compared to male patients (35.7%).

As per the analysis of family history (hereditary), 19.4% of cancer patient's family members had cancer. Hence, cancer is related to both hereditary and metabolic.

Drinking alcohol and smoking habits cause of cancer. In this research, 26.2% of cancer patients have the drinking alcohol habit and 33.3% of cancer patients have smoking habit. Some other habits are like drinking tea and coffee, 83.5% of cancer patients drink tea and 23.4% of patients prefer coffee. Hence, drinking tea may be the reason for cancer occurrence.

The study results show more number of breast cancer (20.2%) patients followed by cervix (16.3%), stomach (7.3%) and blood (7.3%). Hence, there is a need to control these cancers in the zone 1 regions of AP. 44.2% of patients are house wives which are observed to be more when compared to working women. 56.5% of cancer patients live in urban region, whereas 43.5% of patients live in rural region. According to this, there are more number of cancer patients in urban region compared to rural region. Most of them are travelling by bus and auto, and most of the cancer patients like banana, non-vegetarian, rose and jasmine

Age, height and weight are some of the factors occurring cancer. In this analysis, most of the patients are in between the age of 35–60 years. Hence, cancer can be an ageing disease. In the height factor, the mean of the height is 153.97 ± 15.09 centimetres. In the weight factor, the mean of the weight is 54.89 ± 9.97 kilograms.

Most of the cancer patients have tension (54.8%). Most of the patients do not have education. Most of them are having low blood pressure compared to high BP.

Some of the cancer patients prefer cooldrinks rarely (61.3%). Some of the patients prefer to take ghutka (chewing tobacco). About 46.6% of cancer patients have pain in body. The study also shows that most of the patients have hair loss (73.9%).

Some interesting results that 91.7% of patients are married. Most of the patients are related to O+ blood group. There is more number of diabetic (21.2%) patients associated with cancer.

The most of the patients preferred allopathic (89.5) as primary treatment. Previously, most of the patients did not use drugs. The patients underwent treatment by chemotherapy. Most of the diagnosis was conducted by scanning and biopsy. Some other facts from the study that above 26% patients used to take palm oil, 58.1% do not take fast foods and 54.2% use steel as cooking vessel.

It shows that most of the patients do not go for morning walk (63.5%), does not have any habit of meditation (82.9%) and even playing games (64.9%). Above 84.5% of cancer patients are using mosquito repellents. About 71.4% of patients did not have major injuries in their lifetime. 66.7% of patients have medium speaking level, 53.6% of patients see TV medium and 54.8% of patients think medium.

The main cluster centroid uses k-means clustering algorithm by class and gender. The model is constructed with full training data and takes the 0.5 s. The two clustered instances are cluster 0 and cluster 1. The cluster 0 contains 42% instances, whereas cluster 1 contains 58% instances. The centroid values of full data attributes are 47.4802 (age), urban (living), yes (tension), 153.9742 (height), 54.8869 (weight), etc. The cluster 0 centroid's values are 55.3302 (age), rural (living), no (tension), 155.5991 (height), 55.5425 (weight), low (BP), yes (pains), more (hair loss) and so on. The cluster 1 centroid's values are 41.7808 (age), urban (living), yes (tension), 152.7945 (height), 54.411 (weight), normal (BP), no (pains), less (hair loss) and so on.

Figure 1 shows the k-means cluster analysis related to age attribute and cancer type attribute. In this, the blue colour elements represent the cluster 0 and the red colour elements represent the cluster 1. The square symbols represent the male elements, and the cross symbols represent the female elements. The clusters 0 and 1 are formed with patients' age between 35 and 70 years and patients of breast and lung cancers.

Figure 2 shows the scatter plot for cancer types with blood groups at regional level. Most of the cancer instances from Visakhapatnam and Vizianagaram are seen in A +ve, B +ve and O +ve blood groups. K-means cluster algorithm is used to analyse the scatter plot.

Figure 3 shows the multidimensional scaling (MDS) plot. As per the analysis, it has been observed that more members from Visakhapatnam are affected by cancer due to tensions.

Figure 4 shows the complete link hierarchical clustering for the cancer data set of zone 1 districts of Andhra Pradesh. In this, first compute the pairwise distances of two clusters named as C1, C2 and then choose the largest distance of all pairwise distances. In other words, if the cluster C1 contains m elements and the C2 cluster contains n elements, then compute the m*n pairwise distances and construct the hierarchical link for chosen largest distance for that element.

In this experiment, the complete link hierarchical clustering is grouping the ages of cancer instances with respect to the cancer type attribute. The distance between

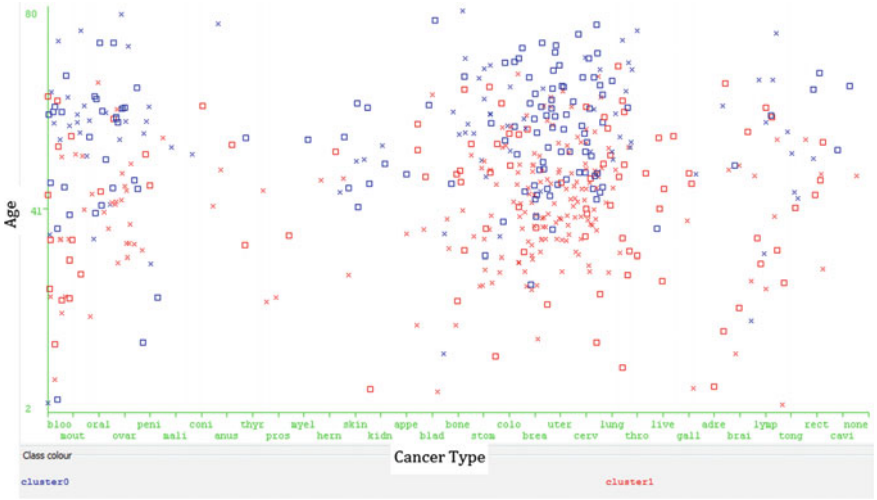


Fig. 1 K-means clustering related to cancer type and age

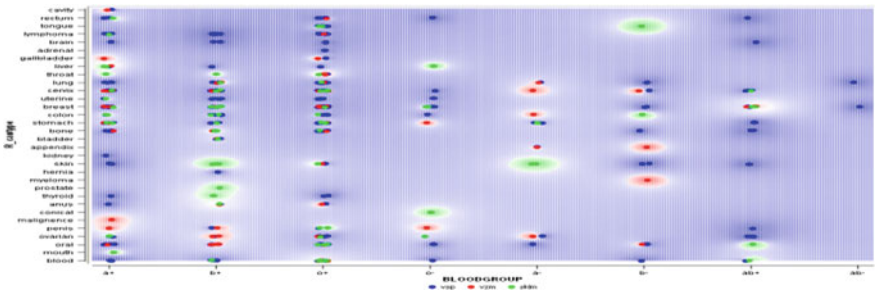


Fig. 2 Scatter plot for relationship to blood group, cancer type and place using k-means clustering

two clusters is determined as the maximum of pairwise distances. The maximum distance of two cluster elements (age 9 and age 9) is connected in hierarchical way at the distance 3.886854 (shown in green colour hierarchical cluster block). In the same way, all elements in green colour cluster elements are constructed in hierarchical manner. The maximum height of the green colour hierarchical cluster block is 5.646441. As well as, the maximum height of the pink colour cluster structure is 6.10689, and the blue colour cluster height is 6.4366. The blue and pink colour clusters are connected at the maximum height 6.591434 and treated as one cluster. It is connected with the green colour cluster at height 6.766052 in-depth value 5.

In the same way, complete link hierarchical clustering can be applied to the cancer type attribute (Fig. 4). The maximum distance of two cluster elements (blood and bone) is connected in hierarchical way at the distance 3.886854 in green

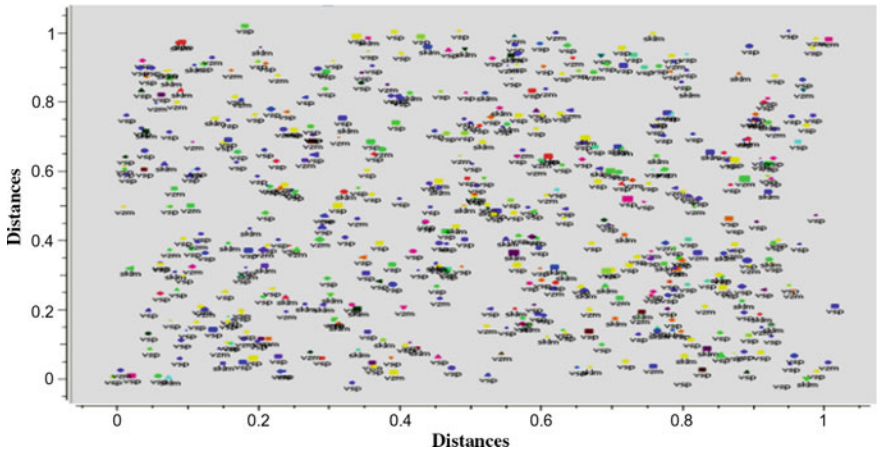


Fig. 3 MDS Plot

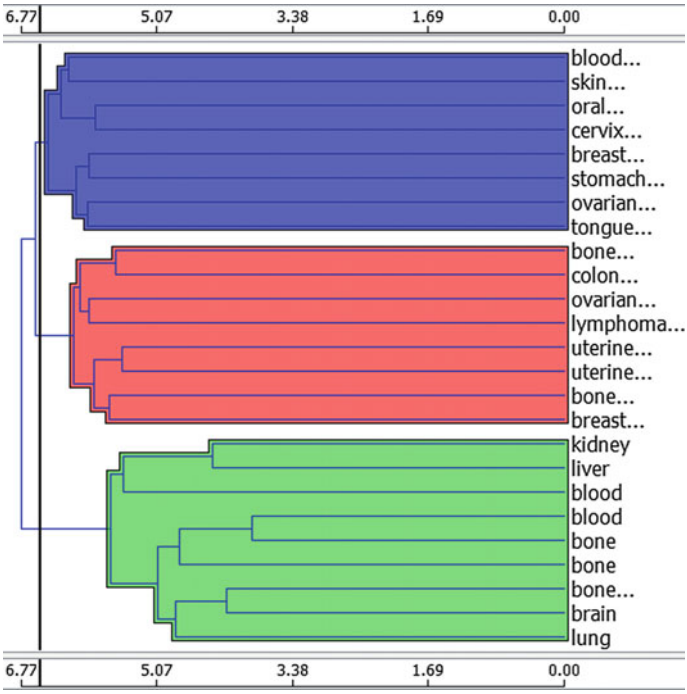


Fig. 4 Complete link hierarchical cluster with respect to cancer type attribute of cancer data set

colour cluster block. In the same way, all elements in green coloured cluster are constructed in hierarchical manner. The height of the green colour hierarchical cluster is 5.646441. As well as, the height of the pink colour cluster is 6.10689, and the blue colour cluster height is 6.4366. The blue and pink colour clusters are connected at the maximum height 6.591434 and treated as one cluster. It is connected by the green colour cluster at height 6.766052 in-depth 5.

Figure 5 shows the linear projection by using hierarchical clustering for cancer data set. The projections are measured with five parameters (projections with maximum five attributes) that are cancer type (cantype), smoke, drink, gutkha (chewing tobacco) and job. The most projections are interrelated with these attributes that this combination is the first rank with 92.98% predictable score of all

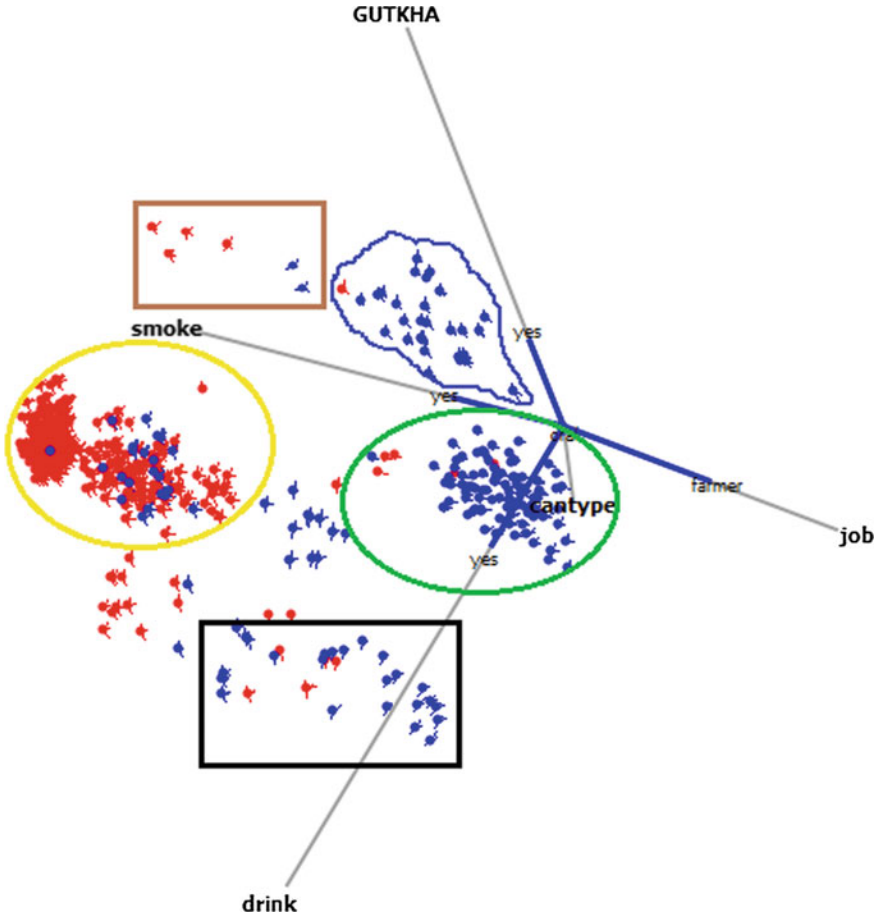


Fig. 5 Projections analysis of cancer data set

other tested possible placements of five visualized attributes in the projection list. For this, k-nearest neighbour method is used for projection evaluation and supervised principal component analysis for optimizing separation analysis. The study clearly describes the cause of occurring cancer in zone 1 districts of Andhra Pradesh that the main combination factors are smoke, drink, gutkha (chewing tobacco) and related job.

The figure shows the projections analysis of cancer data set. The red colour plotting points represent the female instances, and blue colour plotting points represent the male instances of cancer data set. The green colour circle block contains the most of male instances and complete instances in the block related to habit of smoking (yes), drinking (yes) and gutkha (yes). In this, many instances are associated with mouth, lung, colon and stomach cancers as per the parameter cancer type (cancer type attribute) computations, and it's data points indicates the lower level job status like farmer, auto driver related to job parameter (attribute).

The complete instances of yellow colour circle block correspond to the no habits like smoke (no), drink (no) and gutkha (no). Most of the data points of this yellow circle block are red colour that indicates the female instances of cancer data set. Most of the exemplars specify the breast and colon cancer on the cancer type (cantype) parameter and also describes the job parameter that corresponds to housewife, worker, and office assistant.

The black-bordered block data points show that drink parameter value is yes and remaining smoke and gutkha attributes values are no. The blue-bordered block data points show that drink parameter value is no and remaining smoke and gutkha parameters values are yes. The brown-bordered block data points show that drinking habit is yes and remaining smoke and gutkha attribute values are no.

4 Conclusion

The analysis with attributes age, sex, smoke, mosquito repellents, gutkha (chewing tobacco) thinking levels, tensions and food habits are major factors in cancer occurrences. In this, scatter and MDS plates are analysed using k-mean cluster algorithm. It gives the good result causing cancer. As well as, projections analysis is used by hierarchical clustering with Kruskal's algorithm and display attribute projections rankings of cancer data set. Further analysis will be conduct on prediction and curing of cancer.

Note: Authors have taken the consent from the concerned person/authority to use the materials, etc., in the paper. Authors will be solely responsible if any issues arise in future with regard to this.

References

1. Cruz, Joseph A., David S Wishart.: Applications of machine learning in cancer prediction and prognosis. Cancer inf. **2** (2006)
2. Hara., Ichimura, T.: Data mining by soft computing methods for the coronary heart disease database. In: Fourth International Workshop on Computational Intelligence and Application, IEEE SMC Hiroshima Chapter, Hiroshima University, Japan, 10–11 December (2008)
3. Rajkumar, Reena, G.S.: Diagnosis of heart disease using datamining algorithm. Glob. J. Comput. Sci. Technol. **10**(10) (2010)
4. Lenert, L., Lin, A., Olshen, R., Sugar, C.: Clustering in the Service of the Public's Health http://www-stat.stanford.edu/olshen/manuscripts/helsinki.PDF
5. Srinivas,K., Rani, B.K., Govrdhan, A.: Applications of data mining techniques in healthcare and prediction of heart attacks. Int. J. Comput. Sci. Eng. (IJCSE). **02**(02), 250–255 (2010)
6. Yan, H.: Development of a decision support system for heart disease diagnosis using multilayer perceptron. In: Proceedings of the 2003 International Symposium, vol. 5, pp. 709–712 (2003)
7. Sitar-Taut, V.A.: Using machine learning algorithms in cardiovascular disease risk evaluation. J. Appl. Comput. Sci. Math. (2009)
8. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI Mag. **17**(3), 37 (1996)
9. Balasubramanian, T., Umarani, R.: An analysis on the impact of fluoride in human health (dental) using clustering data mining technique. In: Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, 21–23 March 2012
10. De Falco, I.: Differential Evolution for automatic rule extraction from medical databases. Appl. Soft Comput. **13**(2), 1265–1283 (2013)
11. Belciug, S., Gorunescu, F., Salem, A., Gorunescu, M.: Clustering-based approach for detecting breast cancer recurrence. In: 10th International Conference on Intelligent Systems Design and Applications (2010)
12. Vital, T., Panduranga, et al.: Data collection, statistical analysis and clustering studies of cancer dataset from viziayanagaram District, AP, India. ICT and critical infrastructure. In: Proceedings of the 48th Annual Convention of Computer Society of India-Vol II. Springer, Cham, (2014)
13. Douglas, P.K., Harris, S., Yuille, A., Cohen, M.S.: Performance comparison of machine learning algorithms and number of independent components used in fMRI decoding of belief versus disbelief. Neuroimage **56**(2), 544–553 (2011)
14. Rekha Saxena, B.N.: Nagpal, M.K. Das, Aruna Srivastava, Sanjeev Kumar Gupta, Anil Kumar, A.T. Jeyaseelan, and Vijay Kumar Baraik. A spatial statistical approach to analyze malaria situation at micro level for priority control in Ranchi district, Jharkhand. Indian J. Med. Res. **136**(5), 776–782 (2012)
15. Andreeva, P.: Data modelling and specific rule generation via data mining techniques. In: International Conference on Computer Systems and Technologies – CompSysTech (2006)
16. Hamerly G., Elkan C.: Learning the K in K-means. In: Proceedings of the 17th Annual Conference on Neural Information Processing Systems, British Columbia, Canada (2003)

Performance Analysis of Array Multiplier Using Low-Power 10T Full Adder

D. S. Shylu Sam and A. Christina Roseline

Abstract Multipliers are used in arithmetic and logic unit (ALU) and the performance of multiplier greatly depends on the number of adder cells. A study and performance investigation of array multiplier using low-power 10T full adder is presented. One of the crucial circuits in all digital devices is the adder cell and it is used in ALU. In this work, various performance parameters including power consumption, delay, and power-delay product are compared with full adder, array multiplier using 10T full-adder and static energy-recovery full (SERF) adder-based techniques. Array multipliers realized that using 10T-based full-adder circuit consumes less power when compared to SERF techniques. Simulation results show that the power consumption of array multiplier using 10T full-adder and SERF technique is 673.7 μW and 1.042 mW, respectively, which is less than the power consumption of conventional array multiplier. The power consumption and delay of proposed Array Multiplier are obtained using 180nm CMOS process.

1 Introduction

In present electronic period, various calculations are executed using low-power adder cells [1]. In this aspect, multiplier plays a vital role in arithmetic unit including microprocessors, microcontrollers, and various digital signal processors (DSPs). In order to reduce the power consumption, the usage of transistors in adder cells is reduced which reduces the overall power consumption of array multiplier.

D. S. Shylu Sam (✉)
Department of ECE, School of Engineering & Technology, Karunya Institute of Technology and Sciences, Karunya Nagar, 641114 Coimbatore, Tamil Nadu, India
e-mail: mail2shylu@yahoo.com

A. C. Roseline
Department of ECE, School of Engineering & Technology, Karunya Institute of Technology and Sciences, Karunya Nagar, 641114 Coimbatore, Tamil Nadu, India
e-mail: chriztinarozeline@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_28

Series of adder together forms the multiplier circuit [2, 3]. The main factor in digital-based CMOS circuits can be categorized as: dynamic power and static power.

Dynamic power of a logic gate can be defined as the power consumption of the logic gate in ON state. The dynamic power can be minimized by reducing the switching activity and load capacitance. Selection of logic styles, transistor scaling and design optimizations reduces the load capacitance of the design. At all the levels of design abstraction, the switching activity can be reduced.

During the circuit operation, short-circuit current flows from the supply and rails to the ground when the transistor switches between the states. This static power dominates more in submicron technology.

Static power causes a rise in device leakage current and static current when the design is in steady state. When the signals switch its state effects in a main power consumption component referred as dynamic power.

The paper is organized as follows: Section 2 provides full description of 10T full-adder cells. Section 3 provides a thorough explanation of array multiplier. In Sect. 4, performance analysis is done based on the simulated results. Conclusion and references are discussed in Sects. 5 and 6, respectively.

2 Circuit Description

2.1 Full Adder

Full adder is a circuit that adds three input binary numbers and produces two outputs which are shown in Fig. 1. Let A, B, and C be the three input ports which are provided with logic 0s and 1s as inputs [4, 5]. Let SUM and COUT be the two output ports as shown in Eqs. 1 and 2. Output is said to be high when it is logic 1, and low when it is logic 0. The full-adder Boolean equation is as follows:

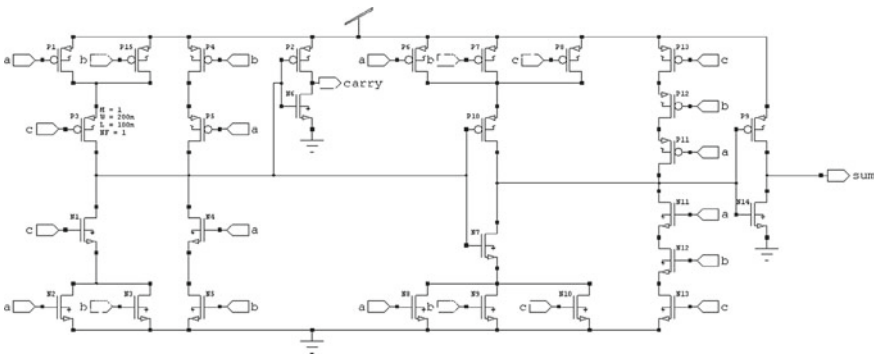


Fig. 1 Conventional full adder [4]

$$\text{SUM} = A \oplus B \oplus C \tag{1}$$

$$\text{COUT} = AB + BC + CA \tag{2}$$

Full adder is formed by the arrangement of PMOS and NMOS transistors in different positionings. PMOS is said to be ON, when 0 is given as input, and it is said to be OFF, when 1 is given as input. The bulk of PMOS transistor is connected to VDD. NMOS is said to be in ON condition, when 1 is given as input, and in OFF condition, when 0 is given as input. The bulk of NMOS is connected to GND.

2.2 SERF Full Adder

The initialism of static energy-recovery full adder is SERF [6]. It operates at high frequency, and fails to work at the voltage below 0.3 V. The number of transistors is reduced to ten due to which there is a drastic decrease in power consumption [7–9]. It is designed by XNOR gate based full-adder circuit. It has no direct path to the ground due to which there is a reduction in power consumption. As shown in Fig. 2, the input ports A and B are given to XNOR gate, whereas C is given to 2T multiplexer, and the output is driven from the output ports SUM and COUT based on the working of transistors as mentioned above.

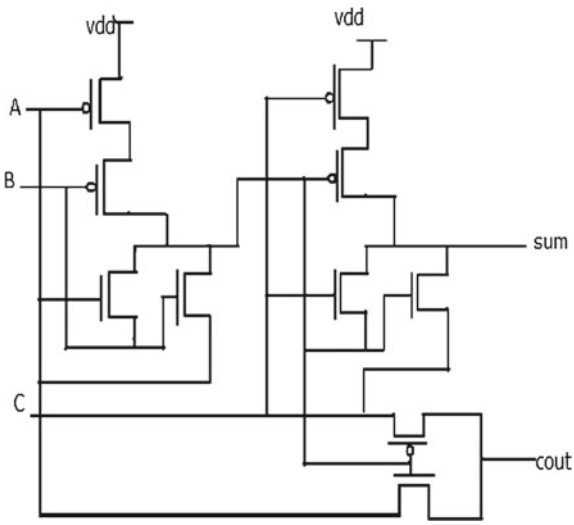


Fig. 2 SERF adder [6]

2.3 10T Full Adder

This 10T full adder consists of XOR gate and 2T multiplexer which is known as data selector. The multiplexer is implemented by the combination of NAND, OR, and NOT logic gates as shown in Fig. 3. Input ports A and B are given to XOR gate, whereas C is given to 2T multiplexer, and finally, the outputs are driven as SUM and COUT based on the working of transistors as mentioned above. [10]

2.4 Multiplier

Multiplier is the important block in digital signal processing (DSP) and in other applications. VLSI design engineers designed the multipliers to meet the design specifications including low power, less area, and high speed with the regularity of layout for compact VLSI implementation. Parallel multiplier comprises of partial products to be added which is the foremost parameter that defines the performance of multiplier.

2.4.1 Array Multiplier

Array multiplier is based on the traditional multiplication and is preferred over other multipliers like Wallace tree multiplier, Booth algorithm-based multiplier, and serial–parallel multiplier due to its simple structure. P0, P1, P2, P3, P4, P5, P6, and P7 are the outputs of array multiplier as shown in Fig. 4. Figure 5 shows the methodology of array multiplier.

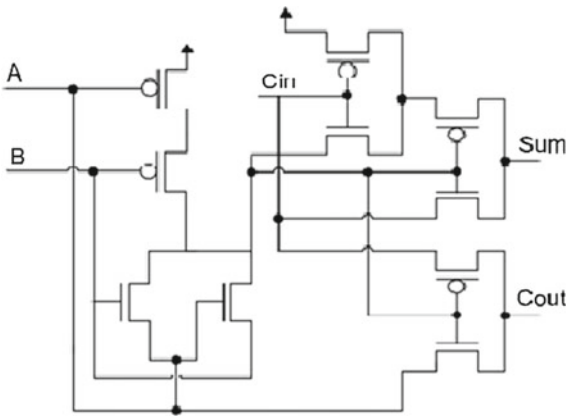


Fig. 3 10T full adder [10]

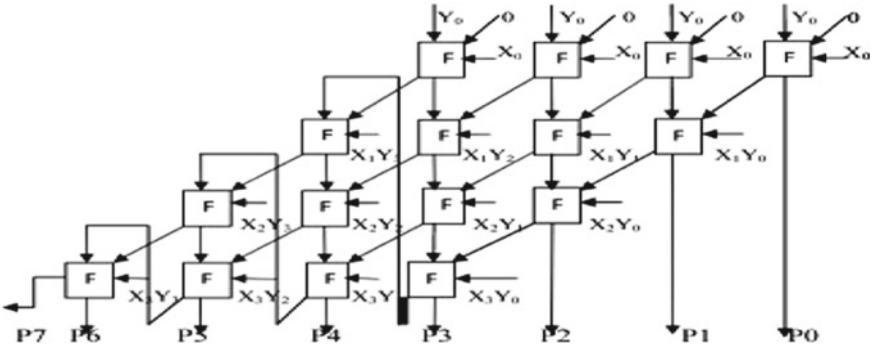


Fig. 4 Structure of array multiplier

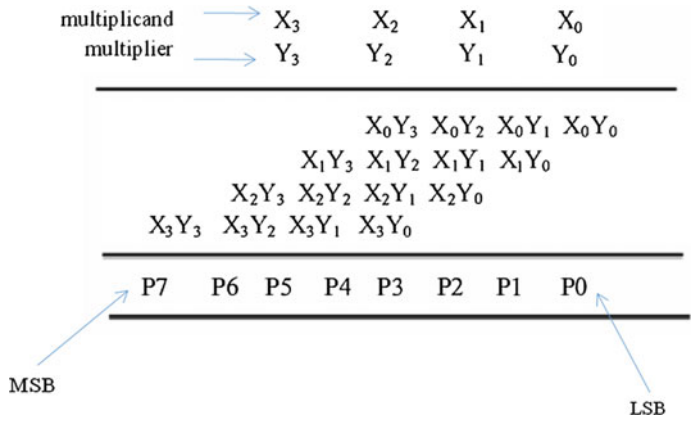


Fig. 5 Methodology of multiplication

2.4.2 Operation of Array Multiplier

Working of array multiplier is as simple as basic multiplication. Generally, the first row of array multiplier can be of half adder or full adder. In the present work, the whole array multiplier circuit is designed using full-adder circuit (SERF and 10T). With the inputs given in bit stream, the carry term generated from each adder circuit is diagonally preceded to the next row of the adder [11, 12].

3 Proposed Array Multiplier

As discussed in above chapters, an array multiplier using low-power 10T has been designed as shown in Fig. 6. The performance of the designed array multiplier has been compared with that of the array multiplier using SERF adder. It is found that the performance of the array multiplier using 10T full adder is far better than the array multiplier using SERF adder. The reason for the increased performance of 10T full adder is due to the positioning of transistors.

Thus, by implementing the array multiplier using 10T full adder in any system can reduce the power and improve the speed of that system. The 10T full adder consumes less power [13, 14]. As the power consumption of adder used in array multiplier is reduced, this in turn reduces the power consumption of array multiplier. This array multiplier can be used in various devices like CRO, microprocessor, and microcontroller [15]. Thus, for the designed array multiplier, it can be concluded that the number of transistors and positioning of transistors play a major role.

4 Results and Discussion

4.1 Simulation of Array Multiplier Using Full Adders

Array multiplier using 10T full adders (SERF and 10T) is simulated and designed using Virtuoso tool in Cadence. The design is simulated using 180 nm CMOS process, and bit-streaming is done for all the combinations of inputs along with rise time and fall time of 1 picosecond, the time period is provided as 10 ns. At the end of the transient analysis, power and delay are calculated and waveform is obtained as follows.

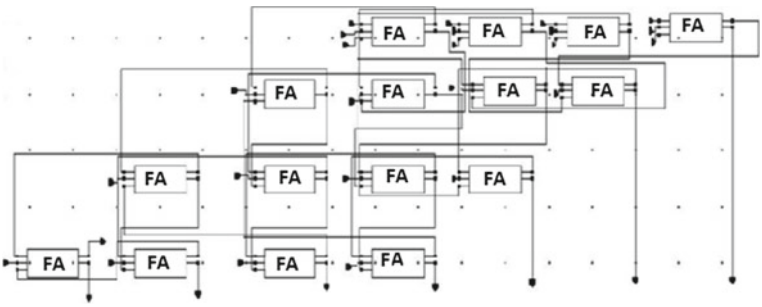


Fig. 6 Array multiplier using low-power 10T full adder

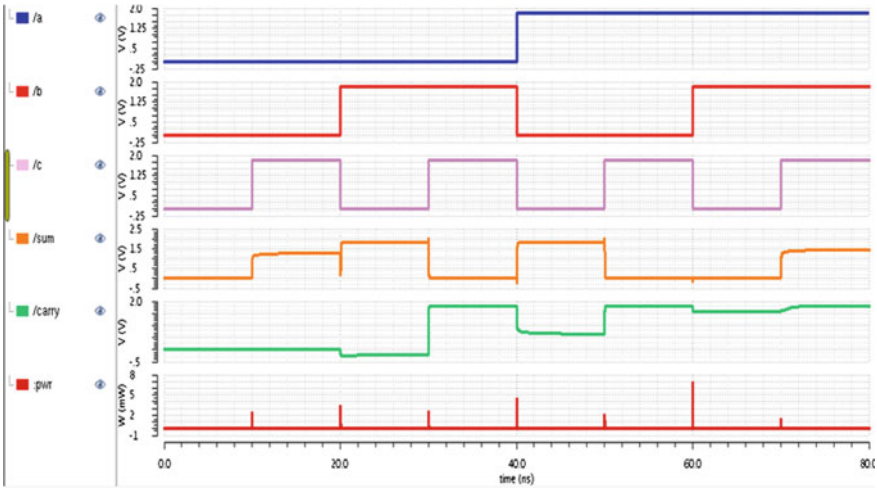


Fig. 7 SERF Full adder

4.2 *Waveform of Full Adder*

The waveform of full adders is obtained as shown in Figs. 7 and 8 with A, B, and C as inputs, SUM, and COUT as outputs.

4.3 *Performance Analysis*

Power consumption, delay and power-delay product are the factors which determine the performance of the array multiplier.

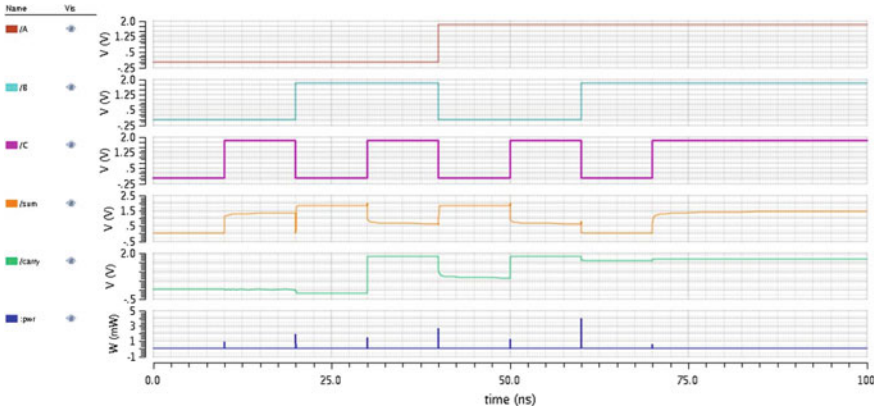


Fig. 8 10T full adder

Table 1 Full-adder circuit

Full-adder circuit	Power consumption (w)	Delay (s)	Power-delay product (J)
SERF	5.281E-6	658.8E-6	3.479E-9
10T	4.325E-6	20.54E-9	8.883E-13

Table 2 Array multiplier circuit

Full-adder circuit	Power consumption (w)	Delay (s)	Power-delay product (J)
SERF	1.099E-3	20.26E-12	2.226E-14
10T	673.7E-6	20.09E-9	1.3534E-10

Thus, from Table 1, it is found that the power consumption of 10T full-adder circuit is reduced in comparison with SERF adder-based circuit. The delay obtained for 10T-based adder is also less when compared with the conventional full adder. As shown in Table 2, the power of designed array multiplier with 10T full-adder cell is reduced and can be used for processors.

5 Conclusion

CMOS design is based on various design parameters including power and delay analysis. In the present work, the above-mentioned parameters for SERF-based full-adder circuit are compared with 10T full-adder circuit. Simulation results show that the designed array multiplier consumes less power and can be used as the building blocks for various processors.

Acknowledgements The authors are grateful to the management of Karunya Institute of Technology and Sciences (KITS) for providing necessary facilities in VLSI laboratory to carry out this work.

References

1. Ravi1, N., Satish, A., Dr. Jayachandra Prasad, T., Dr. Subba Rao, T.: A new design for array multiplier with trade off in power and area. IJCSI, **8**, 1–5 (2011)
2. Vamsi Krishnai, B., Dhaunjaya, K.: Analysis and modeling of low power array multipliers using cadence virtuoso simulator in 45 nm technology. IJECIERD. **3**(4) (2013)
3. Jhamb, M., Garima, Lohani, H.: Design implementation and performance comparison of multiplier topologies in power-delay space. Jestech. 355–363 (2015)
4. Shams, A.M., Darwish, T.K., Bayoumi, M.A.: Performance analysis of low-power 1-Bit CMOS full adder cells. IEEE Trans. On VLSI Syst. **10**, 20–29 (2002)

5. Mathew, K., Asha Latha, S., Ravi, T., Logashanmugam, E.: Design and analysis of an array multiplier using an area efficient full adder cell in 32 nm CMOS technology. IJES. **2,** 8–16 (2013)
6. Navi, K., Maeen, M., Hashemipou, O.: An energy efficient full adder cell for low voltage. IEICE. **2**(2), 553–559
7. Manjunath, K.M.: Analysis of various full-adder circuits in cadence, 30–38 (2015)
8. Navi, K., Kavehei Shahid, O.: Low-power and high-performance 1-Bit CMOS full-adder cell. J. Comput. **3**(2) (2008)
9. Yadav, M., Gupta, A., Rai, S.: A study and comparative analysis of low power hybrid-CMOS 1-bit full adders in deep-submicron technology. In; IEEE International Conference in Power, Control and Embedded Systems, pp. 1–7, ICPCES (2017)
10. Bui, H.T., Wang, Y., Jiang, Y.: Design and analysis of low-power 10-transistor full adders using novel XOR XNOR gates. IEEE Trans. Circuits Syst.-II: Analog Digital Signal Proc. **49**, 25–30 (2002)
11. Kumar Pankaj and: R. K. Sharma: Low voltage high performance hybrid full adder, Engineering Science and Technology an. International Journal **19**(1), 559–565 (2016)
12. Hernandez, M.A., Linares-Aranda, M.: CMOS full-adders for energy-efficient arithmetic applications. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **19**(4) (2011)
13. Chang, C.-H., Gu, J., Zhang, M.:A review of 0.18 μm full Adder performances for tree structured arithmetic circuits. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **13**(6), 686–695 (2005)
14. Goel, S., Kumar, A., Bayoumi, M.A.: Design of robust energy-efficient Full Adders for deep-sub micrometer design using hybrid-CMOS logic style. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **14**(12), 1309–1321 (2006)
15. Zhang. M., Gu, J., Chang, C. H.: A novel hybrid pass logic with static CMOS output drive full-adder cell. Proc. IEEE Int. Symp. Circuits Syst. 317–320(2003)

A Secure and Lightweight Authentication Protocol for Multiple Layers in Wireless Body Area Network

N. V. Abiramy and S. V. Sudha

Abstract Wireless body area network (WBAN) is one of the emerging technologies in IoT, and in healthcare service, people experience better life and effective medical service through this; however, the medical system and privacy of patient's health information are very sensitive, and it is significant to give privacy and security in WBAN. In previous studies of WBAN, according to the features of network structure, the authentication protocol was not designed appropriately. In this paper, we design authentication protocol for multiple layers. First level is between the sensor nodes and mobile terminal using anonymous authentication protocol, and the second level is between the mobile terminal and application provider using certificateless authentication without pairing protocol. In the proposed work, to reduce computational cost, elliptic curve cryptography (ECC) algorithm is used. It is proved that the proposed protocol is secure and efficient by security and performance analysis.

1 Introduction

People's life is becoming very convenient because of the development of Internet of things (IoT). The IoT has extensive growth in applications, for example, smart grid, smart city, telemedicine, industrial Internet, intelligent transportation and smart farming. Particularly, it carries greater invention in smart healthcare systems [1]. The WBAN is the evolving network that can deliver emergency medical assistance, remote medical treatment and remote monitoring. Wireless sensor nodes are used in WBAN that can be implemented in three ways, and they are (1) implantable (2) wearable and (3) invadable sensors [2]. For instance, a wearable pulse oximeter

N. V. Abiramy (✉) · S. V. Sudha
Computer Science and Engineering, Computer Science and Engineering, Coimbatore, Tamil Nadu, India
e-mail: abigoldenmani@gmail.com

S. V. Sudha
e-mail: svsudha.mvenki@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_29

sensor is used to check the blood oxygen level or pulse rate, an implantable blood glucose sensor is used to check the glucose level in blood, and an ECG sensor is used to detect the heart rate of patients and so on [3].

The wireless body area network depends on the wireless sensor networks [4]. But, the variance in WBAN is that it has sensors deployed in, on and around the body a mobile terminal and a remote control centre [5]. There are different types of sensors in WBAN that monitor various physiological information in the human body without disturbing their comfort and forward the collected information to remote control centre through mobile terminal, and at last, the remote control centre examines and reports the information to meet diverse requirements [6].

Nowadays, WBAN is becoming the most popular research topic. But, the major problems in WBAN are privacy and security, robustness and scalability [7]. WBAN collects and forwards sensitive information of patient's health data, so the security is more important, and reducing computational cost is a challenging task [8].

This paper is proposed with lightweight authentication protocol for multiple layers using an ECC algorithm [9]. Here we consider two protocols, anonymous authentication and certificateless authentication without pairing protocol.

2 Related Work

In previous papers of WBAN, public key cryptography is used as an authentication protocol between MT (mobile terminal) and AP (application provider). To avoid complications in handling public key cryptosystem, an identity-based cryptography is created. Even though the digital certificate is not needed for the identity-based cryptography, it becomes diffident when KGC is attacked. After that an certificateless cryptography was introduced in WBAN [10].

A new public key cryptography called certificateless public key cryptography was first introduced by Al-Riyami and Paterson in 2004 [11]. Reference [12] A lightweight and effective certificateless authentication protocol that depends on certificateless signature method was introduced in 2013 by Jingwei Liu et al. However, during the session phase, this protocol allows the attackers to track the information of users [13]. The two protocols that are designed above have the difficulty of public replacement attack, and [14] to overcome that, a lightweight certificateless cryptography was proposed by Hu Xiong et al. Later, a cloud assistant-based certificateless public auditing scheme for WBANs which provide protection for data integrity was introduced by Debiao He et al. in 2015 [15]. By implementation, pairing operation in certificateless authentication will reduce the efficacy, as it takes high computational cost. To solve the above-stated problem, this paper proposed a certificateless authentication without pairing protocol, and the major benefit of certificateless authentication is to overcome the key escrow problem. The proposed two protocols are built on the encryption of ECC algorithm which uses squat key size, and it is also more efficient.

3 Preliminaries

The notations, ECC, and intractability problems are discussed below.

3.1 Elliptic Curve Cryptography

Related to Rivest–Shamir–Adleman (RSA) which is a conventional public key cryptosystem, the research displays that elliptic curve cryptography needs key size of 160 bits to give the highest security, but RSA needs a key size of 1024 bits to give the same security level [16–18]. ECC wants less storage space and low bandwidth, and compared to RSA, the ECC was very fast and effective in the computational time. In this paper, the elliptic curve is used to do encryption and authentication process. The elliptic curve is a plane curve that is defined by $EC: y^2 = x^3 + ax + b$ mod P over prime finite field Fp with the discriminant $4c^3 + 27d^2 \neq 0$.

3.2 Intractability Problems

To guarantee the proposed protocols in this paper, we took three intractability problems that are shown in Eq. 1, 2, 3 that are based on ECC.

Discrete logarithmic elliptic curve problem (DLECP): Assume that a point X is on $EC(Fp)$, Y is the various point of X on $EC(Fp)$, an existing integer $m > 0$, forms $Y = mX$. Under the basis of known X and Y, m is difficult to compute.

$$Prob[m \in Fp | (X, Y) \in EC(Fp) \wedge \mathrm{Y} = \mathrm{mX}] = \varepsilon \tag{1}$$

Here the minimum value ε is negligible.

Diffie–Hellman Computation Problem (DHCP): Assume that the points $X, m.X$ and $n.X$ lying on $EC(Fp)$ where $m, n \in Zx$ in polynomial time, it is hard to find $m.n.X$ by the above conditions.

$$\mathrm{Prob}[\mathrm{m}, \mathrm{n} \in \mathrm{Zx} | (\mathrm{X}, \mathrm{m.X}, \mathrm{n.X}) \in \mathrm{EC(Fp)} \wedge (\mathrm{m.n.X}) \in \mathrm{EC(Fp)}] = \varepsilon \tag{2}$$

ε is negligible minimum value during polynomial time.

Factorization in Elliptic Curve Problem (FECP): Assume that two points X, Y on elliptic curve $EC(Fp)$ where $Y = m.P + n.P$ with $m, n \in Zx.mX, nP$ on $EC(Fp)$ is hard to find.

$$Prob[m, n \in Zx | (Y = smX + nX) \in EC(Fp) \wedge (mX, nX) \in EC(Fp)] = \varepsilon \tag{3}$$

The minimum value ε is negligible.

4 Model

Here we consider two levels of communication in WBAN to achieve a reliable communication and multilayer authentication protocol. Subsequently, we will show the process of communication in WBAN and will discuss about the attacks that the system come across during the communication.

4.1 Architecture of the System

Here, we divide two levels of communication in WBAN: first level involves inside authentication between sensor nodes (SN) and mobile terminal (MT), and the next level involves outside authentication between mobile terminal and application provider (AP).

Figure 1 shows the architecture of the system. Here two different protocols were designed for each level. The first level of communication is between the sensors and mobile terminal; clearly, the first single sensor should be considered with limited resources, which makes the design of the protocol. The mobile terminal has sufficient storage and computational resources, and here, it acts as a cluster head. The protocol in first level has artificial interference, but the second level does not have, and it is completely wireless communication so that it is necessary to give authentication between mobile terminal and application provider.

5 Proposed Authentication Protocol

Here, we commenced two proposed authentication protocols for better security in WBAN.

5.1 Anonymous Authentication Protocol

We propose a scalable anonymous authentication protocol between SN and MT. Here, we consider two phases. The first is the initialization phase in which parameters are generated, and the second is the registration phase in which private and public keys are generated and then group keys are established; these are shown in Fig. 2.

(1) **Initialization:** Let us consider n is a prime number produced by MT and it distributes parameters $(En, F/En, N)$ then MT selects master private key randomly by $p \in Zm^*$ and public key is computed as $N_{MT} = p.N$. MT selects a

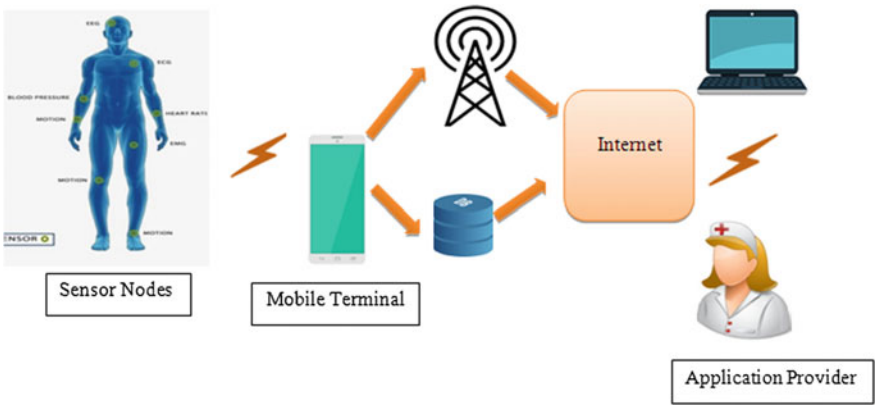


Fig. 1 System architecture

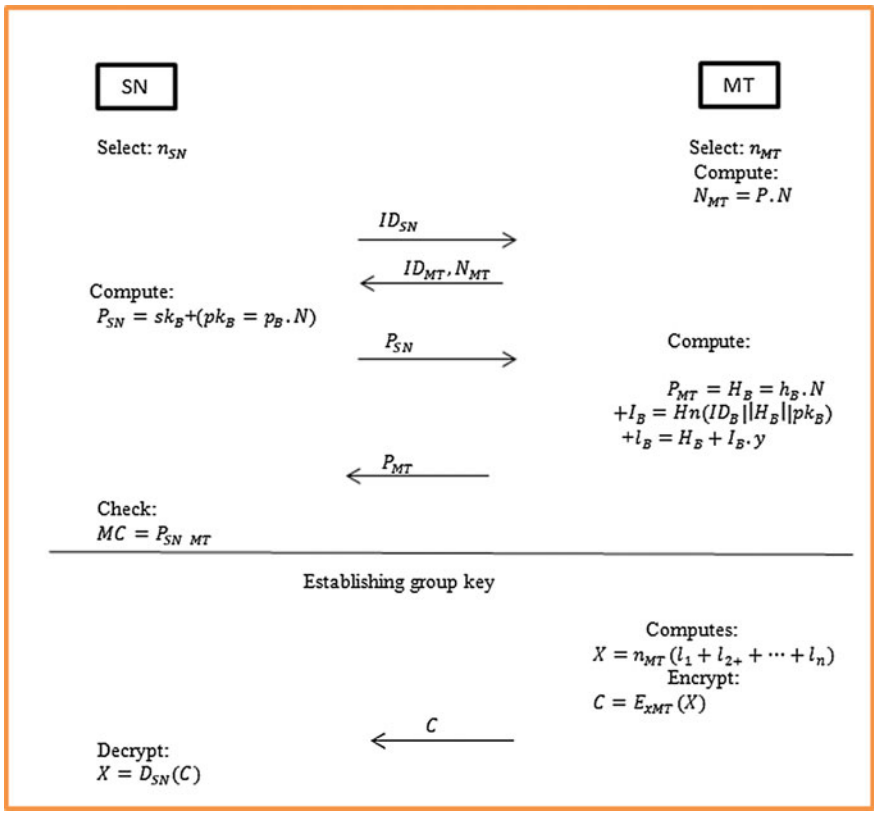


Fig. 2 Authenticating and establishing group keys between SN and MT

Table 1 Notations and their representations

Notations	Representations
SN	Sensor nodes
MT	Mobile terminal
AP	Application provider
PP	Public key pair
PR	Private key pair
SK	Session key
A	A generator point on E
Hn	Hash function

message authentication code and secured hash function $Hn: (0,1)^* \rightarrow Zm^*$ then the parameters are distributed as system parameters and send to SN.

(2) **Registration:** The SN with the identity ID_B selects secret value $p_B \in Zm^*$ and defines secret key as sk_B and public key as $pk_B = p_B.N$. The public key and identity of SN are sent to MT.

After receiving message from SN, the MT chooses random value as $h_B \in Zm^*$ later MT computes $H_B = h_B.NI_B = Hn(ID_B||H_B||pk_B)$, $l_B = H_B + I_B.y$. (as mentioned in Table 1)

(3) **Establishing group key:** The group key of MT is calculated by $X = n_{MT}(l_1 + l_{2+} + \ldots + l_n)$ and use C as a secret value to encrypt the data, then value of C is sent to SN and it decrypts C using its secret value to get X.

5.2 *Certificateless Authentication Without Pairing Protocol*

We proposed a novel security protocol called certificateless authentication without pairing and session key establishment between MT and AP. There are three phases in this authentication protocol. Initialization is the first phase in which system parameters are developed by KGC. The next is the registration phase, in which MT and AP are indexed to KGC and also they independently generate public and private keys, and last is the authentication phase in which MT and AP discuss the session key and they authenticate each other; these are shown in Fig. 3.

(1) **Initialization:** The system initialization and generating security parameters are done by KGC. Consider a prime number $x, y, y|x-1$ and choose elliptic curve $EC: y^2 = x^3 + ax + b$ mod P where $4c^3 + 27d^2 \neq 0$, further $c, d \in Fp$. Next select hash function, $Hn1: (0,1)^* \times [m,n] \rightarrow Zx, Hn2: (0,1)^* \times (0,1)^* \times [m,n] \times [m,n] \rightarrow G$. $m \in Zx$ is system master key chosen by KGC, system public key is computed as $Q = mA$ and m is kept secret, $(x, y, A, EC(Fp), Q, Hn1, Hn2)$ are the public parameters.(as mentioned Table 1)

(2) **Registration phase:** MT and AP index themselves to KGC. A random integer s_{MT} is selected by KGC after MT gets request, $PP_{MT1} = s_{MT}A, PR_{MT1} = s_{MT} +$

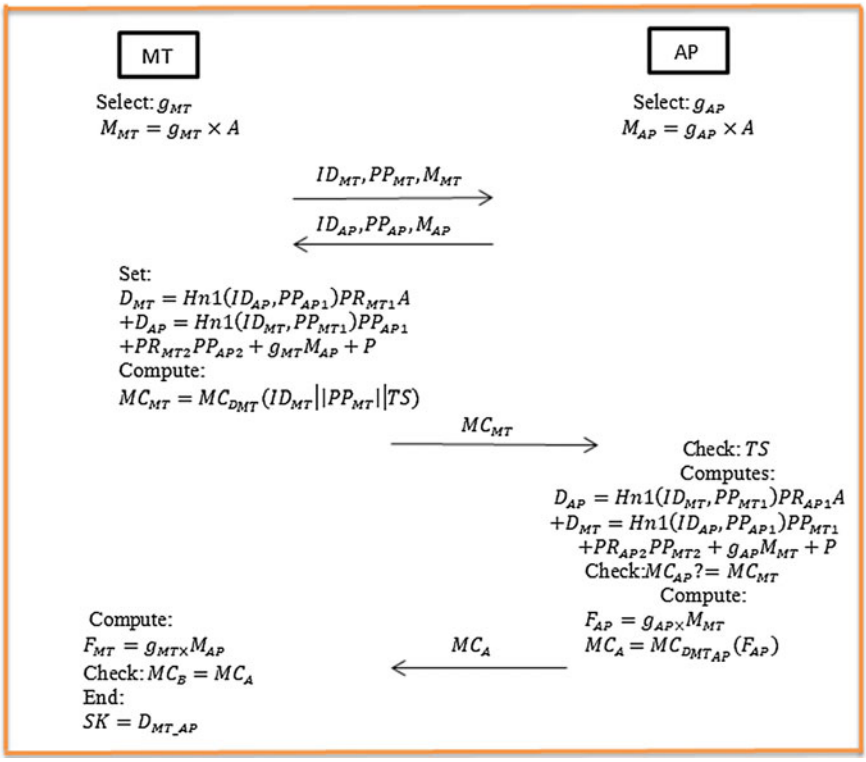


Fig. 3 Authentication between MT and AP

$mHn1(lC_{MT1}, PP_{MT})$ are the computed partial public and private keys sent to MT, then it produces public and private keys. Take a secret value by considering the random integer r_{MT} and set private key $PR_{MT} = (PR_{MT1}, PR_{MT2}) = (PR_{MT}, r_{MT}), PP_{MT} = (PP_{MT1}, PPMT2) = (PP_{MT1}, r_{MT}A)$.it is similar for AP, take s_{AP} as random integer and r_{AP} as secret value, then AP sets private and public keys as $PR_{MT} = (PR_{AP1}, PR_{AP2}) = (PR_{AP}, r_{AP}), PP_{AP} = (PP_{AP1}, PP_{AP2}) = (PP_{AP1}, r_{AP}A)$.

(3) **Authentication phase:** MT and AP authenticate each other and generate a session key. The random integers selected by MT and AP are g_{MT} and g_{AP}. For authentication, MT computes encryption key, then with timestamp TS it produces message authentication code (MC) MC_{MT} and sends to AP. Then, AP checks TS and it computes encryption key and using that it encrypts MC and checks MC_{MT} is equal to MC_{AP}, if it is equal, the session key is generated.

6 Security Analysis

Forward security: Even though the user's private key is ruined, the session key will not be compromised.

Session key secrecy: The only requested MT and AP will have accurate session keys. In the proposed protocol, public and private keys are used to generate session keys between MT and AP.

Key escrow resilience: It is given by certificateless protocol to avoid malicious KGC that eavesdrops MT and AP public and private keys. In the proposed work based on DHCP, computationally KGC is not able to act as an authorized user.

7 Performance Analysis

As we mentioned in Sect. 3.2, ECC is better than RSA in computation, key size and security. Key size of 160 bits takes 1.61 s in ECC for point multiplication; meanwhile, RSA with key size of 1024 bits takes 22 s for modular exponentiation. Table 2 displays that the first authentication protocol, which based on ECC algorithm uses hash operation with a 3-point multiplication (PM) and 1 hash function (Hn), compared to Li's protocol which uses $(3 + SN)$ exponentiation (E) and 9 hash functions, anonymous protocol reduces computational cost [19]. Table 3 displays that using pairing operation causes great computational cost, so here we design certificateless authentication without pairing protocol in which pairing operation (PO) is not done, and it takes 2E,6PM and 2Hn as compared to Xiong's protocol which uses RSA and 4PO,14E,4PM and 6Hn for computation; the certificateless authentication without pairing protocol reduces the computational cost [20].

Table 2 Computation cost for anonymous authentication protocol

Algorithm	Computation cost of each node
Li's protocol	(3 + n) E + 9 Hn
Anonymous authentication protocol	3PM + 1Hn

Table 3 Computation cost for certificateless authentication without pairing protocol

Algorithm	Computation cost of each node
Xiong's protocol	4PO + 14E + 4PM +6Hn
Certificateless authentication without pairing protocol	0PO + 2E + 6PM + 2Hn

8 Conclusion

In medical care, WBAN is the small size network, and it has an extensive variety of applications. In this paper, we use ECC algorithm to give secure authentication protocol for multiple layers. The proposed work security is based on three intractability problems, namely DLECP, DHCP and FECP, and it will protect confidentiality, integrity and authenticity of data in WBAN. Accordingly, it gives mutual authentication between SN and MT and between MT and AP. The comparisons prove that proposed protocol is efficient, reduce computational cost and have greater flexibility.

References

1. Nidhya, R., Arunachalamand, V.P., Karthik, S.: A study on requirements, challenges and applications of wireless body area network. Asian J. Electr. Sci. **6**(2), 30–36 (2017)
2. Seyedi, M., Kibret, B., Lai, D.T.H., Faulkner, M.: A survey on intrabody communications for body area network applications. IEEE Trans. Biomed. Eng. **60**(8) 2067–2079 (2013)
3. Latr, B. Braem, I.Moerman, C.Blondia, P.Demeester.: A Survey on wireless Body Area Networks. In: J.Wirel. Netw, pp. 1–8 (2011)
4. Zhou, B., Hu, C., Wang, H., Guo, R.: A wireless sensor network for pervasive medical supervision. In: International Conference on Integration Technology, pp. 740–744 (2007)
5. Hanson, M.A., Powell, H.C., Barth, A.T., Ringgenberg, K., Calhoun, B.H., Aylor, J.H., Lach, J.: Body area sensor networks: challenges and opportunities. J. Comput. **42**(1), 58–65 (2009)
6. Xia, Z., Wang, X., Qin, Z., Ren, K.: A privacy-preserving and copy-deterrence content-based image retrieval in cloud computing. IEEE Trans. Forensics Sec. 2594–608 (2016)
7. Sudha, R., Nivetha, P.: A study on security in wireless body sensor networks. Int. J. Innovative Res. Comput. **4**, 1376–1380 (2016)
8. Liao, X., Shu, C.: Reversible data hiding in encrypted images based on absolute mean difference of multiple neighboring pixels. J.Vis. Commun. 21–28 (2015)
9. Nidhya, R., Arunachlam, V.P., Karthick, S.: Multitier security architecture with energy requirement for medical wireless sensor network. Int. J. Printing Packag. Allied Sci. **4**(5), 3885–3897 (2016)
10. Tan, C., Wang, H., Zhong, S., Li, Q.: IBE-lite: A lightweight identity-based cryptography for body sensor networks. IEEE Trans. Inf. Technol. Biomed. **13**, 926–932 (2009)
11. Al-Riyami, S.S., Paterson, K.G.: Certificateless public key cryptography, pp. 452–473 Springer, Berlin
12. Zhang, L., Liu, J., Sun, R.: An efficient and lightweight certificateless authentication protocol for WBAN. In: International Conference on INCS. pp. 637–639 (2013)
13. Jaiswal, P., Kumar, A.: Design of queue-based group key agreement protocol using elliptic curve cryptography, pp. 167–176. Springer (2009)
14. Xiong, H.: Cost effective scalable and anonymous ceretificateless remote authentication protocol. In: IEEE Trans. Inf. Forensics Secur. 2327–2339 (2014)
15. He, D., Zeadally, S., Wu, L.: Certificateless public auditing scheme for cloud assisted wireless body area networks. IEEE Syst. J. 1–10 (2015)
16. Agarwal, H., Badadapure, P.R.: A survey on elliptic curve cryptography. Int. Res. J. Eng. Technol. 2014–2018 (2016)
17. Chaudry, H.A., Naqvi, H., Mahmood, K., Ahmad, H.F., Khan, K.: An improved remote user authentication scheme using elliptic curve cryptography. Wirel. Commun. 1–19 (2016)

18. Chaudhry, S.A., Khan, M.T., Khan, M.K., Shon, T.: A multiserver biometric authentication scheme for tmis using elliptic curve cryptography. J. Med. Syst. 40–54(2016)
19. Li, M., Yu, S., Lou, W., Ren, K.: Group device pairing based secure sensor association and key management for body area networks. ICC Commun. 2651–2659 (2010)
20. Xiong, H., Qin, Z.: Revocable and scalable certificateless remote authentication protocol with anonymity for wireless body area networks. IEEE Trans. Inf. Forensics (2015)

Energy Characterization of Bluetooth in Opportunistic Mobile Crowdsensing Platform

Kalyani Sahoo, Ramesh K. Sahoo and Srinivas Sethi

Abstract Smart environment services provide latest information and communication technologies to boost the different activities of citizens. Sensing is one of the necessary parameters to observe present status of infrastructures and environment. In this paper, it has been planned to analyze the performance of the Bluetooth with its characterization in mobile crowdsensing platform. Bluetooth has emerged as a promising platform for short-range wireless networking. It has been proposed a model to find out the energy spent and power consumption by considering three parameters: Bluetooth, Wi-Fi, and LTE.

1 Introduction

Crowdsensing is a technique that enables to assemble a huge quantity of data by permitting an extensive variety of data sources from the crowd to contribute data, and sometimes Crowdsensing is referred as mobile Crowdsensing (MCS)[1, 2]. The sensing and computational ability of mobile devices increase with the development of mobile technology. According to involvement, Crowdsensing is classified into 2 types: participatory Crowdsensing and opportunistic Crowdsensing. The users voluntarily involve by contributing information in participatory Crowdsensing, whereas the data can be sensed, collected, and shared without any user intervention in opportunistic Crowdsensing. Data collection, data storage, and data upload are three processes in mobile Crowdsensing. All three processes are interrelated to each other.

K. Sahoo (✉) · R. K. Sahoo · S. Sethi
Department of CSEA, IGIT Sarang, Dhenkanal, Odisha, India
e-mail: kalyanisahoo94@gmail.com

R. K. Sahoo
e-mail: ramesh0986@gmail.com

S. Sethi
e-mail: srinivas_sethi@igitsarang.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_30

In mobile Crowdsensing, user contributes data which are created from sensors embedded in mobile devices and may be in the form of IoT devices [3]. Accelerometer, magnetometer, GPS, and camera are just descriptive of a set of sensors which are nowadays employed to work different applications in many domains. The information acquired through mobile Crowdsensing platforms is aggregated and delivered to a collector located in the cloud. This enables so-called sensing as a service model. This makes the collected public data available to developer and end user. IoT and MCS are key enablers in so-called sensing as a service model. Efficiency of this model is defined in terms of the revenues obtained from selling data versus the cost of the sensing campaign.

MCS leverages active participation of citizens in sensing activities which are required to contribute data from sensor of their mobile devices. This contribution is reported using commonly available cellular, Bluetooth, or Wi-Fi communication technology to a data collector usually placed in the cloud for data processing and its analysis [2]. Data acquisitions in MCS can be either participatory or opportunistic. Bluetooth is designed for an inherently low-power operation in communication era. The emphasis on low-power operation stems from portable devices, which strongly depend on the efficient use of their battery. In fact, Bluetooth allows making a trade-off between data transfer rate and power consumption. Nowadays, Bluetooth technology has emerged as a favorable platform for short-range wireless communication that operates in the 2.4-GHz ISM band. Its purpose is offering an alternative to connect portable devices without requiring drawing cable between them. A service discover protocol (SDP) is a dedicated protocol that can be used to exchange the information about all the services available at each node after starting the connection before connection node(s) can find their nearby nodes after inquiry procedure. Consumption of energy caused by wireless data transmission on a smartphone increases very fast with the growing capability of Internet application that requires network connectivity [4], as a result shrinks the battery life and is not able to keep up energy demand of application.

The rest of the paper describes starting from background of the proposed model in Sect. 2, followed by proposed model definition and description in Sect. 3. Section 4 analyzes the results and concludes the paper in Sect. 5.

2 Background

In MCS [1, 2], a group of people called crowd can collect and share information through their devices and those devices are enabled with advance sensing and computational capability. For proper functioning of MCS system, a large number of participants are required for the sensing, collection, and uploading of the data. MCS is a relatively new technique in which users of smartphones use the rich sensing capabilities of their devices to collect and share information. Crowdsensing simulation aims to simulate the crowd behavior, estimate the sensing patterns, and help researchers and also provide solution for the developer to choose what to sense as

well as to identify the minimum user threshold necessary for an application to collect sufficiently big data so that some useful information can be found. MCS has different application areas such as environment, infrastructure, health, citizen collaboration, traffic, and social networking.

To efficiently utilize the limited radio resources, the universal mobile telecommunication system (UMTS) and radio resource control (RRC)[1, 5] protocols introduced a state machine associated with each user equipment (UE)[2, 3]. The UMTS network comprises of three subsystems such as user equipment (UE), UMTS terrestrial radio access network (UTRAN), and core network (CN). The UTRAN permits connection between UE and CN. It comprises of two components: base station (Node-BS) and the radio network controllers (RNC). The RNC control multiple numbers of Node-BS. The features of UTRAN are radio resource control packet scheduling and control implemented at the RNC. The centralized CN is the master of the whole network or cellular network. The Gateway GPRS Support Node (GGSN) within the core network serves as a gateway and hides the UMTS internal infrastructure from the external network.

IDLE, CELL DCH, and CELL FACH are three states in RRC [6] state machine. IDLE of RRC state machine is in default state when UE is turned on. CELL DCH state allows a user equipment to utilize the radio resources for data transmission. User equipment can access high-speed packet access mode. When there are large numbers of users, some user equipments use low power though they are still on the CELL DCH state.

CELL FACH is established with no dedicated channel which is allocated to the user equipment. Hence, there is a connection, but without user equipment. The UE can transfer data through shared low-speed channel. It is developed for low-data throughput. It consumes less radio resource than CELL DCH[2, 3].

Recently, Bluetooth has emerged as a promising platform for short-range wireless networking. Energy aware system design and evaluation of network protocol for an omnipresent networking environment require practical knowledge of the energy consumption behavior for the wireless devices [5]. It may be desirable looking into the impact of energy efficient devices that directly affect the overall performance of the system. A Bluetooth node can be in active mode when it is actively participating in the channel, otherwise it can reduce the power consumption [7] by eventually using the low-power mode, that is the sleep mode. The Bluetooth controller operates in two major states, that is, either in the active mode or in the sleep mode. The sleep mode is the default low-power mode in Bluetooth; when the device is in sleep mode, the clock is running for a fixed period and in that period no transmission is possible. When one node needs transmission at the time of sleep mode, there is a fraction of time required for changing the sleep mode to active mode and that fraction of time is called delay time. A Bluetooth node in the system follows the inquiry and page procedures to transfer from sleep to active state [8]. When the device is in active mode, there is no delay for data transmission.

3 Proposed Model

Generally the collector is located in the cloud and the sensing as a service paradigm, the cloud notifies the sensor(s) and the mobile device(s) located in specified area of the cloud. They achieve the request and gather the information. The cloud collector selects which samples need to be gathered based on the sensing interest of it. It may assume to be known in advance. Requests may be asked from different agents or applications, and it may need samples from different locations and type of sensor(s). The cloud collector screens the observed region in a set of locations. In each location, the mobile device creates samples from a set of sensor(s). The spent [9] energy (E_s) of the mobile devices in contributing data is qualified to sensing $\left(E_s^c\right)$ and reporting operations $\left(E_s^r\right)$ [1, 2]:

$$E_s = E_s^c + E_s^r \tag{1}$$

The sensing contribution $\left(E_s^c\right)$ may be considered, only if the sensor S is not used by any other request or application. This is defined by utility context of S that is U_s. E_s^c is defined as below:

$$E_s^c = E_s^c . U_s \tag{2}$$

Here, E_s^c is considered for energy spent by the sensor S.
The utilization context (U_s) is defined as [1, 2]:

$$U_s = \begin{cases} 0 & \textit{if the sensor s is used by another application} \\ 1 & \textit{otherwise} \end{cases} \tag{3}$$

Delivering the information which is gathered from the group of sensors and then communicating the data to the cloud collector in a wireless mode. Reporting data is made at beginning of time-slot t for samples gathered during time-slot t-1. Energy cost of sensor S related to communication $\left(E_s^r\right)$ depends on the working technology, that is, LTE, WI-FI, or Bluetooth, which is defined as below:

$$E_S^r = \begin{cases} E^{L_3}, \textit{wifi disable, LTE disable and bluetooth connected} \\ E^{L_4}, \textit{wifi disable, LTE idle and bluetooth connected} \\ E^{L_5}, \textit{wifi disable, LTE idle and bluetooth idle} \end{cases} \tag{4}$$

When both Wi-Fi and LTE [10] interfaces are active, transmission takes place via Wi-Fi as it is more energy efficient. The user(s) do not consume their plan of the data as they pay to the cellular operator. The energy spent [9] during transmission time T_{tx} is defined as:

$$E^{L_3} = \int P_{tx}^B dt \tag{5}$$

Where P_{tx}^{B} is the power consumed for transmission of Bluetooth packets created at rate λ_g, while Wi-Fi-disabled and LTE-disabled but Bluetooth connected is defined as:

$$P_{tx}^{B} = P_{SL} + P_{tx}^{b} * T_{tx}^{b} + \gamma_{xg} * \lambda_g \tag{6}$$

The parameters P_{SL}, γ_{xg}, and P_{tx}^{b} represent energy in sleep mode, energy cost, and transmission power to elaborate a generated packet.

The *RRC* state machine [6] and its implementation in the proposed model are used to LTE power consumption. LTE [10] has *RRC* IDLE and *RRC* CONNECTED in *RRC* states. In the state of *RRC* CONNECTED, user equipment can be mentioned in one of the three modes in the system. Continuous reception, short *DRX*, and long DRX are three modes in the system. In the state of *RRC* IDLE, user equipment is only in *DRX* mode. Whenever the smartphone is in idle state and desires to communicate, it transits to the connected state from idle state. It goes into the tail state before returns back to the idle state after the transmission. In this way, the energy consumption for the smartphone during reporting is derived from [1] with two new parameters P_{tx}^{b} *and* T_{tx}^{b}, while Wi-Fi is disable, LTE idle and Bluetooth connected. This can be defined as:

$$E^{L_4} = P_p * T_p + P_{tx}^{L} * T_{tx}^{L} + P_{tx}^{L} * DRX_{IT} + P_{DRX} * RRC_{IT} + P_{tx}^{b} * T_{tx}^{b} \tag{7}$$

Where T_p and P_p are promotion delay and power, T_{tx}^{b} is time transmission, P_{tx}^{b} is power transmission, DRX_{tx} is discontinuous reception inactivity timer. P_{DRX} is the power consumed when the smartphones are in one of the two *DRX* modes and *RRC* is the *RRC* inactivity timer.

In Bluetooth connectivity, P_{tx}^{b} and T_{tx}^{b} are power and time transmission.

When the Wi-Fi is disabled, LTE is idle, and Bluetooth is idle, the energy consumption equation is as follows:

$$E^{L_5} = P_p * T_p + P_{tx}^{L} * T_{tx}^{L} + P_{tx}^{L} * DRX_{IT} + P_{DRX} * RRC_{IT} + P_{p}^{b} * T_{p}^{b} + P_{tx}^{b} * T_{tx}^{b} \tag{8}$$

Where T_p^b and P_p^b are promotion delay and power, when Bluetooth connectivity is in idle mode. P_{tx}^{b} and T_{tx}^{b} are power and time transmission.

4 Result and Discussion

This section illustrates the result of the proposed model.

Figure 1(from Eq. 5) reflects the result of energy spent for different power consumption values when Bluetooth is enabled and other two parameters such as LTE and Wi-Fi are disabled. Bluetooth is considered as in connected state. This figure places the power consumption and energy spent by Bluetooth to generate

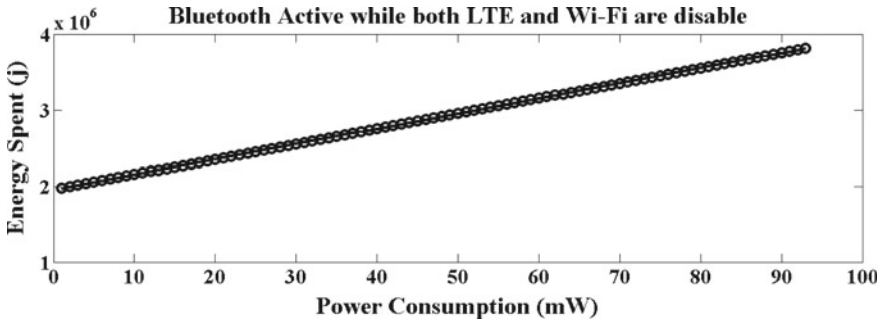


Fig. 1 Energy spent versus power consumption when both LTE and Wi-Fi are disabled

packets at the rate of λ_g during transmission time T_{tx}. The energy spent is increasing with respect to corresponding power consumption. It linearly increases the value of energy spent with respect to power consumption.

Figure 2(from Eq. 7) shows the energy spent with respect to power consumption while Bluetooth is connected, Wi-Fi is disabled, and LTE is in idle state. Bluetooth is considered only in the connected state. The figure between energy spent and power consumption discusses the energy spent by the smartphone during reporting.

Figure 3(from Eq. 8) shows the energy spent with respect to power consumption while Wi-Fi is disabled, LTE is in idle state, and Bluetooth is also in idle state. When the device is in idle state and it needs to communicate, then it transits to connected state. The transition from idle state to connected state requires some power and time that is called promotion delay.

Figure 4 shows the comparative study of energy spent with respect to power consumption when the Bluetooth of the device is in active and idle state. From the figure, it can be seen that when device is in connected state, it spends more energy and power. But when the device is in idle state, it spends less energy and consumes more power as compared to energy. By comparing the active and idle states in figure, it concludes that there is a promotion delay when the device transits from

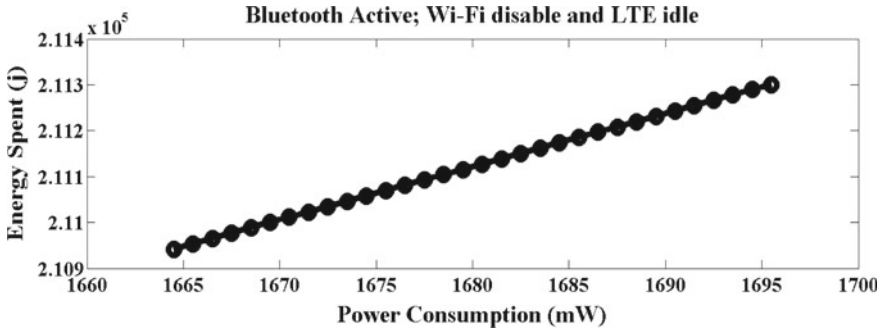


Fig. 2 Energy spent versus power consumption when Wi-Fi is disabled, LTE is in idle state, and Bluetooth is in active state

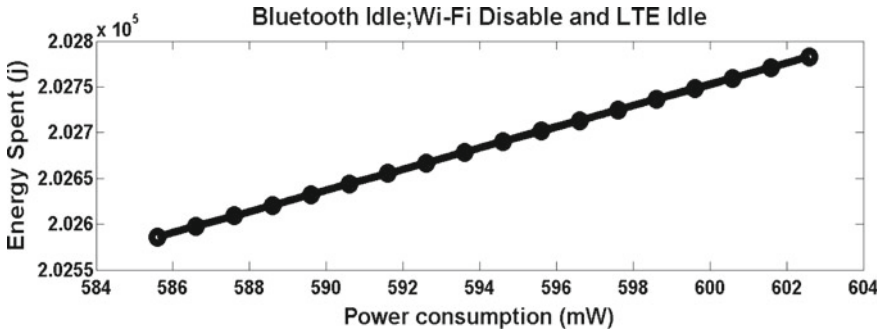


Fig. 3 Energy spent versus power consumption when Wi-Fi is disabled, LTE is in idle state, and Bluetooth is also in idle state

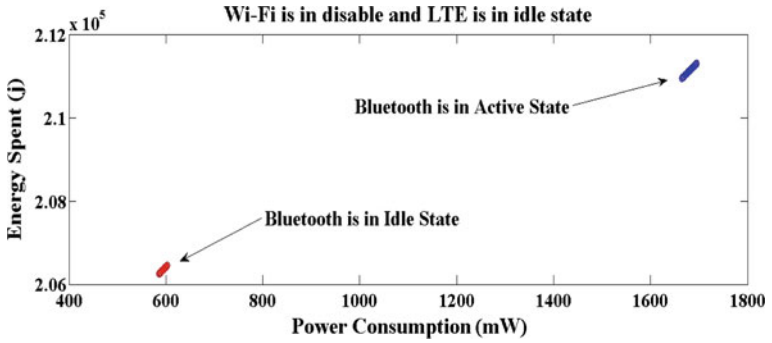


Fig. 4 Energy spent versus power consumption when Wi-Fi is disabled and LTE is in idle state

idle to connected state. Due to the promotion delay, starting point of the idle state figure varies from the connected state figure. When the device is in idle state, it minimizes the energy spent and power consumed. But when the device transits from idle to connected state, it requires some extra energy in terms of promotion delay.

5 Conclusion

In this paper, it has been proposed and analyzed the energy efficient model for WI-FI, LTE, and Bluetooth in opportunistic mobile Crowdsensing environment for data acquisition. It has been analyzed the total energy spent by Wi-Fi, LTE, and Bluetooth during the connected state and sleep mode. Also figure out the comparison between the energy spent of the active state and the idle state. By minimizing the energy spent during data collection from prominent user that can directly impact on the overall performance of the mobile Crowdsensing system.

References

1. Capponi, A., Fiandrino, C., Kliazovich, D., Bouvry, P., Giordano, S.: Energy efficient data collection in opportunistic mobile crowdsensing architectures for smart cities. In: 3rd IEEE INFOCOM Workshop on Smart Cites and Urban Computing- Atlanta (2017)
2. Capponi, A., Fiandrino, C., Kliazovich, D., Bouvry, P., Giordano, S.: A cost-effective distributed framework for data collection in cloud-based mobile crowd sensing architectures. In: IEEE Transactions on Sustainable Computing (2017)
3. Fiandrino, C., Kantarci, B., Anjomshoa, F., Kliazovich, D., Bouvry, P., Matthews, J.: Sociability-driven user recruitment in mobile crowdsensing internet of things platform. In: IEEE Global Communications Conference (GLOBECOM), Washington, DC. (2016). https://doi.org/10.1109/glocom.2016.7842272
4. Yu, X., Cui, Y., Savolainen, P., Siekkinen, M., Wang, A., Yang, L., YlaJaaski, A., Tarkoma, S.: Modeling energy consumption of data transmission over Wi-Fi. In: IEEE Transactions on Mobile Computing (2014)
5. Cano, J.-C., Cano, J.-M., Carlos Calafate, T., González, E., Manzoni, P.: Trade-off between power consumption and performance in bluetooth. Ad Hoc Sens. Wirel. Netw. **8,** 141–159. Old City Publishing, Inc. 2009
6. Foddis, G., Garroppo, R., Giordano, S., Procissi, G., Roma, S., Topazzi, S.: LTE traffic analysis for signalling load and energy consumption trade-off in mobile networks. In: IEEE International Conference on Communications (ICC), pp. 6005–6010 (2015)
7. Manzoni, P.: Power characterization of a bluetooth-based wireless node for ubiquitous computing. In: 2006 International Conference on Wireless and Mobile Communications (ICWMC06) (2006)
8. Manzoni, P.: Evaluation of the energetic impact of Bluetooth low-power modes for ubiquitous computing applications. In: Proceedings of the 3rd ACM International Workshop on Performance Evaluation of Wireless Ad Hoc Sensor and Ubiquitous Networks - PE-WASUN 06 PE-WASUN 06 (2006)
9. Qiang, X., Rong, Z.: When data acquisition meets data analytics: a distributed active learning framework for optimal budgeted mobile crowdsensing. In: IEEE International Conference on Computer Communications (INFOCOM) (2017)
10. Huang, J., Qian, F., Gerber, A., Mao, Z. M., Sen, S., Spatscheck, O.: A close examination of performance and power characteristics of 4G LTE networks. In: 10th ACM International Conference on Mobile Systems,Applications, and Services, ser. MobiSys, pp. 225–238 (2012)
11. Huang, J., Qian, F., Gerber, A.: A close examination of performance and power characteristics of 4G LTE Networks. In: 10th International Conference on Mobile Systems, Applications, and Services (MobiSys)-12, Low Wood Bay, 25–29 June 2012. https://doi.org/10.1145/2307636.2307658

Emphysema Medical Image Classification Using Fuzzy Decision Tree with Fuzzy Particle Swarm Optimization Clustering

Swathi Jamjala Narayanan, Rajkumar Soundrapandiyan, Boominathan Perumal and Cyril Joe Baby

Abstract Automated classification of medical images using machine learning methods has portrayed a vital role in the field of medical diagnosis. In the research work presented in this paper, the process of classifying the medical image is done in twofold, feature extraction and classification using fuzzy decision tree (FDT) with evolutionary clustering. The feature descriptors of the images are extracted using local diagonal extrema pattern (LDEP). The extracted features are passed to fuzzy particle swarm optimization (FPSO) clustering algorithm to obtain optimal fuzzy partition space for each attribute, which are then later used for inducing FDT. The proposed method of applying FPSO to develop input fuzzy space for Fuzzy ID3 is tested on emphysema CT image to classify the patient's lung tissue into normal, centribulor emphysema, and paraseptal emphysema. From the results obtained, we observe that our proposed framework improves the classification accuracy of Fuzzy ID3 compared to the other frameworks considered.

1 Introduction

The medical field has seen tremendous improvements in the field of computer-aided diagnosis. Several pattern recognition and prediction tools and algorithms are developed by researchers in order to extract the useful information from medical images in the form of feature vectors which can be further used by machine learning

S. J. Narayanan (✉) · R. Soundrapandiyan · B. Perumal
School of Computer Science and Engineering, VIT, Vellore 632014, India
e-mail: swathi.jns@gmail.com

R. Soundrapandiyan
e-mail: rajkuamrs@vit.ac.in

B. Perumal
e-mail: boomi051281@gmail.com

C. J. Baby
School of Electronics and Communications Engineering, VIT, Vellore 632014, India
e-mail: cyrilbabyjoe@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_31

algorithms for classification [1, 2, 3]. Generally, the set of feature vectors extracted from images influences the performance of the machine learning algorithms.

In the literature, several researchers have come up with strategies to extract efficient features for classification and recognition. The methods like energy histograms [4], shearlet-based energy histograms [5], and local contrast pattern [6] are some of the basic feature extraction methods. The local binary pattern (LBP) feature extraction method developed by Ojala et al. [7] is the most widely used approach, and it has given better results toward texture classification in medical applications. The other variants like local ternary pattern, center symmetric local binary pattern, center symmetric local ternary pattern, data-driven local binary pattern, and local neighboring intensity relationship pattern [8–12] are further proposed by researchers for feature extraction.

As most of these methods suffer due to curse of dimensionality, Dubey et al. [13] have come up with an approach named local diagonal extrema pattern where only two parameters namely the center pixel and its local diagonal neighbors are used for reducing the dimensionality. Once the features are extracted from images, classification algorithms like neural network and support vector machines can be used to classify the image into one of the categories which are predetermined. In this paper, we prefer to use fuzzy decision trees for classification. Choosing a right clustering algorithm has an impact over the partitioning of attributes toward development of fuzzy decision tree. Hence, we prefer to use fuzzy evolutionary clustering, in specific FPSO clustering combined with FDT, for image classification to deal with fuzziness and vagueness present in the extracted features.

The remaining of the paper is organized as follows: In Sect. 2, our proposed framework is presented. In Sect. 3, computational study and the results obtained are given followed by the concluding remarks and the possible extension work.

2 Proposed Methodology

The proposed framework is given in Fig. 1. The approach starts with the extraction of image features using local diagonal extrema pattern approach and then the extracted features are passed onto fuzzy particle swarm optimization (FPSO) clustering technique to obtain optimal fuzzy membership values for each of the features which act as key inputs for development of FDT using Fuzzy ID3. By passing on these obtained fuzzy membership values to the Fuzzy ID3 algorithm, a set of fuzzy rules are generated and product-product-sum reasoning mechanism is applied with max defuzzification method to obtain the output class label for a specified test image. The brief description of each process in the framework is given in next few subsections.

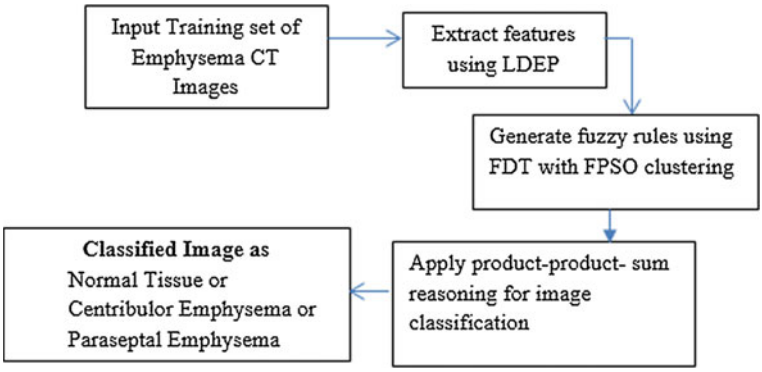


Fig. 1 Proposed framework for emphysema image classification

2.1 Emphysema Dataset

The dataset considered for the experiment is the emphysema CT image database [14] which has 165 data points. The images basically fall under 3 categories namely normal tissue, centribulor emphysema, and paraseptal emphysema. The feature descriptors of these images are extracted using local diagonal extrema pattern (LDEP) method. In this process, we extracted features of length 24 from the images.

2.2 Local Diagonal Extrema Pattern (LDEP) Feature Extraction

The complete process of the computation of $LDEP^{i,j}$ for center pixel $I^{i,j}$ is shown in Fig. 2 [13].

Let $Q_n^{i,j}$ denote the nth diagonal neighbor of any center pixel $Q^{i,j}$ which is at the distance D from a center pixel $Q^{i,j}$, considering $n \in [1, 4]$ and $Q^{i,j}$ is the pixel at ith row and jth column of some gray-scaled image M having m_1 rows and m_2 columns.

Let $I_n^{i,j}$ and $I^{i,j}$ denote the intensity values of $Q_n^{i,j}$ and $Q^{i,j}$, respectively, as given in Fig. 2a. The first-order diagonal derivative is obtained when α is set to 0, 1, and 2 correspondingly as given in Fig. 2b. Figure 2c displays the indexes of the local diagonal extremas, and it is defined as given in Eqs. (1), (2), and (3).

$$\tau_{\max} = \arg\max(sign(\alpha) = 0 \ \forall \alpha \in [0, 2]) \tag{1}$$

$$\tau_{\min} = \arg\min(sign(\alpha) = 0 \ \forall \alpha \in [0, 2]) \tag{2}$$

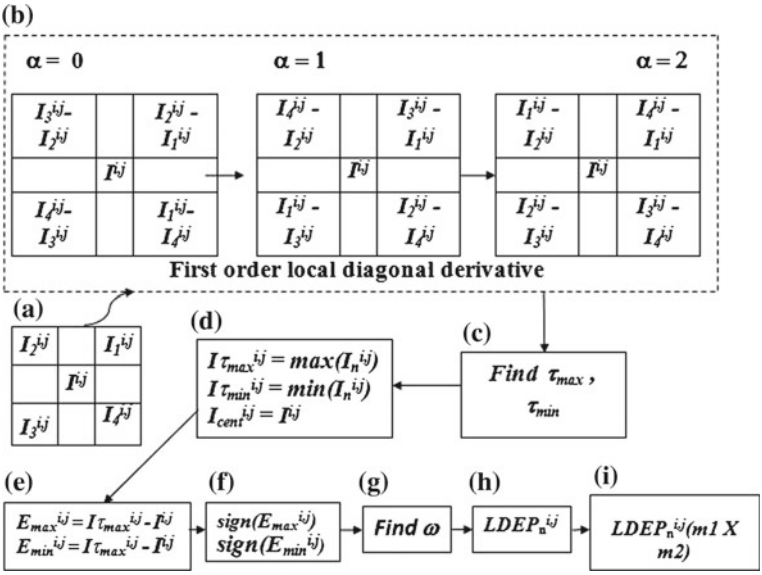


Fig. 2 Computational process of local diagonal extrema pattern [13]

Where,

$$sign(\alpha) = \begin{cases} 1, & \alpha \geq 0 \\ 0, & \alpha < 0 \end{cases} \tag{3}$$

The values obtained in Fig. 2d denote the local diagonal extremas and the center pixel value. Figure 2e represents $E_{max}^{i,j}$ and $E_{min}^{i,j}$ that are considered to be the local diagonal extrema-center difference factor for $I_{max}^{i,j}$ and $I_{min}^{i,j}$. Figure 2f shows the sign of $E_{max}^{i,j}$ and $E_{min}^{i,j}$. Figure 2g depicts the value of ω from the value of Fig. 2f. Figure 2h finds the value of $LDEP_n^{i,j}$ and is defined in Eq. (4).

$$LDEP_n^{i,j} = \begin{cases} 1, & if\ n = (\tau_{\max} + 8\omega)\ or\ n = (\tau_{\min} + 4 + 8\omega) \\ 0, & otherwise \end{cases} \tag{4}$$

Finally, the $LDEP_n^{i,j}$ is the pattern computed over an image of size m_1 X m_2 and is as given in Eq. (5).

$$LDEP_n^{i,j}(m_1 \times m_2) = (LDEP_1^{i,j}, LDEP_2^{i,j}, \cdots, LDEP_{\dim}^{i,j}) \tag{5}$$

2.3 Fuzzy Particle Swarm Optimization (FPSO)

Fuzzy particle swarm optimization proposed by Pang et al. [15] is a hybrid evolutionary optimization algorithm in which the position and the velocity of the particles are denoted as fuzzy relations between variables. The position of the particle represented as **X** in Eq. (6) denotes the relation between set of n data points and c cluster centers.

$$\mathrm{X} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{bmatrix} \tag{6}$$

μ_{ij} denotes the membership values of the ith data object in jth cluster. This is exactly same as the membership values given by FCM clustering approach. The velocity of the particle is denoted as a matrix of size $n \times c$ having the values in range between [–1,1]. The updating of position matrix and velocity matrix is performed using Eqs. (7) and (8).

$$\begin{aligned} &Velocity(t+1) = \\ &weight{*}Velocity(t) + (c1r1) \times (particlebest(t) - \mathrm{X}(t)) + \\ &(c2r2){*}(globalbest(t) - \mathrm{X}(t)) \end{aligned} \tag{7}$$

$$\mathrm{X}(t+1) = \mathrm{X}(t) \oplus Velocity(t+1) \tag{8}$$

The fitness function used for evaluating the solutions obtained is given in Eq. (9)

$$f(x) = \frac{K}{J_m} \tag{9}$$

where K is set to be a constant value and J_m is the optimization (objective) function of the FCM clustering algorithm. When J_m value becomes smaller, the fitness function will have higher value and the clustering effect seen would be better. After updating, the membership values are normalized to bring them back to a specific range.

Modified FPSO Algorithm for FDT

The pseudocode of the modified FPSO algorithm for generating fuzzy partition space for each attribute individually in order to induce FDT is given below.

Input: Set of feature Descriptors
Output: Optimized Fuzzy Partition Space

Algorithm:

Set the parameters
P=10, c1=2, c2=2, w=0.9, *max_Iterations*=1000
For each feature descriptor
Create a swarm (X, *particlebest, globalbest* and V are *nXc* matrices) of P particles
For each particle P
Initialize X, V, *particlebest*
End For
End For
While *max_Iterations* is not met **Do**
For each feature descriptor
Initialize *globalbest* for the swarm
For each particle P
Calculate the cluster centers as in FCM
Calculate the fitness value using Equation (9)
Find *particlebest* for the particle.
End For
Find *globalbest* for the swarm.
For each particle P
Update Equation (7) representing the velocity matrix
Update Equation (8) representing the position matrix
End For
End For
End While

The algorithm is terminated when the maximum iterations is reached or when there is no improvement in *globalbest* in a specified number of iterations. The obtained fuzzy membership values are then passed onto FDT induction process.

2.4 Fuzzy ID3 Induction Process

The inputs for the algorithm are the optimal fuzzy partition space obtained from FPSO, the variable leaf selection threshold βth, and the best node selection equation. The process of developing Fuzzy ID3 is as given in [16]: When there are candidate nodes to work on inducing FDT, choose one of the nodes as root node using fuzzy classification entropy search strategy. Then create its child nodes. The child node which ever meets the leaf threshold is termed as leaf node and the remaining child nodes are considered as a new set of candidate nodes, and the procedure of induction is recurrent until the stopping criteria is met.

Product-product-sum reasoning mechanism with maximum defuzzification function is used for obtaining a single class label for the test data [17].

3 Computational Experimental Results

The process of FDT induction is carried out as given in Sect. 2.4. The leaf selection threshold value is set as 0.75 for FDT, and FPSO algorithm is passed on with parameter values set as P = 10 c1 = 2, c2 = 2, and w = 0.9. During the process of FDT generation, the dataset passed to FDT is divided into tenfolds where nine out of tenfolds act as a training set and remaining onefold acts as test set. This process is repeated for 10 times so that each part acts as training as well as testing. Percentage of classification accuracy is calculated as $\frac{n_c}{n} \times 100$ %,where n denotes the total number of test patterns and n_c represents the number of test patterns classified correctly.

The results presented in Table 1. and Fig. 3 include FCM clustering results, FPSO clustering results, FDT with FCM memberships, and FDT with FCM-FPSO memberships. The results show that FDT-FCM-FPSO has improved the accuracy over FDT_FCM and other clustering techniques. The point to be noted here is FDT induction results are improved if we use evolutionary optimization techniques rather than just FCM technique for obtaining fuzzy membership values which act as basic input for FDT.

Table 1 Clustering and classification accuracy of emphysema CT images

Algorithm	Accuracy %
FCM clustering	53.333
FPSO clustering	66.666
FDT-FCM	78.000
FDT-FCM-FPSO	86.667

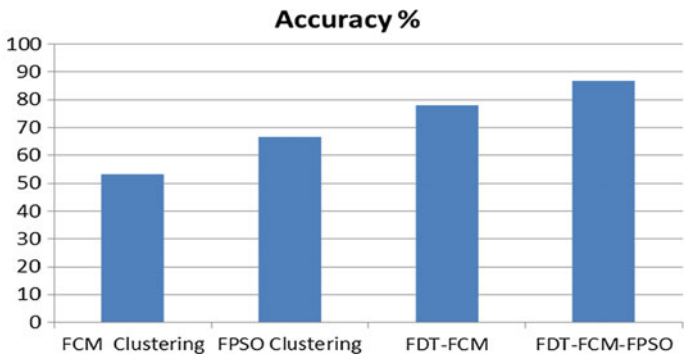


Fig. 3 Clustering and classification accuracy for emphysema CT images

4 Concluding Remarks

In this paper, we have come with a framework of utilizing FPSO clustering for partitioning the input space before inducing fuzzy decision tree using Fuzzy ID3 for medical image classification. Generally, FCM clustering technique is used for generating fuzzy membership values for inducing FDT. In this research, the key point observed is that the optimal fuzzy membership values obtained using the fuzzy evolutionary algorithm along with FCM play a major role in improving the accuracy of FDT. We tested our framework of FDT-FCM-FPSO on emphysema medical image classification and obtained good improvement in accuracy. In future, we prefer to apply few more evolutionary optimization techniques and observe their impact over the classification results of FDT. We also prefer to apply back propagation as a second optimization level after FDT induction process.

References

1. Rajkumar, S., Kavitha, S.: Redundancy discrete wavelet transform and contourlet transform for multimodality medical image fusion with quantitative analysis. In: 3rd International Conference on Emerging Trends in Engineering and Technology, pp. 134–139 (2010)
2. Tiwari, A., et al.: ANN-based classification of mammograms using nonlinear preprocessing. In: Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications, Springer, Singapore (2018)
3. Gautam, A., et al.: An improved mammogram classification approach using back propagation neural network. In: Data Engineering and Intelligent Computing, PP. 369–376, Springer, Singapore (2018)
4. Dong, Y., Ma, J.: Wavelet-based image texture classification using local energy histograms. IEEE Signal Process. Lett. **18**(4), 247–250 (2011)
5. He, J., Ji, H., Yang, X.: Rotation invariant texture descriptor using local shearlet-based energy histograms. IEEE Signal Process. Lett. **20**(9), 905–908 (2013)
6. Song, T., Li, H., Meng, F., Wu, Q., Luo, B., Zeng, B., Gabbouj, M.: Noise-robust texture description using local contrast patterns via global measures. IEEE Signal Process. Lett. **21** (1), 93–96 (2014)
7. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
8. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Process. **19**(6), 1635–1650 (2010)
9. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with local binary patterns. Pattern Recogn. **42**(3), 425–436 (2009)
10. Gupta, R., Patil, H., Mittal, A.: Robust order-based methods for feature description. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 334–341 (2010)
11. Ren, J., Jiang, X., Yuan, J., Wang, G.: Optimizing LBP structure for visual recognition using binary quadratic programming. IEEE Signal Process. Lett. **21**(11), 1346–1350 (2014)
12. Wang, K., Bichot, C.E., Zhu, C., Li, B.: Pixel to patch sampling structure and local neighboring intensity relationship patterns for texture classification. IEEE Signal Process. Lett. **20**(9), 853–856 (2013)

13. Dubey, S.R., Singh, S.K., Singh, R.K.: Local diagonal extrema pattern: a new and efficient feature descriptor for CT image retrieval. IEEE Signal Process. Lett. **22**(9), 1215–1219 (2015)
14. Sorensen, L., Shaker, S.B., De Bruijne, M.: Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE Trans. Med. Imaging **29**(2), 559–569 (2010)
15. Pang, W., Wang, K.P., Zhou, C.G., Dong, L.J.:(2004, September). Fuzzy discrete particle swarm optimization for solving traveling salesman problem. In: 4th International Conference on Computer and Information Technology, pp. 796–800 (2004)
16. Yuan, Y., Shaw, M.J.: Induction of fuzzy decision trees. Fuzzy Sets Syst. **69**(2), 125–139 (1995)
17. Narayanan, S.J., Bhatt, R.B., Paramasivam, I., Khalid, M., Tripathy, B.K.: Induction of fuzzy decision trees and its refinement using gradient projected-neuro-fuzzy decision tree. Int. J. Adv. Intell. Paradigms **6**(4), 346–369 (2014)

Design of LSTM-Based RNN for Prognosis Prediction of High-Risk Diseases from Patient Diagnostic Histories

K. Sathyabama, K. Saruladha and M. Hemalatha

Abstract In the biomedical field, prognosis prediction of high-risk diseases which are common among the adult patient remains a challenge. The risk factor and symptoms of the high-risk diseases like diabetes, blood pressure, and hyperlipidemia are inflated in adult patients and typically lead to medical complications like fatal death. Despite many studies, the prognosis prediction still remains a challenge. In this paper, a model is proposed for prognosis prediction of high-risk diseases by designing LSTM (long short-term memory)-based RNN (recurrent neural network) deep learning network. This work learns diagnostic histories of patients and predicts the incidence of high-risk diseases by using ICD-10 codes. The experimental results depict that the LSTM-RNN provides better performance with respect to precision, recall, and F1-score when compared to RNN-GRU. Further, the proposed model also investigates the effects of the learning parameters with respect to performance.

1 Introduction

An enormous of patient data are available, where deep learning techniques can be used to predict the prognosis of high-risk diseases. The high-risk diseases such as blood pressure, diabetes, and hyperlipidemia are predominantly increased in young-aged people [1]. Though it is a high-risk disease, people are not having much mechanism which will warn the patient at an earlier stage to aid in the diagnosis process. Lots of researchers have shown their interest to predict the prognosis of high-risk diseases [2]. High blood pressure stimulates the risk factor of diabetes. Many methods and models have been developed to predict these

K. Sathyabama (✉) · K. Saruladha · M. Hemalatha
Department of CSE, Pondicherry Engineering College, Puducherry, India
e-mail: sathii_manju@pec.edu

K. Saruladha
e-mail: charuladha@pec.edu

M. Hemalatha
e-mail: lathalavender95@pec.edu

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_32

high-risk diseases [1]. But it is an open issue to predict and classify high-risk diseases and non-high-risk diseases. The main aim of the proposed model is to predict the prognosis of high-risk diseases using RNN-LSTM. This model extracts the relationship sequence of high-risk diseases and explores unknown pattern of patient histories and treatments of medical data [3]. Deep learning techniques like LSTM-based RNN are mainly incorporated to classify and improve the prediction accuracy of the high-risk diseases.

In comparison with the existing model using multiple RNNs, the proposed model uses RNN-LSTM

- In the existing model using multiple RNNs, GRU is used as a recurrent layer and has only two gates, whereas in the proposed model, LSTM is used and has three gates.
- The existing model considers histories of patients only, whereas proposed model considers both histories of treatments and pathological measurements of patient data.
- LSTM easily learns long sequence of data than GRU.
- LSTM provides high expressiveness, better performance, and better prediction accuracy than GRU for patient data.

The remaining portion of this paper is structured as follows. Section 2 describes the literature survey of diseases and machine learning techniques. Section 3 focuses on the limitation of the existing system. Section 4 describes the objectives and features of the proposed model. Section 5 describes the proposed model and also describes evaluation metrics. Section 6 describes the experiment results and conclusion.

2 Literature Survey

Adult diseases are very common in day-to-day life for young-aged people. Since patients having disease are not taking care themselves, it leads to severe problems like death [4]. The health programs were organized on heart and cerebrovascular diseases and to built awareness of high-risk diseases such as cardiovascular and cerebrovascular diseases [5].

Prognosis prediction of cancer survival, cancer susceptibility, and cancer recurrence by using machine learning (ML) techniques were undertaken. ML methods classify the cancer patients into high-risk and low-risk groups [6]. The study of machine learning and data mining applications is predominantly used in the field of medical science to predict and diagnose diabetes mellitus (DM), complications of DM, and genetic background of DM [7]. The hypernetwork models have been used to classify the cardiovascular and breast cancer diseases using pathological measurements [8]. A large volume of data that are classified and categorized in e-commerce Web site using multiple recurrent neural networks is also observed [9].

Aspect-level sentiment classification models determine different parts of sentence with different aspects using attention-based LSTM [10]. The word2vec, dependency-based Word Embeddings, and factual embeddings are performed and the online user comments are classified using LSTM-based approach [11]. Drug name recognition and clinical concept extraction are evaluated and it improves accuracy and reduces time-consuming feature engineering using bidirectional LSTM and bidirectional LSTM-CRF [12].

Similarity between patient data sequence and temporal matching of patients with Parkinson's disease is evaluated using RNN architecture [13]. Clinical data comprising time series data, observations, and laboratory test results are evaluated and classified using LSTM-based RNN [14]. Prognosis prediction of high-risk diseases of adults and classifying high-risk diseases and non-high-risk diseases using simple recurrent neural network(SRNN) and multiple recurrent neural networks (MRNNs) are observed [1].

3 Limitations of the Existing System (RNN-GRU)

This section describes the limitations of the existing system [1]

- Existing system uses only histories of patients as a feature to predict the high-risk diseases.
- Only GRU (gated recurrent unit)-based RNN is used.
- Histories of treatments and pathological measurements are not considered.
- Performance accuracy of GRU is less when large sequence of data is processed.
- GRU-based RNN results in lesser classification accuracy when volume of data increases.

4 Objectives and Features of the Proposed Model

- To enhance performance of RNN to predict accurately the high-risk diseases.
- LSTM is used to process the sequence of input data of variable length sequences.
- Treatment histories and pathological measurements are considered as additional features.
- RNN-LSTM is used for effective classification of high-risk and non-high-risk diseases.
- Diagnostic histories of patients are time series data, and hence, LSTM could be used.

5 LSTM for Prognosis Prediction

Recurrent neural network (RNN-LSTM) is a type of neural network which is designed to operate and process the sequence of input data of flexible length of text sequences.

5.1 Why LSTM Is Used

The core feature of recurrent neural network is that it is capable of retaining the information from a previous state. However, in some cases, the RNN is not able to recall and learn from past or connect the past information (long-term dependencies) due to vanishing gradient problem [15]. In order to overcome the problem of long-term dependencies, LSTM (long short-term memory) and gated recurrent unit can be used [16]. Gated recurrent unit (GRU) consists of two gates with less computational cost compared to LSTM. Although GRU has less computational cost, when data volume increases, the performance and accuracy get reduced to a greater extent. LSTM is a special type of RNN which is able to retain the data for long period of time whereby it improves performance and accuracy of output [17].

5.2 LSTM-Based Prognosis Prediction

LSTM consists of three gates, namely input gate, output gate, and forget gate. These three gates are used to control the flow of data and protect and retain the value [17].

The proposed model LSTM-based RNN for prognosis prediction of high-risk diseases is shown in Fig. 1. It consists of three layers, namely

- LSTM layer
- Word Embedding layer
- Fully connected layer.

The raw data are given as input. The embedding layer converts ICD-10 sequence data into real-value vector and directly extracts semantic features of a given data. The ICD-10(International classification of diseases) is a common code which is used to classify the diseases of different types with unique code [18]. The LSTM layer learns long sequence of data from the embedding layer and overcomes the vanishing gradient problem [18]. The output vector of the LSTM layer is passed to the fully connected layer. The fully connected layer consists of real-value vector of ICD-10 code which helps to classify the disease into high-risk diseases and non-high-risk diseases via softmax layer [19]. The softmax function is expressed in Eq. (1)

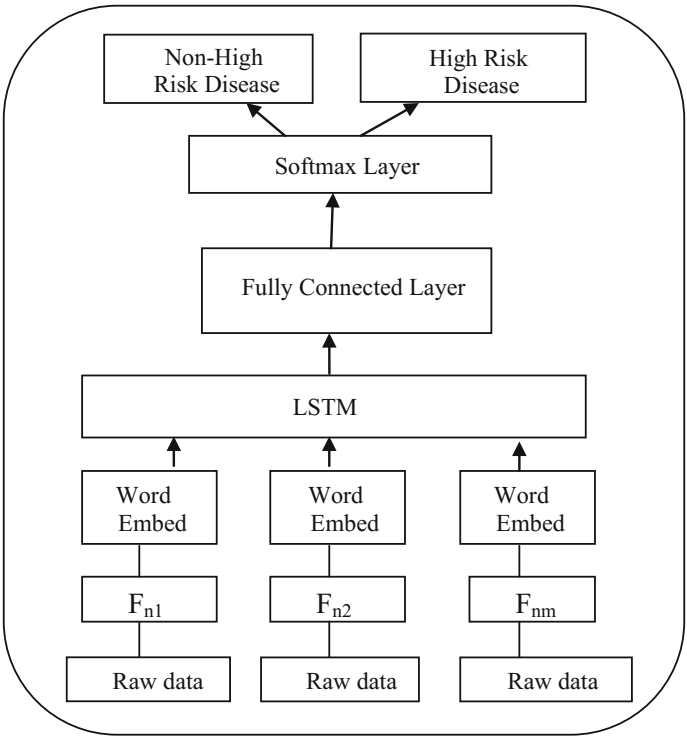


Fig. 1 Design of proposed model for LSTM-based RNN for prognosis prediction. (F_{nm})Feature of (mth) value and (nth) sequence

$$p(y_n|x^k) = \frac{\exp\{f(x^k, y_n)\}}{\sum_{y \in Y} \exp\{f(x^k, y)\}} \tag{1}$$

$$\widehat{y} = \arg max\{p(y_n|x^k)\} \ where \ y \in Y \tag{2}$$

Where y is the score function calculated by LSTM model to predict the disease from x^k of k length sequence given in Eq. (2). The softmax function is used to classify the prognosis prediction of high-risk diseases and non-high-risk diseases.

5.3 Evaluation Metrics

The prognosis prediction of high-risk diseases can be evaluated by using precision, recall, and f-measure. The metrics is evaluated on both training and test data set, and the performance and accuracy of the proposed model are measured.

6 Experimental Setup

RNN-LSTM for prognosis prediction of high-risk diseases is implemented using TensorFlow as a backend and Keras. Keras is the open-source neural network library framework, and the package is written in python. To evaluate the proposed model on the dataset, Pima Indian Diabetes (768 instances) is used along with ICD-10 code sequence data comprising of blood pressure, hyperlipidemia, cardiovascular, and cerebrovascular diseases.

6.1 Experimental Results and Result Analysis

6.1.1 Parameter Setup

The basic learning parameters were set as shown in Table 1, and the experiments were conducted. The dataset is split into 60% for training data and 40% for testing. The loss values of RNN-GRU and RNN-LSTM are shown in Fig. 2a. From the graphs, it is evident that RNN-LSTM outperforms RNN-GRU for the Pima Indian Dataset. The comparison of loss (training and testing data) of the proposed model with respect to the number of epochs is shown in Fig. 2a.

The prognosis prediction of high-risk diseases and non-high-risk diseases with respect to metrics such as precision, recall, f-measure, and accuracy of RNN-GRU and RNN-LSTM is shown in Table 2. As per Table 2, RNN-LSTM gives 11% increase in accuracy in comparison with RNN-GRU. RNN-LSTM provides best performance and prediction accuracy with respect to metrics such as recall and f-measure of high-risk diseases when compared to RNN-GRU. Thus, it indicates that learning capacity of LSTM is higher than GRU. RNN-LSTM also demonstrates

Table 1 Parameter setup

Parameter setup	RNN-GRU	RNN-LSTM
Loss	Binary_crossentropy	Binary_crossentropy
Optimizer	Adagrad	Adagrad
Metrics	Accuracy	Accuracy
Activation function	Sigmoid	Sigmoid
Hidden layer	5	5
Node size	500,300,200,100,50	500,300,200,100,50

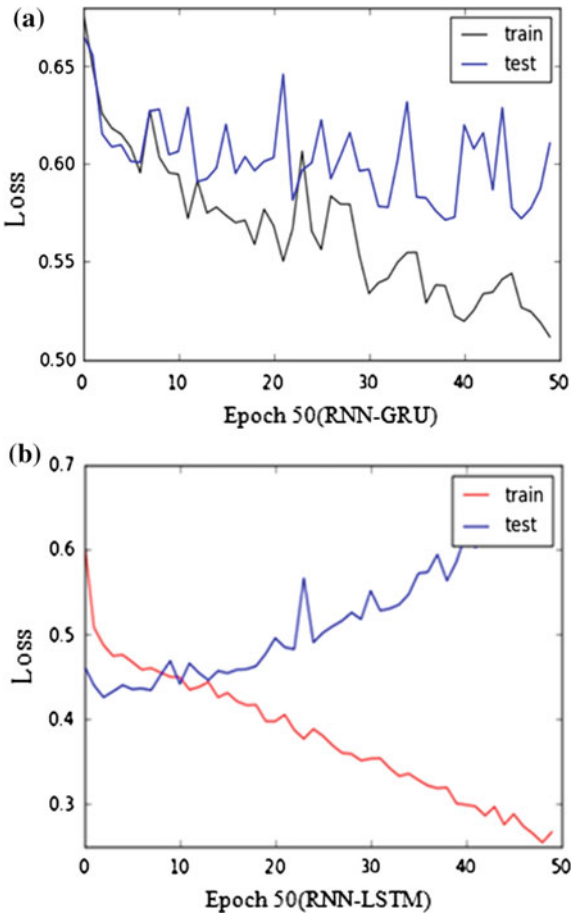


Fig. 2 Loss value of RNN-LSTM (**b**) and RNN-GRU (**a**) with number of epochs (50)

Table 2 Performance Evaluation

High-risk diseases				Non-high-risk diseases			Accuracy
Model	Precision	Recall	f-Measure	Precision	Recall	f-Measure	%
RNN-GRU	0.842	0.89	0.849	0.83	0.72	0.77	78.2
RNN-LSTM	0.84	0.92	0.875	0.84	0.721	0.78	89.23

more precision and f-measure value of non-high-risk diseases against RNN-GRU. Further, the results fortify that RNN-LSTM model learns the long time series data effectively and projects better classification accuracy (high-risk and non-high-risk diseases) than RNN-GRU.

Thus, RNN-LSTM overcomes the limitations of RNN-GRU and establishes better results on performance parameters (Fig. 3).

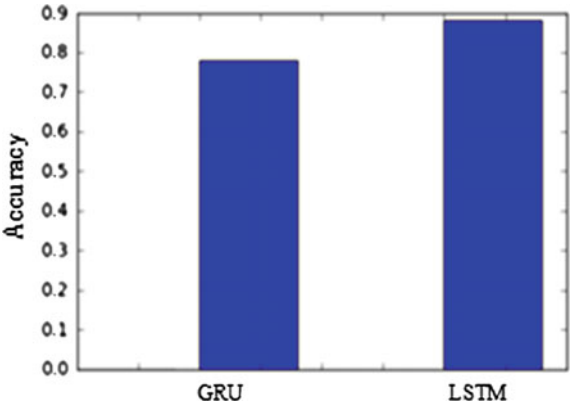


Fig. 3 Comparison of prediction accuracy of RNN-LSTM and RNN-GRU

7 Conclusion

In this paper, RNN-LSTM is used for prognosis prediction of high-risk and non-high-risk diseases. LSTM is used to overcome the limitations identified in the existing model using GRU. The proposed model uses ICD-10 code as sequence data from patient histories to classify the diseases. The proposed model provides better prediction and classification accuracy of 89.23% compared to RNN-GRU. In particular, proposed model provides better prediction with respect to metrics such as precision, recall, accuracy, and f-score. Therefore, LSTM is used to predict the high-risk and non-high-risk diseases accurately and also to improve overall efficiency and accuracy.

References

1. Ha, J.-W.: Predicting high-risk prognosis from diagnostic histories of adult disease patients via deep recurrent neural networks. In: IEEE Conferences Big Comp (2017)
2. Karthikeyan, T., Kanimozh, V.A.: Deep learning approach for prediction of heart disease using data mining classification algorithm deep belief network. Int. J. Adv. Res. Sci. Eng. Technol. **4**(1) (2017)
3. Thiyagarajan, C.: A survey on diabetes mellitus prediction using machine learning techniques. international J. Appl. Eng. Res. ISSN 0973–4562 **11**(3), 1810–1814 (2016)
4. Langley-Evans, S.C.: Nutrition in early life and the programming of adult disease: a review. J. Hum. Nutr. Diet. 1–14, (2015)
5. Go, A., et al.: Heart disease and stroke statistics-2014 update. Circulation **129**(3) (2014)
6. Kourou, K. et al.: Machine learning applications in cancer prognosis and prediction. Comput. Struct. Biotechnol. J. (2014)
7. Iohannis K.: Machine learning and data mining methods in diabetes research. Comput. Struct. Biotechnol. J. (2016)

8. Ha, J.-W., Eom, J.-H., Kim, S. C., Zhang. B.-T.: Evolutionary hypernetwork models for aptamer-based cardiovascular disease diagnosis. In: Proceedings. of the 9th Annual Conference Companion on Genetic and Evolutionary Computation, pp. 2709–2716 (2007)
9. Ha, J.-W.: Large-scale item categorization in e-Commerce using multiple recurrent neural networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 107–115 (2016)
10. Wang, Y.: Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 606–615 (2016)
11. Guggilla. C.: CNN- and LSTM-based claim classification in online user comments, In: Proceedings of COLING 2016. The 26th International Conference on Computational Linguistics, pp 2740–2751 (2016)
12. Unanue, I.J.: Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition (2017)
13. Che, C.: An RNN architecture with dynamic temporal matching for personalized predictions of Parkinson's disease (2017)
14. Lipton, Z.C.: Learning to diagnose with LSTM recurrent neural networks. published as a conference paper at ICLR (2016)
15. https://Medium.Com/@Camrongodbout/Recurrent-Neural-Networks-For-Beginners-Aca4e933b82
16. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Gated feedback recurrent neural networks. In: Proceedings. of the 32nd International Conference on Machine Learning (ICML-15), pp. 2067–2075 (2015)
17. Hochreiter, S., Sepp., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
18. International Classification of Diseases 10th Revision. World Health Organization. 2010. Retrieved February 26, (2010)
19. Veena Vijayan, V.: Prediction and diagnosis of diabetes mellitus–a machine learning approach. In: IEEE Recent Advances in Intelligent Computational Systems (RAICS)
20. Han, J.Y, et al.: Stroke or coronary artery disease prediction from mean platelet volume in patients with type 2 diabetes mellitus. Platelets. **24**(5), 401–406 (2013)

Websocket-Evented Real-Time Online Coding Collaborator

Varun Bhatia, Satej Joshi and Radhika Chapaneri

Abstract This paper outlines a logical prototype for a websocket-dependent, collaborative, online coding platform. It aims to improve seamless code updation, sharing, and collaborating for small to medium development projects—the lack of which leads to massive coding redundancy due to multiple members working on the same project. The paper lists out the foundational features, along with the technologies that can be utilized to realize the objective of creating a platform that allows easier team-based coding. The final proposed platform will allow developers to create and share compiled code across project members without the need of an execution environment, URL sharing or version control, and will implement code-focused ideas like real-time collaboration, multiple code support, online compilation, and administrator room system. It proposes the use of newer, emerging technologies, and stacks like websockets and real-time databases, while also explaining the reason behind the use of these technologies.

1 Introduction

This logical prototype is for an online, code-focused scratchpad using asynchronous frameworks, websockets, and event-driven databases that will allow teams of developers to share common code for testing, storage, and staging purposes. The features stated above will be of huge help to both beginners and professionals alike. Websockets represent an exciting new protocol in Web-based services, as an upgrade that works with HTTP [1]. Using websockets correctly can lead to higher

V. Bhatia (✉) · S. Joshi · R. Chapaneri
Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai, India
e-mail: varunb95@gmail.com

S. Joshi
e-mail: satejjoshi.nmims@gmail.com

R. Chapaneri
e-mail: radhika.chapaneri@nmims.edu

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_33

information sharing across connection endpoints, consequently leading to greater productivity for both the client and the server. The purpose of selecting websockets is to enhance DIRT (data-intensive real-time) system information handling and retrieval [2]. In the context of the proposed platform, websockets are instrumental for instantaneous updating of code and transfer of code values across members (users) participating in that project (common code module). The entire platform works akin to a common client server model; however, by changing the communication protocol from traditional HTTP/HTTPS to websockets (WS), and using specialized event-driven databases, we are able to add newer, stable features that are essential for improving code-based projects' productivity and speed [3, 4].

Section 2 outlines the platform in comparison with other similar online applications. Section 3 explains the reasoning and benefits of using a newer technology stack. Section 4 is the user interface flow, combining all the design elements. Section 5 enlists the methodology of using websockets and RethinkDB for real-time, fast-access use.

2 Platform Review

The proposed platform will be an online coding environment that will use websockets and live databases to introduce faster real-time code editing, better collaboration models in terms of **rooms**, and a code-focused feature set for developers to use. Modelled on the same platform as this [5], but faster, reliable, and more efficient—due to the new technology stack.

While there exist online coding grounds, like JSFiddle [6], they suffer from recurring problems: They are extremely limited in the languages they support; execution of code for HTML, CSS, and JS is done on an i-frame, with little to no server-side support for other languages; collaborative socket-based event systems are not baked in as a primary feature, and they have little to no concept of team-based online coding

3 Underlying Technology Stack and Proposed Features

This section outlines the important aspects that constitute the platform. The features and concepts explained here form the basis for the entire platform architecture. The following technologies are the most essential—each allows the realization of the proposed features that follow it.

As shown in Table 1, unlike GitHub [7], the proposed application is not a version control system. Rather, it is more like Google Docs [8], but remade with faster, newer technology and with a code editor rather than a text editor. Hence, the user works on the code directly, rather than files or repositories (unlike *GitHub).*

Table 1 Comparison of multiple existing platforms in terms of features and technologies offered

Platform	CF	CO	WS	RTC	MLS	EBD
GitHub Gist [7]	✓	✓	✖	✖	✓	–
JSFiddle [6]	✓	✓	✖	✖	✓	–
Codepen.io [11]	✓	✓	✖	✖	✓	–
Google Docs [8]	✖	✓	✖	✓	✖	–
Our Platform	✓	✓	✓	✓	✓	✓

CF: code focused; **CO:** collaborative; **WS:** use of websockets; **RTC:** real-time collaboration; **MLS:** multiple languages support; **EBD:** use of event-based database

3.1 Websockets

Websockets is an emerging communication protocol that re-realizes the base connection between client and server. Unlike the traditional HTTP protocol, websockets allow for a perennial and bidirectional (duplex) connection between the client and the server, allowing for continuous interaction where even the server can send data at any time, instead of waiting for the client to request the sever for information first [9]. This implementation can dramatically reduce load and refresh times, while allowing a continuous stream of information to flow between client and server.

Sockets also allow rapid scalability, especially for one server-many client model. By implementing sockets to act between the client and server, we can pass important information to our server for processing at any time, and this information can be sent from there to any other client, with virtually zero lag.

As shown in Fig. 1, there is no restriction on who can send data to whom, at what time. As long as the connection exists, it is a tunnel through which data can flow in both directions. This means that the server need not wait for the client to request data, and the client need not invoke the server repeatedly to send data. Also, both can *send and receive* data—there is no concept of one-sided data transfer.

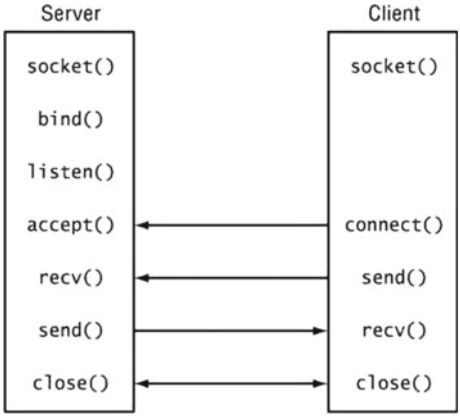


Fig. 1 *Simplified* - One server interacting with a client using the websocket protocol

3.2 Real-Time Database

For this application, RethinkDB [3] has been chosen for the live operations required for collaboration. RethinkDB is an event-driven database which has been optimized for continuous input and processing of small sizes of data. This database emits events whenever a DML (data manipulation language) operation takes place on the data and returns the result in the form of the location of the changed data and the new data. Using these events instead of polling the database whenever changes need to be detected in the database saves a lot of processing time and resources. This type of live database is extremely essential while using the websockets protocol, since it **eliminates the latency** caused by database operations, thereby upholding the real-time feature of the platform. The only drawback of using such a live database is that it creates a lot of metadata due to the storage of various versions of the same data in the memory. For example, for 'X' last made changes to the code in a room, X copies of the entire code will be stored in the live database when the room is active. Thus, the real-time database is not CPU intensive but RAM intensive.

The following are the proposed features of the application that will supplement the **real-time, collaborative, and code-focused** objective of the platform:

(1) **Rooms**: A developer may work on multiple projects, or multiple modules of a same project, with or without additional team members. A collection of members working on a common project, sharing a common file with the common code updated for all in real time, constitutes a room. A member may be a part of multiple rooms. It represents an abstract entity that exists to allow for better scalability, accessibility, and demarcation of code for the user. Create a more streamlined approach to collaboration. This is illustrated in Fig. 2, and the client–server interaction is illustrated in Fig. 3.

(2) **Real-Time Code Updating**: The main feature of this platform is to allow near seamless, low latency code sharing and editing. This means that users working on a common project will be able to see and track the changes made by other users *as the changes take place,* irrespective of time, location, or network of the team members. This is enabled by the use of websockets over HTTP, creating a network of perennial tunnels from the clients to the server. Using the WS protocol using (1), (2), and (3), the application can be coded to reflect changes made on shared code at any time to be updated to all concerned users instantaneously. This allows teams to track and review without having to download code again and again.

(3) **Event-Driven Database**: Most server applications waste time and resources in database lag itself. For a DIRT application, this lag can become unacceptable since real-time updates need to be performed, but the database operations also cannot be trivialized since the data ultimately must exist in the database. To overcome this lag, the proposed application uses event-driven databases over traditional databases for the real-time aspect. The database uses a schema described in (4), (5), and (6).

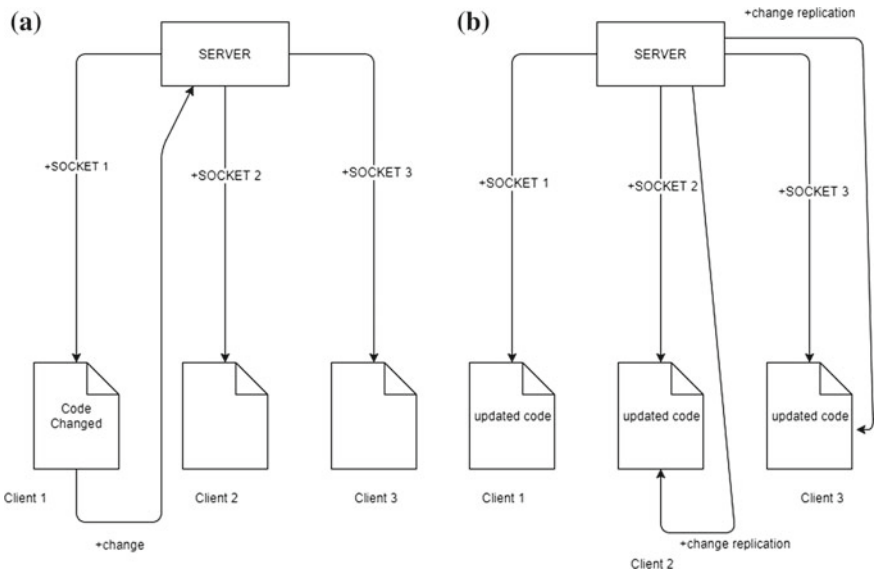


Fig. 2 Initial setup with three connected clients (in one room), and **a**, when client 1 edits the code that is being shared by all the clients. **b** Server reacting to the change be sending the new data to all the connected clients for that room over the websocket protocol

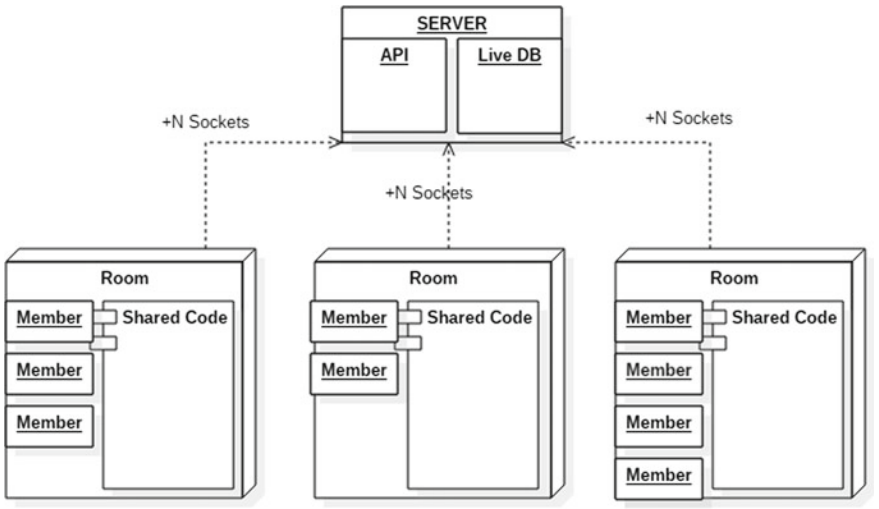


Fig. 3 Abstracted platform architecture with *rooms* and *one main server*. Every room has a separate common document. Changes in one room are not visible to other rooms but are visible to all the members of the concerned room only

(4) **Minified versioning and Code Compiling**: Users can commit code to the server and can then roll back to previously saved versions, which will be served directly from the browser. For code compiling, the entire code will be sent to the server, where it will issue a separate thread for the code to execute by invoking the compiler required. The results will then be sent to the client. Code compilation will **not** be collaborative—each user will have to separately execute the command. Also, the server will protect itself against potentially malicious code compilations.

4 User Interaction Flow

This section is an abstraction of the frontend and backend process flow when the user is interacting with the platform:

1. User registers and signs into the platform. The server is responsible for creating a permanent record of the same on the passive and primary MongoDB database.
2. The user is then directed to the **dashboard**. Here, the server sends data to the frontend to populate the dashboard.
3. The user can then select to join/create a **room—**accordingly, the server stores this data on the passive database.
4. Once the user enters a room, the server creates a **socket connection between the client IP and the server endpoint.** This socket remains in existence till the user leaves the room. The user can invite other users to the room, and a separate socket connection will be made for each. Using these interconnected sockets, each room member will see edits in **real time**. Every edit will be stored in **RethinkDB**, which then **pushes the event edit to all the socket connections**.
5. Additionally, the user can decide to compile the code on the shared document. The code is sent to the server, where a **separate thread is issued** to compile the code in the concerned language. The result is then sent to the client. **Compilation is a local process and individual for every user.**

Technologies Used:

1. Backend: Node.js running the server, RethinkDB for the real-time database, MongoDB for the passive database, and socket.io for creating socketed connections. Compilers for various languages will be loaded (like gcc for C ++).
2. Frontend: Materialize CSS for styling, jQuery, and HTML. Bower components supply the socket.io library for the frontend connection and CodeMirror for the code text editor.

5 Project Implementation

This section details the implementation algorithms for the two major aspects of the project—websocket and event-based database.

5.1 Websocket Information Control

The websocket is initialized on the server side for every client which moves from the dashboard page to the room page. At this instant, the URL of the page gets converted to a GET requests which accepts the room number and user ID as arguments.

Server side: The socket is initialized with the following information:

$$S_1 = \{R_1, V_1, C_1, P_1\} \tag{1}$$

$$S_2 = \{R_2, V_2, C_2, P_2\} \tag{2}$$

$$S_N = \{R_N, V_N, C_N, P_N\} \tag{3}$$

In Eqs. 1, 2, and 3, S represents the initialized socket at the time of a client connection, R represents the room number with which the client has connected, C is the control information in the form of previously committed code, and P is the preferences associated with the user (like choice of language and administrator rights).

This is then emitted to all the active sockets that have been currently initialized. The client that matches the room number accepts the socket information, whereas all the other clients reject the socket. This ensures that at least and at most one client will listen and accept an emitted socket.

Client side: The client listens for and accepts a socket. A socket object is emitted from the client to the server when the client changes the value in the shared document (code). A socket object is accepted when the server sends changed values to the client (code has been changed by some other member).

In both cases, the syntax [10] of the socket object remains the same, as referenced in (1) *However, the value of V in the object will differ, and so will the syntax for the socket object*

```
socket.emit('document-update', msg ;
```

The client emitting a socket instance on document update, with 'msg' being the data object passed to the server

```
socket.on('doc',  function(msg){if(msg.new_val.id === room
&& msg.new_val.user ! = user) {
```

Client waits for a socket object when the document has been updated from some other user. The second line is for the client to decide if the socket instance must be accepted or rejected. These methods are automatically invoked whenever a user is on the editing page for the concerned room. This allows true team-based collaboration irrespective of location and environment.

5.2 Database Management

The key problem arises when multiple concurrent users are present. This will inevitably cause a load on the database, because the implementation requires constant data changes for the value of the document in a particular room, either insertion or deletion. This requires long polling an active database which holds

- the room number/room id
- the value of the document in the room
- username of the last editor

In set format, the long-polled database should be represented as

$$D_1 = \{R_1, V_1, E_1\} \tag{4}$$

$$D_2 = \{R_2, V_2, E_2\} \tag{5}$$

$$D_N = \{R_N, V_N, E_N\} \tag{6}$$

In Eqs. 4, 5, and 6, D is the database object, R is the room number, V is the value of the shared document, and E is the username of the last editor of the document.

For such object-based documents, a NoSQL database is more flexible. Two databases are used:

- MongoDB (*for permanent data storage*)
- RethinkDB (*for real-time code storage*). Refer Table 2 for the schema for this database.

RethinkDB allows developers to minimize long-polling by adding a feature where the database itself emits an event whenever an edit is made to the database document, returning a cursor to the edit. By using RethinkDB and not MongoDB as the main active long-polled database, the long-polling latency is significantly reduced [3].

Table 2 RethinkDB active long polling. *Value* is the code in the common document

ID	Value	User	Room no
147502	console.log("x");	A	1
147513	def method:	B	2
148716	int x	C	3
150152	float y	D	3

5.3 Code Compilation, Commits, and Rollback

Every user will have the option to compile their code and run it on the browser itself. The document contents will be sent to the server, where the server will start a separate thread to run the code on its compiler. The results will then be sent to the client.

Commits and rollback are minified version controlling systems. If room members feel that a code is stable, they commit it, which saves the code on the passive database (not the real-time database). Rollback reverses this process—allowing users to load committed code from the passive database onto the real-time database and replacing the document contents.

6 Conclusion

This code collaborator is being modelled to empower small-scale teams and projects that value time, efficiency, and choice over heavy noncode-focused features. As the number of developers and teams grow, it is no longer convenient or desirable to code at a single location. Therefore, collaboration has become one of the main features to test the reliability of the development cycle for any software. Good collaboration allows for faster turnover and less error. While some platforms have made this possible, speed, especially for smaller projects, is hindered because the number of extra features overshadows the essentials. Therefore, this platform aims to bridge that gap by user newer and faster technologies, while sticking to developer friendly features. It focuses on three aspects: real time, collaborative, and Code-First. The idea is to bridge distances and time constraints to build a lightweight, modern editing environment that can be used (primarily) for small-scale team projects to code in, all in real time, using the power of socketed connections. Real time brings a new dimension to collaborative projects. It allows developers to reflect the changes in their code instantaneously across their team members.

For example, a student wanting to practice Python can open or join a room where hundreds of members post code on the shared document and ask their doubts and queries. If a small-scale team wants to create a markup of their project which is available to all project members, then project head can create a room to store this markup, and all members can be asked to join that room. This will ensure that the core markup of the project to be coded can remain shared and updated in real time, and members can comment and upload the progress they have made on the specific modules of the markup. Other applications include interview testing—an alternative to whiteboard coding where the interviewer and interviewee can be in different locations and the test can be invigilated easily and efficiently.

References

1. Puranik, D.G., Feiock, D.C., Hill, J.H.: Real-time monitoring using ajax and websockets. In: 20th IEEE International Conference and Workshops on the Engineering of Computer Based Systems (ECBS) (2013)
2. Wessels, A., Purvis, M., Jackson, J., Rahman, S.: Remote data visualization through websockets. In: Eighth International Conference on Information Technology: New Generations (ITNG) (2011)
3. Walsh, L., Akhmechet, V., Glukhovsky, M.: Rethinkdb-rethinking database storage (2009)
4. Kopka, H., Daly, P. W.: Responsive Real-Time Collaborator with Audio Channel, Addison-Wesley, Harlow, England (2015)
5. L. a. M. Goldman, Real-time collaborative coding in a web IDE. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (2011)
6. JSFiddle About Page, JSFiddle, https://www.jsfiddle.net/ (2017). Accessed 30 July 2017
7. "GitHub About Page," Github,Inc, 29 July 2011. https://www.github.com. Accessed 2017
8. "Google Docs About," Google, https://www.google.co.in/docs/about. Accessed August 2017
9. S. M. J. M. P. G. P. R. Rohit Arora, Efficient E-learning management system through web socket. In 3rd International Conference on Computing for Sustainable Global Development (INDIACom) (2016)
10. D. Anderson/Modulo.io, How Node.js can accelerate enterprise application development, Modulo.io/Progress (2014)
11. Creating a Pen, Codepen.io, https://www.codepen.io/hello. Accessed July 2017

Analytical Survey on Parameters for Designing an Efficient 5G Antenna System

M. Benisha, R. Thandaiah Prabu and V. Thulasi Bai

Abstract Nowadays, the recently using 3rd generation (3G) and 4th generation (4G) cellular frequency bands which are below 6 GHz have been overcrowded due to the increased number of mobile users. This requires a new generation so-called 5th generation (5G) with the expectations, and it is pushed to work on greater than 28 GHz frequency band. Hence, the vision of communication researchers is designing mobile antennas for this new generation system. Many system and antenna parameters are to be considered during the design of antennas to get better reliable mobile system. In addition, we should consider about the memory and system complexity during simulation. In this paper, we are presented an analytical survey on the various antenna and system parameters to be considered to get a more efficient 5G mm-wave wireless system.

1 Introduction

The cellular technology (1st generation) was introduced with only voice applications by using analog transmission channels. After the evolution of digital technology, 2G systems with more applications (messaging, circuit switched data access) were introduced. To overcome the insufficiency of mobile data at 2G systems, 3G was evolved which provides more capacity voice calls and fast data services with multimedia transmission capabilities. Nowadays, 4G–LTE is

M. Benisha · R. Thandaiah Prabu
Anna University, Chennai, India
e-mail: benishaxavier@gmail.com

M. Benisha · R. Thandaiah Prabu (✉)
Jeppiaar Institute of Technology, Sriperambuthur, India
e-mail: thandaiah@gmail.com

V. Thulasi Bai
KCG College of Technology, Chennai, India
e-mail: thulasi_bai@yahoo.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_34

becoming reality with very high data speed and capacity. All the above cellular technologies are operating with the frequencies below 6 GHz.

Today's Internet world has massive number of Internet users with an emergence of various popular smart phones. This causes the available spectrum to be congested. In future, the number of mobile Internet users will increase beyond the limit for sure. Thus, we are in a situation to move onto the next-generation 5G with the basic two requirements of "zero latency" and "gigabyte experience" [1]. Zero latency provides not only a no-latent case, but also allows the machine-level communication without human interruption also called "Internet of Things", and a very fast human interaction. The transmission and reception will be done in gigabits per second is noted as "gigabyte experience", and this will satisfy the user data rate and network capacity as demanded.

To achieve these basic requirements of 5G, we should move onto the mm-waveband of 30–300 GHz such that the carrier frequency moved from the saturated 750–2600 MHz [2]. Some of the possible center frequency bands which support 5G within the mm-waveband are 24.65–27, 27–29.5, and 40–50 GHz under the sub-band of 24–50 GHz [3]. So, the designed 5G system may be a narrow band with the desired operating band listed above, or a wideband system which can operate in more desired bands.

The current 4G–LTE systems are having some basic restrictions to improve the performance such as limit in bandwidth expansion; that is, there is performance degradation due to carrier aggregation, and limit in using more number of antennas. But in 5G technology, the performance can be increased by some disruptive technologies like use of high frequencies (>6 GHz), updated user-centric virtual-cell formation, use of filter bank multicarrier modulation schemes, enhancing a real-spectral efficiency, use of MIMO technique, and employing interference alignment [4].

The most important parameters tell the system performance of the 5G systems are carrier frequency, bandwidth, modulation scheme, duplexing method, number of antenna's, type of antenna, size of the antenna array, half power beam width (HPBW) of antenna, antenna polarization, channel coding, achievable data rate, transmission range, channel estimation and measurement, and cost. This contribution presents an analytical survey on the above parameters and explains the ranges where we can get better performance to provide a highly reliable wireless mobile technology.

2 Antenna Parameters

While designing antenna, various parameters should be considered such as operating frequency, bandwidth and antenna size, polarization, manufacturing costs, and other probable interference effects with the handheld device; then only the designed antenna will be suitable for the handset. Antenna efficiency is a crucial thing we need to consider while antenna designing, and targeted gain for mobile device is

12 dB and base station is 25 dB, in case of 5G systems. But gain cannot come at the expense of coverage, instead of omnidirectional antenna, directional antenna with beam steering capability such as phased-array antenna could be better. Because the radiation pattern is required to cover entire azimuth and major elevation ranges. Multipath signal transmission may cause destructive signal overlay and leads to Rayleigh fading. Thus, antenna diversity-multiple-input multiple-output (MIMO) techniques can be employed to improve the diversity gain. The thickness of the antenna for satellite personal communication network is to be of the order of 10–12 cm and width to be smaller than 3 cm [5], which says that the size of the antenna should be very less, so that the installation would be easier.

A circularly polarized patch antenna is proposed in [5], which is found in the application for S-band systems. Wider the antenna, that is, when the antenna beam width increases, it reduces the size and cost of the antenna. Initially, the circular patch is designed with diameter 24 mm. By folding the patch, the operating frequency is confined to 4.5 GHz. Again, by loading 4–8 slots, it is confined to 3.85 GHz. With more number of slots, the operating frequency is exactly confirmed to the desired one that is 3.7 GHz. Under this circular patch, high dielectric substrate can be used to reduce the size and narrow the bandwidth. Thus, miniaturization is done. Beamwidth enhancement is done in [5] by using the concept that thickness of the surrounded dielectric has the influence on the height of the metallic block. Finally, the proposed antenna came with the half power beam width (HPBW) of 1240, impedance bandwidth of 10%, operating frequency range of 3.715–3.83 GHz, and center frequency of 3.77 GHz. The system provides an axial ratio of 1.15 dB and gain of 5 dBi.

Recently, a new low-profile patch antenna has been substituted for the microstrip patch antenna, which is called dense dielectric (DD) patch where the metal slab in microstrip patch is replaced by the high-permittivity thin dielectric slab [6]. It provides the same characteristics as the microstrip patch antenna at lower microwave frequencies and better performance at high frequencies like high radiation efficiency and low metallic loss, and is more suitable for array designs. Hence, this can be a better antenna for 5G systems in recent days.

A dense dielectric patch antenna using electronic band gap (EBG) ground structure and dielectric substrate is proposed in [7]. This system is operating with frequency of 28 GHz. It provides the antenna impedance bandwidth of 27–32 GHz and increased gain of 16 dBi. It uses a simple feed structure and follows broadside radiation pattern. The HPBW of this antenna is 110 in the horizontal direction and 34.40 in the vertical direction. It offers a higher operating frequency than the metallic patch antenna, where the metallic patch antenna offers low radiation efficiency at high frequencies.

Millimeter wave frequencies require high-gain antennas to reduce the signal attenuation because of the oxygen molecule absorption. The high-gain system is provided by using an antenna array with the good feeding network in [7]. In addition, with that extra-gain enhancement is done via superstrate technology where the dielectric slab of thickness, approximately $\lambda/4$ is kept over the radiating patch at $\lambda/2$, approximately. The multiple reflections generated between the dielectric slab

and the radiating element will boost up the gain of the system. That is, the substrate layer above the substrate will act as a focusing lens for the main beam radiation of the antenna system.

The proposed antenna in [7] offers broadside radiation pattern with small size and very simple feed network. Also, it has less optimization parameters and shows that the reflection coefficient of –10 dB over the frequencies 27–29.5 GHz. For simulation and optimization, CST an industry standard software is used which works on finite integration technique (FIT).

Here, the four identical DD patch antennas form an array and are fed by one-to-four Wilkinson power dividers. In order to reduce the losses due to surface wave, the electromagnetic band gap (EBG) structure is introduced and simulated finally. The performance parameters such as side-lobe level (SLL), front-to-back ratio (FBR), gain, bandwidth, radiation efficiency and total efficiency of the metallic patch antenna, proposed DD patch antenna, and the proposed DD patch antenna with EBG structure are explained in [7], and it shows that the proposed DD patch array antenna is a good candidate for the future 5G mobile communication systems with the increased gain of 16.6 dBi and radiation efficiency of 71% which is low when compared with the conventional methods.

Three different configurations of patch phase array antennas stated in [8] are fed with inset feed line and operating in 28 GHz. The antenna type 1 is designed such that the array feed line in the same direction and types 2 and 3 are designed with the antenna array feed line in the opposite direction. The measured and simulated results were discussed for the antenna excited at phase differences of 0^{Θ}, 360^{Θ}, and 180^{Θ}. The results show that antenna 1 provides constructive interference while excited at 0^{Θ} and 360^{Θ} and achieves the gain of 8.64 dB and gives destructive interference while excited at 1800 and achieves the gain of 6.30 dB since the feed line is in the same direction. Antennas 2 and 3 provide destructive interference while excited at 0^{Θ} and 360^{Θ} and achieve the gain of 7.47 and 7.46 dB, respectively, and give constructive interference while excited at 180^{Θ} and achieve the gain of 7.87 and 8.45 dB, respectively, since the feed line is in the opposite direction. Also, it offers good isolation performance over a wide frequency range from the patch size which is made small.

A novel broadband-printed elliptical-slot antenna and a dual-band slot antenna are given in [9] which are able to cover the wide range of mm-wave frequency band 28–38 GHz. The broadband slot antenna is exited with 50-ohm proximity-coupled feed line, and it offers the impedance bandwidth of 20–40 GHz which covers the required frequency of 28 and 38 GHz for mm-wave applications. The dual-band slot antenna is designed by introducing a notch band of 30–35 GHz in the broadband antenna stated above. The proposed antennas provide the omnidirectional pattern and offer the gain of 5.3 and 5.6 dBi and radiation efficiency of 93 and 94%, respectively.

A wideband circular patch antenna of fractal-like tree structure is presented in [10], which covers five different frequency bands such as 1.6–7.5, 8–10, 10.5–12.2, 12.7–16.1, and 19.3–23 GHz. Since it is operating in five different bands, it is expected to be a good candidate for the ultra wideband (UWB) applications.

A new 4*4 aperture-coupled DD patch antenna array is proposed in [11] and is employed by printed circuit board (PCB) technology. It is operating at 60 GHz, provides gain up to 16 dBi, and follows a broadside radiation pattern. It is realized with three PCB substrates with substrates 1 and 2 with permittivity 2.24 and substrate 3 with high permittivity of 10.2, and introducing resonance by proper aperture feeding will improve the impedance bandwidth wider than 20%. Substrate 1 will act as a feeder offers parallel feeding network to do aperture-coupling, and the DD patch array is realized by the substrate 3. Simulation is done with a high-frequency structure simulator (HFSS) software, and parameters are measured and simulated. It provides the impedance bandwidth of 23.7% and 3-dB gain bandwidth of 32.5%. The DD patch array delivers high radiation efficiency of the order of 71–83% upon the operating frequency band.

A novel wideband perforated dense dielectric patch array antenna with single-layered substrate integrated waveguide (SIW) feed network is projected in [12] and provides good radiation characteristics and wide bandwidth. It is working on 60 GHz operating frequency and fabricates with PCB technology. At very high frequencies, mm-wave applications require a high-gain antenna because of high propagation loss and are attained by the microstrip patch array nowadays. Wide bandwidth can be obtained by applying a thick substrate between the patch and the ground plane, but it could cause heavy surface waves which will affect the radiation performance. It can be suppressed by the following methods—cavities inside the substrate [13–15], soft-surface technique [16], and uniplanar compact EBG method [17]. In addition, with that, multilayer LTCC technology is adopted in [18], and it is found that it will increase the performance but the manufacturing cost is high. A multilayered PCB DD antenna is discussed in [19], it is found to be the low-cost, high-performance method, and this PCB technology are preferred and used nowadays.

The proposed antenna is made by the substitute of metallic patch that is the dense dielectric material and has simple geometry, and offers high gain of the order of 16 dBi. Backward radiation from aperture-coupling is reduced by the reflector so that the radiation characteristics are preserved. This is done by the single-layered corporate feed network with substrate integrated waveguide (SIW). And, the DD patch array is formed by adopting the perforated method, since these perforations allow high-frequency small patch bands to be placed inside the perforated low-frequency patches, thereby allowing rapid change in current directions and making cross-polarization possible. The proposed antenna is simulated through HFSS and fabricated to verify the simulated parameter values. The simulated and measured bandwidth is 47.9–67.9 GHz and 52.5–66.5 GHz, respectively, gain is 17.9 and 17.5 dBi, respectively, and radiation efficiency is around 60–69%. It follows unidirectional radiation pattern. Because of the excellent parametric values, this antenna is mostly suited for the mm-wave frequency applications. A simple rectangular patch antenna operating at 38 GHz with 4 GHz bandwidth is designed in [20]. The antenna is made as simple as possible to meet the requirement of 5G systems.

3 System Parameters

Antenna design requires initially some set of dimensional formulas to frame the structure and plays a major role in determining the antenna function and parasitic effect and analytical method to be chosen. One should have the knowledge about the materials for antenna to fix the operating frequency, propagation, and bandwidth. Then designing and validating the shape, geometry, and parameters of the antenna using a simulator prior to fabrication gives very good results in designing an antenna because most of the designs require some adjustments or variations in dimensions or geometry to obtain the best performances. While designing antennas operating at high frequencies, and which have array topologies may poses system challenges like memory requirement, system complexity and run time, etc. Some antennas can be fabricated easily, but there will be some complications in simulation and vice versa. For example, the wideband perforated-DD patch antenna presented in [12] is easy to fabricate, and during simulation because of more air holes in the geometry of the antenna it requires more computer memory and simulation time.

The simulator can be selected based on the type and size of the antenna to be designed. Several methods for analytical modeling are as follows: Method-of-moment (MOM) method which is suited for planar structures, finite element (FEM) and finite difference time domain (FDTD) solutions which are fit for 3D structures like horn, pyramidal, waveguide, and dish. The various simulators used are FEKO, Ansoft HFSS, CST, Agilent ADS, MATLAB, and ANTENNA MAGUS.

A high-frequency structural simulator (HFSS) is an advanced recent 3D EM simulator and operates with finite element method (ELM) which can be useful for designing linear circuits, filters, antennas, transmission lines, etc. The circularly polarized patch antenna described in [5], broadband and dual-band slot antenna proposed in [9], the wideband circular patch antenna of fractal-like tree structure proposed in [10], the 4*4 aperture-coupled DD patch array antenna given in [11], and the wideband perforated dense dielectric patch array antenna presented in [12] are simulated using HFSS, and the parameters are verified by the fabrication results.

CST is a unique CAD software which is used to design antennas and components which support high frequencies, and it is configured to use both MOM and FDTD methods of analytical modeling depending on the need of the designer. The dense dielectric patch antenna with EBG structure given in [7], three different configurations of patch phase array antenna stated in [8], and broadband and dual-band slot antenna proposed in [9] are simulated with CST configured with finite integrated (FIT) method which is equivalent to the FDTD method, and also it offers finite array analysis for designing small and large finite-array antennas which are popular recently. Antenna and array synthesis are provided by Antenna Magus. Agilent advanced design system (ADS) is a broad, application-specific integrated-set of accurate, fast, and easy to use interface simulator, which provides full standard-based design and verification used for designing RF, microwave, and high-speed digital applications.

A method of providing wide bandwidth is given in [21] called beamforming technique, it shows that greater than 500 MHz bandwidth is achievable at 28 GHz operating frequency with the given technique, and the important thing is that it helps to reduce the system complexity very greatly. Analog beamforming is a simple and effective method for providing high gain and is less flexible, whereas digital beamforming offers a vast degree of freedom and better performance at the expense of high cost and complexity. Hence, the hybrid beamforming technique can be adopted with the positives of the above techniques which can provide a considerable reduction in the system complexity. And to make a highly secure and reliable access technique by providing a fast-handoff between adjacent beams, one can deploy 2D N*M patch array antenna to achieve space division multiple access (SDMA) also termed as beam division multiple access (BDMA) as explained in [2].

4 Discussion

The various antennas proposed recently are discussed in the previous chapters and we came to know that for a reliable antenna system the size of the antenna should be small enough because the antenna gain is a function of frequency and not as a function of physical size and is illustrated with an example in [21]. Our aim is to find a good antenna for the 5G systems; hence, the operating frequency must be of the order of GHz to support mm-wave frequencies. Since efficiency is the cost of gain, it should be greater than 12 dBi to offer high efficiency. Hence, it is possible to have antenna arrays instead of antenna elements as given in [21] to increase the antenna gain and is adopted in various proposed methods we presented above.

Wireless communication antennas must survive with the varying environment due to the destructive signal overlap during multipath transmission, which leads to adopting diversity techniques and inbuilt multiple-input-multiple-output (MIMO) antennas. These antennas should be designed with some factors to achieve good diversity. Some of the factors are as follows: They should be small and efficient like antenna element and should have some degree of cross-polarization discrimination, less coupling effects, and with perfect matched termination. In order to provide the expected data demand for 5G mobile systems, the most effective method is having a wide bandwidth antenna which can be achieved with beamforming technique and we can have a wideband antenna operates in different frequency bands as discussed in [10].

The 5G system should have a high spectral efficiency, which is the factor of modulation method and the multiple access scheme involved, so it must be designed with the techniques which have better performance than the schemes used in 4G–LTE. The various schemes which have better performance than the previously used orthogonal frequency division multiplexing (OFDM) and orthogonal frequency division multiple access (OFDMA) are as follows: the generalized frequency division multiplexing (GFDM), bi-orthogonal frequency division multiplexing (BFDM), universal filtered multicarrier (UFMC), and filter bank multicarrier (FBMC) stated in [2].

To increase the antenna performance, system-level simulation with antenna analysis and antenna synthesis is very important before fabrication of the proposed work, and it leads to an efficient antenna design. Various modeling and simulators are discussed above for validating the antenna performance, and the thing is that the simulator should be chosen to have less system complexity and less memory requirements. Different antennas with microstrip patch, dense dielectric patch antennas, operating at different mm-wave frequency bands following different polarization effects are discussed above, and the parameters are presented in Table 1 to show the performance comparison. Most of the antennas are operating at 28 GHz and unlicensed at 60 GHz, which may be the optimal frequency band for the future generation. And, most of them provide good gain, bandwidth, and radiation efficiency.

Table 1 Performance comparison of the discussed antennas

Parameter/ Antenna	BW (GHz)	Operating frequency (GHz)	Gain (dBi)	Radiation efficiency (%)	Simulator
CP micro strip patch antenna [5]	3.58–4	3.77	5	_	HFSS
DD patch antenna [7]	27.1–29.5	28	12.48	92	CST
DD patch array antenna with EBG [7]	27–>30	28	16.3	71.8	CST
Patch phase array antenna- type 1 [8]	1.4	28	8.64	–	CST
Patch phase array antenna- type 2 [8]	1.4	28	7.87	–	CST
Patch phase array antenna- type 3 [8]	1.4	28	8.45	–	CST
Broadband-printed slot antenna [9]	20–40	28 & 38	5.3	93	CST & HFSS
Dual-band printed slot antenna [9]	20–41	28 & 38 Notch at 33	5.6	94	CST & HFSS
Aperture-coupled patch antenna [11]	–	60	16	71–83	HFSS
Perforated-DD patch array antenna [12]	52.5–66	60	17.5	60–69	HFSS
Simple rectangular patch antenna [20]	35.92–39.98	38.14	31 dB	–	HFSS

5 Conclusion

In this work, many mm-wave antennas proposed for 5G cellular communication are studied and their important performance parameters such as gain, bandwidth, operating frequency, and radiation efficiency are compared to give an idea about the various antenna parameters to be considered while designing the reliable antenna. In addition, with that the other antenna parameters such as size, antenna array, and system parameters like system complexity and simulation time are also discussed to design an efficient antenna suitable for 5G mobile communication since the design of the antenna is a major challenge for the upcoming mobile technologies.

References

1. White paper - 5G use cases and requirements, Nokia Solutions and Networks (2014)
2. Mitra, R.N., Agrawal, D.P.: 5G mobile technology: a survey. ICT Express 1 (2015)
3. Tan, H., Li, W., Wang, T., Fang, J., Feng, Z.: The analysis on the candidate frequency bands of future mobile communication systems. Commun. China **12**(Supplement), 140–149 (2015)
4. Benn, H.: Vision and key features for 5th generation (5G) cellular, Samsung R&D Institute UK (2014)
5. Mak, K.M., Lai, H.W., Luk, K.M., Chan, C.H.: Circularly polarized patch antenna for future 5G mobile phones. IEEE Access **2** (2014)
6. Lai, H.W., Luk, K.-M., Leung, K.W.: Dense dielectric patch antenna—a new kind of low-profile antenna element for wireless communications. IEEE Trans. Antennas Propag. **61** (8) (2013)
7. Haraz, O.M., Elboushi, A., Alshebeili, S.A., Sebak, A.-R.: Dense dielectric patch array antenna with improved radiation characteristics using EBG ground structure and dielectric superstrate for future 5G cellular networks. In: Special Section On 5 g Wireless Technologies: Perspectives Of The Next Generation Mobile Communications And Networking (2014)
8. Yua, L.C., Kamarudinb, M.R.: Investigation of patch phase array antenna orientation at 28 GHz for 5G applications. In: 2016 International Electrical Engineering Congress, iEECON2016, Chiang Mai, Thailand, Procedia Computer Science, vol. 86, 2–4 March 2016
9. Ashraf Q1, N., Haraz, O.M., Mahmoud Ali, M.M., Ashraf, M.A., Alshebili, S.A.S.: Optimized broadband and dual-band printed slot antennas for future millimeter wave mobile communication. AEU - Int. J. Electron. Commun. **70**(3) (2016)
10. Hamdy, S., El-Khouly, A., Zaki, A., El-Khamy, S.E.: A new fractal- like tree structure of circular patch antennas for 5G multi-band and wide-band applications. In: Progress in Electromagnetic Research Symposium (PIERS), Shanghai, China, 8–11 (2016)
11. Li, Y., Luk, K.-M.: A 60-GHz dense dielectric patch antenna array. IEEE Trans. Antennas Propag. **62**(2) (2014)
12. Li, Y., Luk, K.-M.: Wideband perforated dense dielectric patch antenna array for millimeter-wave applications. IEEE Trans. Antennas Propag. **63**(8) (2015)
13. Lamminen, E.I., Säily, J., Vimpari, A.R.: 60-GHz patch antennas and arrays on LTCC with embedded-cavity substrates. IEEE Trans. Antennas Propag. **56**(9), 2865–2874 (2008)
14. Yeap, S., Chen, Z.N., Qing, X.: Gain-enhanced 60-GHz LTCC antenna array with open air cavities. IEEE Trans. Antennas Propag. **59**(9), 3470–3473 (2011)
15. Xu, J.F., Chen, Z.N., Qing, X., Hong, W.: Bandwidth enhancement for a 60 GHz substrate integrated waveguide fed cavity array antenna on LTCC. IEEE Trans. Antennas Propag. **59** (3), 826–832 (2011)

16. Wang, L., Guo, Y.X., Sheng, W.X.: Wideband high-gain 60-GHz LTCC L-probe patch antenna array with a soft surface. IEEE Trans. Antennas Propag. **61**(4), 1802–1809 (2013)
17. Lamminen, E.I., Vimpari, A.R., Säily, J.: UC-EBG on LTCC for 60-GHz frequency band antenna applications. IEEE Trans. Antennas Propag. **57**(10), 2904–2912 (2009)
18. Zhang., Titz, D., Ferrero, F., Luxey, C., Zhang, Y.P.: Integration of quadruple linearly-polarized microstrip grid array antennas for 60-GHz antenna-in-package applications. IEEE Trans. Compon. Packag. Technol. **3**, 1293–1300, (2013)
19. Li, Y.J., Luk, K.-M.: A 60-GHz Dense Dielectric Patch Antenna Array. IEEE Trans. Antennas Propag. **62**(2), 960–963 (2014)
20. Jajere, A.M.: Millimeter wave patch antenna design antenna for future 5G applications. Int. J. Eng. Res. Technol. (IJERT), **6**(02) (2017)
21. Roh, W., Seol, J.-Y., Park, J., Lee, B., Lee, J., Kim, Y., Cho, J., Cheun, K.: Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results. In: 5G Wireless Communication Systems: Prospects and Challenges. IEEE Communications Magazine (2014)
22. Chu, H., Guo, Y.X., Wang, Z.L.: 60-GHz LTCC wideband vertical off-center dipole antenna and arrays. IEEE Trans. Antennas Propag. **61**(1), 153–161 (2013)

A Comprehensive Survey of Machine Learning-Based Network Intrusion Detection

Radhika Chapaneri and Seema Shah

Abstract In this paper, we survey the published work on machine learning-based network intrusion detection systems covering recent state-of-the-art techniques. We address the problems of conventional datasets and present a detailed comparison of modern network intrusion datasets (UNSW-NB15, TUIDS, and NSLKDD). Recent feature-level processing techniques are elaborated followed by a discussion on supervised multi-class machine learning classifiers. Finally, open challenges are pointed out and research directions are provided to promote further research in this area.

1 Introduction

Cyber-attacks against individuals, businesses, and nations hit headlines almost every day as computers, mobile communication devices, IoT devices, and networks have become an integral part of life. Unfortunately, these devices are attacked by network intruders and cyber-criminals with malicious intent. Powerful security defense mechanisms involving intrusion detection systems are essential for detecting such cyber-attacks. Since the administrator cannot manually inspect each packet entering or leaving the network, it is desirable that the machine learns itself the behavior of incoming packets and detects any suspicious or anomalous behavior.

The aim of this paper is to present a comprehensive survey of machine learning-based network intrusion detection. There are multiple surveys previously published in this area: [1] focusses on four major categories (statistical, information theory, classification, and clustering), [2] covers various methods, systems, and

R. Chapaneri (✉) · S. Shah
Mukesh Patel School of Technology Management and Engineering,
NMIMS University, Mumbai, India
e-mail: radhika.chapaneri@nmims.edu

S. Shah
e-mail: seema.shah@nmims.edu

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_35

validity measures of network anomaly detection, [3] presents various distance and similarities measures, and [4] compares available commercial NIDS. Even though all these surveys provide a valuable contribution, this paper covers various facets of machine learning-based NIDS focusing on recent trends in terms of datasets, feature analysis, and machine learning techniques. The **contributions** in this paper are (a) comparison of modern datasets with conventional datasets relative to current network scenarios and attacks, (b) recent features transformation and selection techniques are compared along with feature analysis of recent dataset, (c) overview of various machine learning techniques is used in the field of network intrusion detection in last 5 years, and (d) open challenges in this area are pointed out and research directions are suggested to tackle these challenges.

Section 2 provides an overview of intrusion detection systems, Sect. 3 presents comparative analysis of various datasets, Sect. 4 focuses on feature-level processing, and Sect. 5 surveys various machine learning techniques. Section 6 presents open challenges to be solved in this area followed by conclusions in Sect. 7.

2 Overview of Network Intrusion Detection System (NIDS)

Performing an intrusion or attack is a multistage process consisting of (a) reconnaissance, (b) vulnerability identification and scanning, (c) gaining access and compromising system, (d) maintaining access and creating backdoors, and (e) covering tracks.

Table 1 presents the characteristics of various intrusions; the detailed taxonomy of tools and attacks can be found in [5]. IDS is used to classify incoming data either as malicious or as normal traffic. There are three approaches for deployment of IDS: host-based (HIDS, machine-specific), network-based (NIDS, network-specific and its interfaces), and hybrid IDS [6]. The detection approaches can be categorized into signature-based and anomaly-based. In signature-based detection (also known as misuse detection) approach, a database of the malicious pattern (signature) is created; the matching algorithm is applied on incoming data, and if the pattern is matched with known signature, an alarm is raised. A major advantage of this approach is the low false positive rate (FPR), but it fails to detect novel zero-day attacks; also, the signature database is required to be constantly updated. In contrast, the anomaly-based approach creates a model (i.e., learns the characteristics) of normal behavior and detects deviations or anomaly stimulated due to the attack. With this approach, generalization of behavior can be achieved in the testing phase allowing identification of attack in real time. A drawback of this approach is that it suffers from high FPR and hence low detection accuracy considering modern attacks on network. Therefore, this area is growing and a challenging area of research.

Table 1 Network intrusion characteristics

Intrusions	Characteristics
Probe/ Reconnaissance	Attacker gathers information about the network, including number of machines, type of OS.
Exploit	Exploits known vulnerability of networks
Backdoors	Security mechanism is bypassed stealthily to access network data
U2R	Gain illegal access to root account to manipulate confidential resources
R2L	Gain local access as a user of a targeted machine remotely
Worms	Malicious code replicates itself to spread to other computers on the network
DoS/DDoS	Attacker disrupts the normal computing operation and makes the service unavailable to legitimate users

3 Comparative Analysis of Network Intrusions Datasets

Datasets play an important role in evaluating the performance of IDSs since they can be used for repeating experiments and validating new techniques. The authors in [8] have set several guidelines for creating NIDS datasets: (a) realistic network and traffic, (b) labeled dataset, (c) total interaction capture, and d) diverse intrusion scenarios. IDS datasets typically consist of following categories of features: (a) **basic features** which include features of TCP/IP connection, (b) **time-based features** which capture temporal characteristics of network data; for example, features that examine all connections having the same destination in last 2 s so that DoS and probe attack (which involves sending a huge amount of traffic data to the same host) can be easily discriminated and identified, and (c) **content features** which are extracted using domain knowledge to capture suspicious behavior in the data, for example unauthorized logged in as root or number of failed login attempts.

3.1 Conventional Datasets

The first benchmark dataset of IDS was KDD99 released by DARPA. Extensive test bed network was created, and different types of attacks were simulated and categorized into four groups: DOS, U2R, R2L, and Probe. The dataset consists of 41 features and a label to specify either normal or attack (and type of attack). Even though widely used, this dataset has inherent flaws as pointed out in [8]. To avoid the limitations of KDD99 dataset, NSL KDD dataset was created having the following advantages [9]: (a) Redundant records were removed to decrease classifier bias, (b) duplicate records were removed, (c) new train and test sets were created containing reasonable amount of samples for full set training and testing, and (d) a number of samples are inversely proportion to difficulty level of the particular attack category.

3.2 Modern Datasets

Since the complexity of network and types of attacks have changed over time, the NSL KDD dataset is therefore not relevant for the current evaluation of modern network scenarios. Hence, efforts are geared toward creating benchmark datasets that can capture modern attack traces either via simulation or via real-time network traffic. Recent datasets are discussed followed by a comparative analysis of two specific datasets with NSL KDD in Table 2.

Table 2 Comparison of benchmark and recent datasets (UNSW-NB15, TUIDS)

Parameters	NSL KDD	UNSW-NB 15	TUIDS
Year of creation	KDD99—1999 NSL KDD—2009	2015	2015
Modern network traffic and attacks	No	Yes	Yes
Type of network traffic	Simulated	Simulated	Real
Attack traffic generation	Simulated	IXIA perfect storm tool	Modern tools in test bed network
Duration of data collected	7 weeks	16 h and 15 h	7 days
Feature extraction tools	BroIDS	BroIDS, Argus, C# algorithms	Distributed feature extraction framework
Feature categories	Basic, content, time, connection	Basic, content, time, connection	Basic, content, time, connection
# Features	41	47	50
# Attack categories	4	9	2
Number of samples (training set)	**N**:67,343; **D**:45,927; **P**:11,656; **U**:52; **R**:995	**N**: 56,000; **F**: 18,184, **A**: 2,000; **B**: 1,746; **D**: 12,264; **E**: 33,393; **G**: 40,000; **R**: 10,491; **S**: 1,133; **W**: 130	**N**: 71,785; **D**: 42,592; **P**: 7550
Number of samples (testing set)	**N**:9,710; **D**:7,458; **P**:2,422; **U**:67; **R**:2887	**N**: 37,000; **F**: 6,062, **A**: 677; **B**: 583; **D**: 4,089; **E**: 11,132; **G**: 18,871; **R**: 3,496; **S**: 378; **W**: 44	**N**: 47,895; **D**: 30,613; **P**: 7757
Publicly available	Yes	Yes	Yes

N: Normal; **D**: DoS; **P**: probe; **U**: U2R; **R**: R2L; **A**: analysis; **B**: backdoor; **E**: exploits; **F**: fuzzers; **G**: generic; **R**: reconnaissance; **S**: shellcode; **W**: worms

(a) **UNSW NB15** [10]: This dataset represents modern low footprint attacks and was developed by Australian Center for Cyber Security (ACCS). IXIA perfect storm tool was used to create hybrid of modern network scenarios and nine classes of attack traffic. Since it contains information about new attack types updated continuously from CVE (dictionary of publicly known information security vulnerabilities), the dataset captures modern network intrusion characteristics. Two simulation scenarios were conducted, first for 16 h with 1 attack/sec and second for 15 h with 10 attacks/sec. Labeling of the dataset was done via the IXIA report. From network traces, 43 features were extracted that include basic (14), content (8), time (9), connection (7), and additional (5) features. The description of each feature is provided in [10], and the distribution of number of samples of each category is provided in Table 2. UNSW NB15 is a complex dataset [11] as it represents modern network and attack traffic and can be used for reliable evaluation.

Figure 1a illustrates the analysis of feature values of this dataset where all features are normalized via z-score. We observe that some features are redundant; for example, values of feature indices 14 to 19 are almost similar for all 10 categories and hence such features need to be discarded via some feature selection technique. Also, using boxplot analysis of Fig. 1b, we observe that there is a significant variation of feature values among the ten categories which will result in better discrimination for classifiers.

(b) **TUIDS** [12]: A test bed network was developed at Tezpur University consisting of 5 different networks by configuring 250 hosts across several VLANs, multiple switches, wireless controllers, and routers. Real-time network traffic was generated based on daily activities of students, faculties, system administrators, etc., and attack traffic was generated by launching attacks in test bed network using modern tools such as targa, nmap, brutessh. Network traffic of 7 days was captured using libpcap, and gulp and fast distributed framework

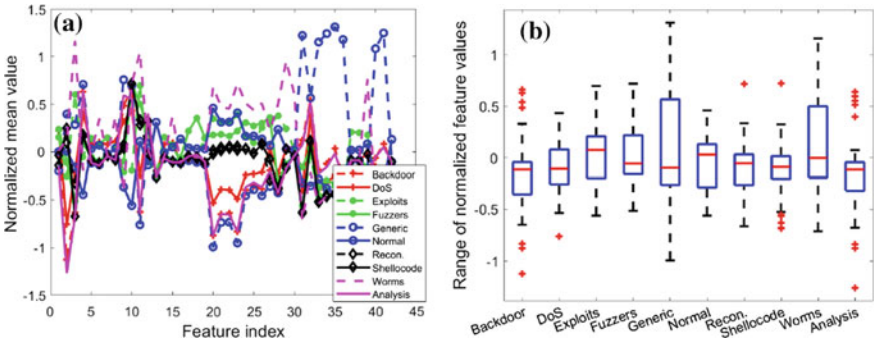


Fig. 1 Feature analysis of UNSW NB15 dataset: **a** normalized average 43 feature values for 10 categories and **b** boxplot showing feature variations of 10 categories

were used to extract 50 features consisting of basic (25), content (5), time (10), and connection (10)-based traffic features. Due to unavailability of training and test files, we could not do feature analysis of this dataset.

(c) There are few other datasets such as **CAIDA**, **LBNL,** and **ISCI,** but they have the problem of anonymity and payload is completely removed because of which there is less utility of these datasets to researchers. **KYOTO** dataset is made using Honeypots consisting of email and DNS traffic only. **IRSC** dataset contains both simulated and real attack data, but it is not publicly available.

(d) In [8], authors have designed a systematic approach to generate dynamic datasets which keeps on updating current traffic pattern and newer types of attacks. Profiles are developed that reflect an abstract representation of characteristics of attacks and their features and aims at reproducibility of experiments to simulate attack traffic.

(e) In [7], a tool-based method was developed to create a controlled and reproducible environment for generation of intrusion datasets. The dataset aims at generating known, similar, and new attacks.

From Table 2, we observe that the recent two datasets UNSW NB15 and TUIDS satisfy many guidelines sets in [8]. Since UNSW-NB15 benchmark dataset is publicly available, we recommend its use for further research in this area.

4 Feature-Level Processing of Network Intrusion Datasets

Typically, the data of network intrusion datasets is huge and high dimensional and all attributes may not be relevant in building the IDS model. Preprocessing of feature matrix $\mathbf{X}$ is a basic step for preparation of adequate feature vectors, wherein either normalization given by Eq. (1) (left) or standardization by z-score given by Eq. (1) (right) is performed, where $\mathbf{X}_{uw}$ is the uth feature value of $\mathbf{x}_u^{\mathrm{T}}$ corresponding to the wth sample, $N_{\min}$ and $N_{\max}$ are the desirable range of feature values (e.g., [0, 1]), and $\mathbf{x}_u^{\mathrm{T}}$ and s_{xu} are the mean and standard deviation of feature vector $\mathbf{x}_u^{\mathrm{T}}$.

$$\mathbf{X}'_{uw} = \frac{\mathbf{X}_{uw} - \min\left(\mathbf{x}_u^T\right)}{\max\left(\mathbf{x}_u^T\right) - \min\left(\mathbf{x}_u^T\right)} \cdot \left(N_{\max} - N_{\min}\right) + N_{\min}\,;\; \mathbf{X}'_{uw} = \frac{\mathbf{X}_{uw} - \bar{\mathbf{x}}_u^T}{s_{xu}} \qquad (1)$$

Next, to achieve the discriminative capability of classifiers, it is desirable to reduce the dimensionality of features either through feature transformation or through feature selection. In the former, techniques such as principal component analysis (PCA) and independent component analysis (ICA) are used due to which the resulting features are projected onto a transformed feature space, wherein only the most important feature dimensions are retained; however, by doing so, the interpretability of feature dimension is lost. In the latter approach of feature selection, redundant and irrelevant features are discarded due to which only the prominent features are retained which are interpretable. A feature x is relevant if its

removal from the full feature set F leads to a decrease in classification performance, i.e., $P(y = y'|F) > P(y = y'|F\backslash\{x\})$ and $P(y \neq y'|F) < P(y \neq y'|F\backslash\{x\})$, where y is the correct label and y' is the predicted label. The task of feature selection is to find the optimal set of features represented by q^* given by Eq. (2), where F is the full feature set, $\Phi(F, q)$ is the set with selected features, q is a binary vector of length F which indicates whether a feature x is selected ($q_x = 1$) or not ($q_x = 0$), and $\mathbf{Q}$ is measure that describes the success of feature selection.

$$q^* = \arg\min_q \mathbf{Q}\left(y, y', \Phi(F, q)\right) \tag{2}$$

The feature selection evaluation strategies can be classified into filter, wrapper, and ensemble methods. For D-dimensional features, the possible number of feature sets is $2^D - 1$ which can be searched in either sequential, exhaustive or random manner. Filters evaluate feature sets using relevance criteria such as correlation, information gain without considering classifier performance. Wrappers evaluate feature sets by measuring the quality (e.g., accuracy) of classifier and thus often achieve better results than filters but have drawbacks of overfitting and time-consuming relative to filters. Ensemble methods are directly integrated into the classifier algorithm (e.g., decision trees) which can help improve the performance, but the drawback is that they cannot be applied to other classification algorithms.

Table 3 lists the state-of-the-art techniques used for feature-level processing of network intrusion datasets. PCA is widely used for extracting relevant information as it derives set of uncorrelated features from the set of correlated features. PCA sorts the principal components by its variance and selects eigenvectors with higher eigenvalues. But it may be possible that these eigenvectors are not discriminative, hence authors in [13] select projections with higher FDR values. Even after the

Table 3 Feature selection/transformation techniques for network intrusion detection

Paper	Method	Approach	Dataset used	#Features
[13]	Feature transformation	Principal component analysis with Fisher Discriminant Ratio (FDR)	NSL KDD	8
[14]	Feature transformation	Kernel principal component analysis	KDD99	–
[15]	Filter	Ensemble of filtered, consistency, correlation-based feature selection techniques	NSL KDD, Kyoto	21, 11
[16]	Filter	Combination of correlation coefficient and information gain	NSL KDD	13
[17]	Filter	Pearson correlation coefficient	NSL KDD	17
[18]	Feature representation	Cluster center and nearest neighbor	KDD99	–
[19]	Filter and wrapper	Combination of different filters and stepwise regression wrapper approach	NSL KDD	16

reduction of features to eight, they found that the accuracy is not reduced. As PCA can only extract linear information, authors in [14] have used kernel PCA which transforms the features in high-dimensional space due to which they found approximately 5% increase in detection rate as compared to using only PCA. In [15], irrelevant features are removed using ensemble of three methods: filtered, consistency, and correlation-based feature selection techniques due to which features are reduced by 48.78% for NSL KDD and 31.25% for Kyoto dataset. In [16], initially correlation-based feature selection (CFS) is applied resulting in 10 features, but since CFS does not guarantee selection of all relevant features, information gain is used to rank the features not selected by CFS and high-ranked features are considered resulting in 13 features. In [17], authors have used multilayer approach; in the first layer, Pearson correlation coefficient is applied between the features, and in the second layer, correlation coefficient is calculated between features selected by the first layer and the class followed by C4.5 used as a classifier, resulting in 1% accuracy improvement. In [18], original dataset is transformed into one-dimensional distance by summing up the distance between a specific data sample and cluster centers and distance between data sample and its nearest neighbor. In [19], multistage feature selection is done by a combination of filters and stepwise regression wrappers to remove the bias by any specific method. The cost of every feature is analyzed and eliminated 13 very costly features. It is observed that the number of prominent features is different for various techniques as the whole dataset is not considered for evaluation. Also, the dimensionality reduction techniques for modern datasets need to be explored.

5 Machine Learning Techniques for NIDS

Supervised classification techniques are widely used for learning and generalizing the behavior of each category, e.g., normal, DoS attack, fuzzer attack. Formally, a classifier is a map f: $\mathbf{X} \rightarrow Y$, where $\mathbf{X}$ is the processed feature matrix of dimensions N (number of samples) by D (feature dimensions), Y is the set of categories of dimension $N \times 1$ indicating the class of each input sample, and f is the multi-class classifier algorithm that assigns each input sample to a class. Several well-known algorithms such as SVM, decision trees (DT), Naïve Bayes (NB), artificial neural networks (ANN) are widely used for modeling IDSs. Recently, extreme learning machines (ELMs) have been used for IDS modeling due to its specific advantage that ELMs are faster to train and can handle multi-class classification inherently. To evaluate the performance of classifier, standard metrics such as accuracy, precision, recall, F-score, area under curve (AUC) of receiver operating characteristics (ROC) are used. In this section, we survey only notable recent works (2013–2017) as earlier works are detailed in previous surveys [2] [4].

In [20], semi-supervised classification using restricted Boltzmann machines (RBMs) is used to combine generative and discriminative approaches of machine learning; the experiments concluded that performance suffers when system is

evaluated on a real network traffic instead of traces from training dataset. This indicates that transfer learning methodology should be used to improve and generalize IDS performance. In [21], *k*-best Bayesian network (BN) classifiers were averaged resulting in Bayesian model averaging (BMA) classifier performing significantly better (96% accuracy) compared to NB (92% accuracy) on NSL KDD dataset. In [15], instead of sampling the entire dataset, beta profiling is used to reduce size of dataset and an online sequential version of ELM is used to achieve 96% accuracy on NSL KDD with a significant reduction in testing time (2.6 s). Multiple kernel learning framework is used in [22] by integrating multiple kernel boosting (instead of adaptive boosting) in the kernel version of ELM with its applicability to scale for large datasets and improved accuracy with low false alarms. In [13], PCA is used to de-correlate the features and a probabilistic version of self-organizing map (SOM) is used to provide fuzzy response of the classifier. In [14], kernel PCA is used for feature reduction and SVM as a classifier where RBF function is used in KPCA and SVM is modified by embedding mean value and mean square difference values of attributes to reduce the noise caused by feature difference. In [23], simple rule sets are first determined using classification and regression trees (CARTs) followed by applying discrete wavelet transform (DWT) to transform the features and using SVM classifier to build the IDS model; the DWT features are shown to be more discriminative than standard features of NSL KDD. A modified version of SVM is proposed in [24] which uses ramp loss instead of the hinge loss of conventional SVM to address the problem of class imbalance and tackle the presence of outlier samples in dataset and shows improved performance on both NSL KDD (98.6% accuracy) and UNSW NB15 (93.5% accuracy) datasets.

In [25], a modified version of *k*-means is used to initially reduce the number of samples per class of NSL KDD dataset, followed by applying multi-level ELM and SVM algorithms to classify the input test sample; this strategy is applicable for real-time training since the computational cost of training the classifier is significantly reduced by initial clustering of samples with modified *k*-means clustering. In [7], multi-objective feature selection method is applied followed by new evaluation mechanism that classifies the anomaly into known, similar, and new attacks; this is a new paradigm that enables better performance evaluation of classifiers to detect not only known attacks but also modern zero-day attacks.

Table 4 compares the accuracy of the recent machine learning techniques used for detecting network intrusions. It is evident from the results of [24] that even after applying the same model, the accuracy on UNSW dataset is less (~93%) as compared to NSL KDD dataset (~98%). This implies that the state-of-the-art techniques that perform well for NSL KDD may not work accurately for the modern UNSW dataset; thus, further research is required to model modern attack scenarios. Also, the detection rates of low frequent attacks (R2L and U2R in NSL KDD and worms, shellcode in UNSW) are low as demonstrated by various researchers in their work.

Table 4 Accuracy of recent machine learning techniques for network intrusion detection

Paper	Technique	Description	Dataset	Acc. %
[21]	BMA	Prediction by averaging k-best BN classifier Better predictive power even with small training dataset	NSL KDD	96
[15]	OS-ELM	Fast and can process network packets one at a time or in chunk. Alpha and beta profiling to reduce the effect of imbalanced dataset	NSL KDD, Kyoto	98.66 96.37
[24]	Ramp-KSVCR	KSVCR is used for multi-class classification Ramp loss function is implemented to reduce effect of noise and outliers	NSL KDD, UNSW	**98.68** **93.52**
[25]	Multi-level (SVM, ELM)	Modified K-means to build high-quality training dataset Multi-level model using SVM and ELM	KDD99	95.75
[23]	CART, DWT, SVM	Combination of CART, DWT, and SVM to identify intrusion Visual analysis using interactive PCA to understand intrusion and their relationship	NSL KDD	95.5

6 Open Challenges in NIDS Research

Although many techniques have been developed to classify the packet into normal or attack, still NIDSs face various challenges which need to be addressed so that these systems can be widely deployed commercially. Some of these challenges are:

(a) **Huge volume of data**: IDS should have low computational complexity for training (to be able to learn behavior of new attacks) as well as for testing (for real-time performance). To solve this problem, active learning strategies can be used to select relevant input samples for training the classifier rather than using the full input training dataset.

(b) **High cost of false alarms**: Most IDSs have high false positive rate (FPR) which can be potentially disastrous for the network. If the classifier generates more false positives, the attacker can easily exploit vulnerabilities of the network; in case of false negatives, an alarm is raised even if the packet is normal leading to waste of time and effort for the network administrator. To solve this problem, probabilistic machine learning techniques must be used. With deterministic techniques, an input test sample is always classified into one of the known categories resulting in hard decisions, whereas with probabilistic techniques the decision is soft in terms of probability for each class (e.g., the sample is 80% normal, 15% fuzzer, 5% DoS) due to which the network administrator can take suitable actions.

(c) **Class imbalance**: Since the anomalies are defined as rare instances among network traces, it is obvious that the normal and anomalous class distribution will be skewed. Thus, the classifier should model this phenomenon while building the model via a cost-sensitive variation where the cost reflects the number of samples per class. To solve this problem, weighted extreme learning machine (ELM) can be implemented to improve the performance.

7 Conclusions

The constantly evolving nature of network intrusions demands IDSs that can reliably learn the behavior of incoming traffic pattern. In this work, we have presented the survey of ongoing and recent efforts addressing following aspects of machine learning-based NIDS: (i) recent datasets, (ii) feature-level processing techniques along with feature analysis of UNSW-NB15 dataset, (iii) recent machine learning algorithms including extreme learning machines and multi-level hybrid approaches. Compared to other surveys, this paper provides a comparison of conventional and modern network intrusion datasets, state-of-the-art feature selection, and machine learning techniques and highlights the challenges which can be beneficial for NIDS research community.

For further research in this area, we recommend to consider the use of modern datasets over conventional ones so that performance can be evaluated for the current network scenarios and attacks. To improve the classifier accuracy, appropriate feature selection techniques on recent datasets need to be developed so as to determine prominent features. Unsupervised learning techniques can also be explored and analyzed to solve the problem of clustering various network packets. Active learning strategies and probabilistic machine learning techniques can be implemented to improve the state of the art of NIDSs.

References

1. Mohiuddin, A., Mahmood, A., Hu, J.: A survey of network anomaly detection techniques. J. Netw. Comput. Appl. **60**, 19–31 (2016)
2. Bhuyan, M., Bhattacharyya, D., Kalita, J.: Network anomaly detection: methods, systems and tools. IEEE Commun. Surv. Tutor. **16**, 303–336 (2014)
3. Weller-Fahy, D., Borghetti, B., Sodemann, A.: A survey of distance and similarity measures used within network intrusion anomaly detection. IEEE Commun. Surv. Tutor. **17**, 70–91 (2015)
4. Garcia-Teodoro, P., Diaz-Verdejo, J., Fernandez, G.: Anomaly-based network intrusion detection: techniques, systems and challenges. Comput. Secur. **28**, 18–28 (2009)
5. Hoque, N., Bhuyan, M., Baishya, R., Bhattacharyya, D., Kalita, J.: Network attacks: Taxonomy, tools and systems. J. Netw. Comput. Appl. **40**, 307–324 (2014)

6. Scarfone, K., Mell, P.: Guide to intrusion detection and prevention systems (IDPS). NIST Spec. Publ. **800**, 94 (2007)
7. Viegas, E., Santin, A., Oliveira, L.: Toward a reliable anomaly-based intrusion detection in real-world environments. Comput. Netw. **127**, 200–216 (2017)
8. Shirvai, A., Shirvai, H., Tavallaee, M., Ghorbani, A.: Toward developing a systematic approach to generate benchmark datasets for intrusion detection. Comput. Secur. **31**(3), 357–374 (2012)
9. Tavallaee, M., Bagheri, E., Lu, W.: A detailed analysis of the KDD CUP 99 dataset. In: Compuational Intelligence for Security and Defense Applications (2009)
10. Moustafa, N., Slay, J.: UNSW-NB15: a comprehensive data set for network intrusion detection systems. In: Military Communications and Information Systems Conference (MilCIS) (2015)
11. Moustafa, N., Slay, J.: The evaluation of network anomaly detection systems: statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Inf. Secur. J.: Global Perspect. **25**(1–3), 18–31 (2016)
12. Bhuyan, M., Bhattacharyya, D., Kalita, J.: Towards generating real-life datasets for network intrusion detection. Int. J. Netw. Secur. **17**, 683–701 (2015)
13. De la Hoz, E., De La Hoz, E., Ortiz, A., Ortego, J., Prieto, B.: PCA filtering and probabilistic SOM for network intrusion detection. Neurocomputing **164**, 71–81 (2015)
14. Kuang, F., Xu, W., Zhang, S.: A novel hybrid KPCA and SVM with GA model for intrusion detection. Appl. Soft Comput. **18**, 178–184 (2014)
15. Singh, R., Kumar, H., Singla, R.: An intrusion detection system using network traffic profiling and online sequential extreme learning machine. Expert Syst. Appl. **42**(22), 8609–8624 (2015)
16. Wahba, Y., ElSalamouny, E., ElTaweel, G.: Improving the performance of multi-class intrusion detection systems using feature reduction. IJCSI Int. J. Comput. Sci. Issues **12** (2015)
17. Eid, H., Hassanien, A., Kim, T., Banerjee, S.: Linear correlation-based feature selection for network intrusion detection model. Adv. Secur. Inf. Commun. Netw., 240–248 (2013)
18. Lin, W., Ke, S., Sai, T.: CANN: An intrusion detection system based on combining cluster centers and nearest neighbors. Knowl.-Based Syst. **78**, 13–21 (2015)
19. Iglesias, F., Zseby, T.: Analysis of network traffic features for anomaly detection. Mach. Learn. **101**, 59–84 (2015)
20. Fiore, U., Palmieri, F., Castiglione, A., Santis, A.: Network anomaly detection with the restricted Boltzmann machine. Neurocomputing **122**, 13–23 (2013)
21. Xiao, L., Chen, Y., Chang, C.: Bayesian model averaging of Bayesian network classifiers for intrusion detection. In: Computer Software and Applications Conference Workshops (COMPSACW) (2014)
22. Fossaceca, J., Mazzuchi, T., Sakrani, S.: MARK-ELM: application of a novel multiple kernel learning framework for improving the robustness of network intrusion detection. Expert Syst. Appl. **42**, 4062–4080 (2015)
23. Ji, S., Jeong, B., Choi, S., Jeong, D.: A multi-level intrusion detection method for abnormal network behaviors. J. Netw. Comput. Appl. **62**, 9–17 (2016)
24. Bamakan, S., Wang, H., Shi, Y.: Ramp loss K-support vector classification-regression; a robust and sparse multi-class approach to the intrusion detection problem. Knowl.-Based Syst. **126**, 113–126 (2017)
25. Al-Yaseen, W., Othman, Z., Nazri, M.: Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Expert Syst. Appl. **67**, 396–303 (2017)

Modified Ant Colony Optimization Algorithm for Task Scheduling in Cloud Computing Systems

G. Narendrababu Reddy and S. Phani Kumar

Abstract Cloud computing is the development of distributed computing, parallel computing, and grid computing, or defined as commercial implementation of such computer science concepts. One among the day-to-day challenges in cloud computing environment is task scheduling (TS). TS is the process of allocating cloudlets to virtual machines (VM) in a cloud architecture with a concern of effective load balance and efficient utilization of resources. With the aim of facing challenges in cloud task scheduling, many non-deterministic polynomial time-hard optimization problem-solving techniques and many meta-heuristic (MH) algorithms have been proposed to solve it. A task scheduler should adapt its scheduling strategy to changing environment and variable tasks. This paper amends a cloud task scheduling policy based on modified ant colony optimization (MACO) algorithm. Main contribution of recommended scheme is to minimize makespan and to perform multi-objective task scheduling (MOTS) process. MACO algorithm will improve performance of task scheduling by reducing makespan and degree of imbalance comparatively lower than basic ACO algorithm.

1 Introduction

In recent years, data centers play an important role in global computing platforms such as sending frequent mail and other operations like search, read, and write. Cloud computing is internet-based computing system, which is highly dynamic. Task scheduling (TS) and resource allocation are main issues in cloud computing [1]. Based on the aspects of Platform as a Service (PaaS), Software as a Service (SaaS), and Infrastructure as a Service (IaaS), these computing platforms provide a

G. Narendrababu Reddy (✉)
G. Narayanamma Institute of Technology & Science, Hyderabad, India
e-mail: gnbreddy25@gmail.com

S. Phani Kumar
School of Technology, GITAM University, Hyderabad, India
e-mail: phanikumar.s@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_36

collection of host networks through the Internet. According to Information Technology Laboratory, definition of cloud computing is the process of sharing their computer processing resources and data to other devices which are on demand [2]. Cloud computing is an on-demand resource for networks, servers, programs, and services with high speed and little effort in interaction of users. One of the recent challenges in a computing platform is TS [3]. The general issue in TS is NP-hard optimization problems like travelling sales man problem and combinatorial problems like integer programming and addressing problems. These issues occur in the allocation of hundreds and thousands of virtual machines (VM) to cloud resources that cause delay in the performance of the TS [4]. To overcome these problems in TS, ant colony optimization (ACO) algorithm has been implemented [5].

In TS process, data centers, brokers, VM, and cloudlets have been created for performing the task required by using CloudSim toolkit to satisfy customer's demand [6]. CloudSim is a framework for simulation of cloud infrastructure, where data centers are resource providers, broker's help in creation and destruction of VMs, cloudlets perform and submit tasks to provide information about RAM size, bandwidth, and number of CPU allocations [7]. Here traditional ACO algorithm finds best shortest path by representing graphs. Major purpose of this algorithm is to lessen the makespan of performing task [8]. This algorithm has the capability to adopt according to dynamic applications; for example, it can modify its character from an artificial ant to real ant. Compared to FCFS (first come first serve) and RR (round-robin) algorithms, ACO finds best solutions to travelling salesman problems and reduce makespan [9]. During runtime, overhead task load gets increased along with lack of rapid adaptability which follows through a consequence of increase in execution time and decrease in convergence rate that will enact as a major pitfall to basic ACO [10].

In this article, a MACO optimization algorithm is put forth in order to achieve anticipated performance in CloudSim framework by means of selecting multiple processors, reducing makespan, and high convergence speed in minimum time. Hence, a CloudSim system can be able to perform humongous load of task with greater efficiency.

This paper is composed as follows. Section 2 gives survey about some recent existing CloudSim framework strategies. Modified ACO algorithm with its standard structure, workflow together with its fundamental functions, and specific modifications are given in Sect. 3. Section 4 presents result and experimental setup, which shows comparison study between ACO and modified ACO. Conclusion with future work is presented in Sect. 5.

2 Literature Review

Cui et al. [11] presented a multi-objective model with constraints for tackling TS issue in cloud computing. Multi-objective model takes task processing mode of cloud system into account and illustrates scheduling performance in terms of makespan, flow time, and reliability by using queuing theory and Markov process.

In the proposed method of genetic algorithm-based chaotic ant swarm (GA-CAS) algorithm, four operators and natural selection are applied to solve this embarrassed multi-objective optimization problem. The author uses more number of operators to solve TS problem, which results in complexity and inefficient functioning of GA-CAS algorithm.

HE Hua et al. [12] have presented PSO-based adaptive multi-objective task scheduling (AMOTS) strategy. Objective of advanced strategy is maximizing the resource utilization and minimize the task completion time, energy consumption, and average cost. The adaptive acceleration coefficient was implemented to show improved results of PSO algorithm, but the algorithm cost was slightly high, which is a typical pitfall to AMOTS strategy.

Chun-Wei Tsai et al. [13] have presented hyper-heuristic scheduling algorithm (HHSA) to find superior scheduling solutions for cloud computing systems. Diversity recognition and development discovery operators are utilized by the procedure to dynamically conclude that which primary level is to be employed in searching superior candidate results. Some of the drawbacks in this methodology are finding a candidate solution for high-level heuristic which is not possible; moreover, it requires effective operators to achieve a better performance.

Liyun zuo1 et al. [14] have presented the multi-objective optimization (MOO) algorithm to handle TS issues and requirements of biodiversity in cloud computing. Author introduces resource cost model, which mirrors the link between user's resource costs and budget costs. Experimental results showed that MOO method recovered only half percent in finest case development, which is not efficient at all in CloudSim environment.

Medhat Tawfeek et al. [15] have presented an approach whose main goal is minimizing the makespan of a given task set. ACO is one of the random optimization methods, which can be utilized for allocating the jobs to virtual machines. The author is not concerned about one important parameter namely task execution rather than makespan and degree of imbalance, so this strategy lacks in efficiency leading to less accurate results.

In this revision, proposed MACO algorithm helps to decrease makespan significantly in order to resolve the above-mentioned drawbacks.

3 Task Scheduling Algorithm Based on Modified Ant Colony Optimization

TS method creates a highly ordered scheduling mechanism for cloud network in order to optimize use of resources available in network. Cloudlets are assigned to every VM according to their respective allocation policy. Then cloudlets wait for a processor with more cost and span, and some processors sit idle in some VMs. These will increase both manipulation cost and makespan. These are very critical in case of processing huge amount of data. So, in this paper, MACO algorithm is employed to reduce makespan in TS. MOTS process is utilized to reduce makespan

and balance entire system workload, which helps to handle both parameters without affecting one another or without collapsing fundamental structure resulting in efficient task scheduling routine.

3.1 Ant Colony Optimization Algorithms

In ACO scheme, initially ants are placed randomly at all VMs, and all ants are set up with a level of pheromone according to MIPS, bandwidth, and a number of processors with respect to their initial VM. Then ants are allowed to move from one VM to another randomly by a process called selection of the next VM. In this process, ants select their next VM according to other trails or global pheromone by which it will normally go with a belief that as shortest path. Only in initial stage, it selects VM randomly, then compared with the predefined probability, if it is more than that random VM, it will go or select another VM. Every ant will maintain tabu table history, which will be used to check whether the selected next VM is visited or not. In case that VM is already visited means, it will repeat selection process, otherwise visit that selected VM and mark it as visited on its tabu table.

The necessity for a modification in this kind of ACO scheme forces many to formulate a new strategy based on finding an even optimized way of working with bandwidth and traffic, but all of existing left out of initial state of idle processors. And also those are all not concerned about proper feeding of tasks to each and every processor on VM to avoid overloading. These are all aims drawn up for a proposed scheme to solve, and instead of looking upon a cluster of processors as VM, it will try to optimize for each and every single processor inside every VM. A proposed scheme will perform ACO optimization for all processors in VMs based on the capability of every individual processor and store it up as an optimized list by ACO optimization. The advanced algorithm also focuses on the multi-objective task; it will balance multiple metrics.

The basic ACO algorithm dramatically boosts load of tasks in runtime, and it has lack of rapid adaptability. So, ACO increases execution time and decreases convergence rate. Hence, MACO algorithm is utilized to overcome these problems in real-time cloud environment. In addition, it will not take into account individual processors in a machine; thus, leaving some idle processors in VM results in reducing efficiency of virtual machine, and hence, it is followed by many enhancements to get more efficient ACO which is modified ACO algorithm.

3.2 Modified Ant Colony Optimization Algorithm

In this paper, the main objective of the proposed algorithm is to improve the efficiency of TS, reduce makespan, and also balance the multi-objective tasks. Initially, all ants are presented on processors randomly. Then, all the ants provide pheromone, and depending on the pheromone values, ants are allowed to move.

An ant will place pheromone in a processor according to its processing speed, amount of traffic to reach, number of processors present in that processor, etc. Then it registers that VM on its tour history; similarly, all ants will tend to move in a similar way. By this proposed scheme, developer can feed more number of tasks to more processing VM and less number to slow processing VM. Thus, developer need not worry about utilization of resources by framework and left alone of some best working processors in their VM.

Pseudo-code for proposed scheme is given below:

```
Procedure: MACO[A.Vm,k,Cldt]
  Class cldsim[Vm,cldt]
  function Main()
    Vm= new Vm
  Class cldShedMod [mips,numpes] extends cldShed
  rcl ← Rem Cld
  function CldFin(rcl)
    getcloudletfinishedlist().add(rcl);
  function updateVmprocessing
    for rgl ← cloudletTo Finish
      getCloudletExelist().remove(rgl);
      cloudletFinish(rgl)
    end for
  function cloudletSubmit(cld, filetranstime)
    for i ← MACO.k
      rcl.setMachineAndPeId(H-id,i);
  class MACO[A,k]

  A ← Ant
  pr ← Prob(A)
  function comp_dist()
    for i ∈ Pelist.size() do
      for j ∈ Pelist .size () do
        dist[i][j]= cld.size/pe.Mips;
  function main()
    BT ← BestTour
    k ← index of BT
    for i ← BT.length() − 1
      for j ← BT.length()
        if(BT[i]>BT[j])
          Tmp=k[i];
            k[i]=k[j];
          k[j]=Tmp;
        end if
      end for
    end for
  end class
```

Proposed scheme can be evaluated in several ways from that several strategies can be compared with ACO scheme in order to know about its performance, efficiency, and better utilization of resources against several existing frameworks some of them are makespan, a number of iterations took reduce makespan. MACO process balances makespan and an entire system load, in balancing time without affecting one parameter to another. In previous methods, it will reduce makespan as well as balance system load efficiently. Factors that took influence over the efficiency of a CloudSim process are as follows:

Degree of imbalance *DI*: In order to reduce overloading of tasks to a processor, increase RAM and MIPS, utilization of bandwidth, to reduce traffic for cloudlets. *DI* is used to measure imbalance among processors to spread workload evenly for every processor.

$$DI = \frac{T_{\max} - T_{\min}}{T_{avg}} \tag{1}$$

In Eq. (1), $T_{\max}$ and $T_{\min}$ are highest and shortest time spans took for a cloudlet to process in proposed framework, and T_{avg} is given as average time span took by system to process one cloudlet. Thus, DI will interpret balance of workload among different processors in order of their equally spread working time for overall task.

Execution time (makespan) T_E: Eq. (2) gives the expression for T_E. If workload of *j*th task is hosted on *i*th resource of processing capacity P_i, then execution time of *j*th task is $\frac{W_j}{P_i}$. It is assumed that execution time is proportional to execution cost. X_{ji} indicates binary variable for position of task *j* and if task was hosted on resource *i*.

Hence,

$$T_E = \left(\frac{W_j}{P_i}\right) * X_{ji} \tag{2}$$

4 Experimental Results

The proposed algorithm has been implemented in Java NetBeans 8.2 version and 32-bit operating system, 8 GB RAM as basic requirements. In this paper, MACO achieved reduction in makespan and multi-objective TS process. MACO aims to minimize makespan and balance the workload in TS, which results in proposed algorithm superior to other traditional algorithms such as basic ACO from Mehdat Tawfeek's study [15]. In the following experiments, basic ACO and MACO algorithms with different iterations were associated with makespan of 10–100 task sets.

Figure 1 indicates number of tasks in quantities like 10, 20, up to 100 task sets shown along x-axis and corresponding DI to each set of tasks shown in y-axis.

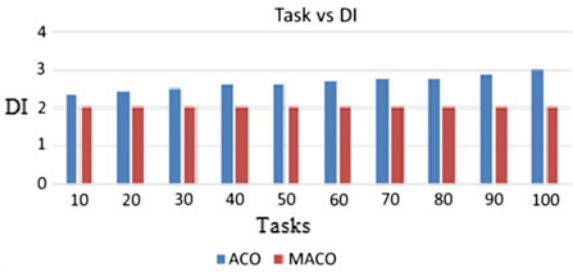


Fig. 1 Number of tasks versus DI

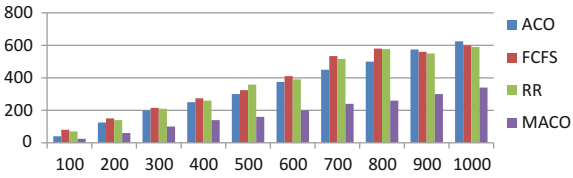


Fig. 2 Number of tasks versus makespan

Graph represents that proposed algorithm is better than basic ACO, and it is evident that MACO balances entire system load perfectly. Figure 2 indicates number of tasks like sets of 10–100 tasks shown along x-axis, and makespan of each task correspondingly ranges between 100–700 ms and is represented along y-axis obtained from Mehdat Tawfeek's study [15]. According to this graph, MACO achieves reduction in makespan better than ACO in terms of effective task scheduling.

Figure 3 represents the number of iterations like sets of 10–50 represented along x-axis with their corresponding makespan interpreted along y-axis. Graph indicates that performance of MACO algorithm is better than basic ACO algorithm, which means that proposed algorithm achieves better performance.

A TS algorithm based on MACO is implemented, and the algorithm is run for 8 problem instances with the number of processors as 8. Ten trials are done for each problem instance with ACO, and the average value of execution time of tasks and utilization of each processor are obtained.

Chart in Fig. 4 will illustrate the utilization of each processor along with the balance of load for each set of processors represented in various colors and number of cloudlets.

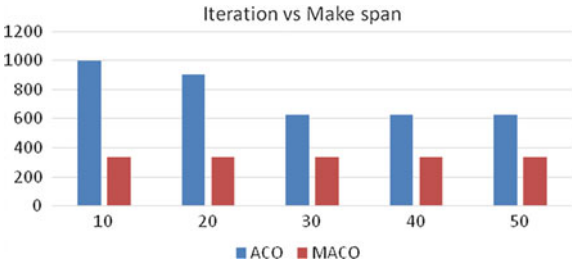


Fig. 3 Number of iterations versus makespan

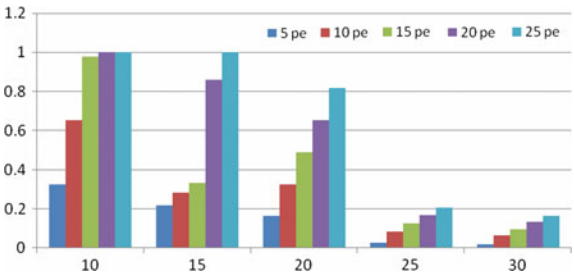


Fig. 4 Utilization of Pe versus DI

As a result, the CloudSim can be able to efficiently utilize the resources of a processor or virtual machine efficiently in order to lessen makespan of the process, and MACO can be able to schedule more number of tasks in lesser span of time.

5 Conclusion

In this paper, MACO algorithm was utilized for achieving TS with reduced makespan in a standard cloud environment. In order to overcome a drawback of ACO algorithm, makespan value may not vary throughout the runtime along with a bit slow in convergence time. Therefore, need for a new strategy arises, which leads MACO to be adopted for solving this problem and to improve the convergence speed. Experimental results showed that MACO reduced makespan effectively in addition to achievement of MOTS process efficiently. Execution of this scheme evaluated in real-time CloudSim framework gives a drastic improvement in efficient resource utilization. Proposed method achieves a better result in terms of reduction in makespan and increased load balancing capability. Even though characteristics of all tasks are same or not, MACO can handle better balance on load. Thus, MACO gave better performance compared to all other existing methodologies, which are in practice for efficient cloud computing scenario. MACO can be further implemented to various types of typical workloads on different combinations of VMs with their variety of processors that give several combinations of processing capability resulting in more efficient and less complex cloud computing frameworks.

References

1. Anagha, Y., Rathod, S.B.: Priority based task scheduling by mapping conflict-free resources and Optimized workload utilization in cloud computing. In: IEEE International Conference on Computing Communication Control and automation (ICCUBEA) (2016)
2. Anuradha, M., Selvakumar, S.: ACO based task scheduling algorithm for hybrid cloud. Int. J. Emerg. Technol. Comput. Sci. Electron. (IJETCSE) ISSN **13**(1), 0976–1353 (2015)
3. Xue, S., Li, M., Xu, X., Chen, J., Xue, S.: An ACO-LB algorithm for task scheduling in the cloud environment. J. Softw. **9**(2), 466–473 (2014)

4. Babukartik, R.G., Dhavachelvan, P.: Hybrid algorithm using the advantage of ACO and cuckoo search for job scheduling. Int. J. Inf. Technol. Converg. Serv. **2**(4), 25 (2012)
5. Deng, W., Zhao, H., Zou, L., Li, G., Yang, X., Wu, D.: A novel collaborative optimization algorithm in solving complex optimization problems. Soft Comput. 1–12 (2016)
6. Suruchi, S., Kuila, P.: Design of dependable task scheduling algorithm in cloud environment. In: Proceedings of the Third International Symposium on Women in Computing and Informatics, pp. 516–521. ACM, New York (2015)
7. Tawfeek, M.A., El-Sisi, A., Keshk, A.E., Torkey, F.A.: Cloud task scheduling based on ant colony optimization. In: Computer Engineering & Systems (ICCES), pp. 64–69 (2013)
8. Li, K., Xu, G., Zhao, G., Dong, Y., Wang, D.: Cloud task scheduling based on load balancing ant colony optimization. In: Sixth Annual Chinagrid Conference (ChinaGrid), IEEE, pp. 3–9 (2011)
9. Razaque, A., Vennapusa, N.R., Soni, N., Janapati, G.S.: Task scheduling in cloud computing. In: IEEE Long Island Systems, Applications and Technology Conference (LISAT), pp. 1–5 (2016)
10. Nizomiddin, B.K., Choe, T-Y.: Dynamic task scheduling algorithm based on ant colony scheme **7**(4) (2015)
11. Hongyan, C., Li, Y., Liu, X., Ansari, N., Liu, Y.: Cloud service reliability modelling and optimal task scheduling. IET Commun. 1–12 (2016)
12. He, H., Xu, G., Pang, S., Zhao, Z.: AMTS: Adaptive multi-objective task scheduling strategy in cloud computing. China Commun. **13**(4), 162–171 (2016)
13. Tsai, C.W., Huang, W.C., Chiang, M.H., Chiang, M.C., Yang, C.S.: A hyper-heuristic scheduling algorithm for cloud. IEEE Transac. Cloud Comput. **2**(2), 236–250 (2014)
14. Zuo, L., Shu, L., Dong, S., Zhu, C., Hara, T.: A multi-objective optimization scheduling method based on the ant colony algorithm in cloud computing. IEEE Access **3**, 2687–2699 (2015)
15. Panda, S.K., Jana, P.K.: Normalization-based task scheduling algorithms for heterogeneous multi-cloud environment. Inf. Syst. Front. 1–27 (2016)

Design of 5G mm-Wave Antenna Using Line Feed and Corporate Feed Techniques

R. Thandaiah Prabu, M. Benisha, V. Thulasi Bai and R. Ranjeetha

Abstract In this contribution, we present three different configurations of Microstrip Patch Array Antennas (MPA's) with three different materials (FR-4, RT-Duroid, and Teflon) to examine their gain with different feed methods at 28 GHz for 5G applications. Measured Return Loss (S11) and gain are presented for all the types. The simulated result showed that the designed antennas are able to operate at 28 GHz. The designs for all the MPA's are simulated using Advanced Design System 2009 (ADS) software.

1 Introduction

The mobile phone technology introduced in the early 1990s are growing in a very tremendously in the short span. The technology begins with 1G and grown as 2G, 3G, now 4G and very soon to be 5G. All these generations of technologies are evolved incompatible with the previous generation and with lots of innovations and new features and facilities. All these mobile communication systems are using the current cellular spectrum in the range of 300 MHz–3GHZ and are being overcrowded. This is why the newly evolving generation 5G is shifted to millimetre range, where the antenna gain is expected to be 12 dBi and bandwidth is 1 GHz,

R. Thandaiah Prabu (✉) · M. Benisha
Anna University, Chennai, India
e-mail: thandaiah@gmail.com

R. Thandaiah Prabu · M. Benisha
Jeppiaar Institute of Technology, Sriperambuthur, India
e-mail: benishaxavier@gmail.com

V. Thulasi Bai
KCG College of Technology, Chennai, India
e-mail: thulasi_bai@yahoo.com

R. Ranjeetha
Prince Dr. K. Vasudevan College of Engineering and Technology, Chennai, India
e-mail: ranji.ae001@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_37

and the operating frequency is nearly 28–30 GHz. Thus the researches are going on for the designing of the printed antenna for 5G mobile communication [1, 2].

Microstrip antennas introduced in 1950s are realised in 1970s after the development of printed circuited Boards (PCB). The verity of Microstrip antennas are preferred for the modern wireless communication since it is easy to fabricate and feed, having low profile, lighter weight and low cost, planar configuration, superior portability, and easy to use with the array or to integrate with other microstrip elements as Monolithic Microwave Integrated Circuits (MMIC's) [3, 4]. Moreover, patch antennas are used as simple and highly preferred in many applications such as Mobile Radio, Broadcast Systems, Radio Frequency Identification (RFID), Global Positioning System (GPS), Surveillance systems and so on because of the following special features such as twofold characteristics, circular polarization, double frequency operation, frequency alertness, wider bandwidth, beam streaming and flexible feed line can be easily obtained from these patch antennas as described already [5]. Even though some of the disadvantages are there for the Microstrip antenna such as narrow bandwidth, spurious radiation from a feed, low power and poor polarisation purity. Thus to avoid these limitations researchers should concentrate on the design with the feed methods.

There are verity of feed methods are available for Microstrip antennas. These may be either Contacting or Non-Contacting i.e. power can be directly fed to the antenna by a contacting material like microstrip line or coaxial probe or EM coupling can be carried out to deliver the required power from the radiating patch. With these methods microstrip antenna can be of a single element or array of elements, to make the array to be an effective radiator, each patch in the array must be suitably fed [6]. The arrangement of feed can have single or multiple lines. If single feed line is used it is called to be series feed network and it is said as corporate feed if there are multiple feed lines which can account for phase change and eventually steer the antenna beam. [7].

Thus in this paper single element, two elements, and 4 elements patch antennas with line feed and corporate feeds are designed with three materials (FR-4, RT-Duroid, and Teflon) to work on the frequency 28 GHz which is suitable for mm wave applications. The analysis is done to find their Return Loss and Gain and is compared for best operational efficiency.

2 Microstrip Patch Antenna

Due to their inherent advantages of Low Cost and Low profile design methods and ease of fabrication, Microstrip patch antennas have widely been used and common within wireless circuit designs.

A generic structure of a Microstrip Patch Antenna (MPA) is show in Fig. 1. The Conductive patch, ground plane of the radiating patch is designed using high conductivity copper material. A dielectric material with permittivity (ϵ_r) is used and the radiating patch has a length (L) and Width (W). The height of the substrate (h) should be smaller than the wavelength of the input signal/but care should be taken that it doesn't become too small.

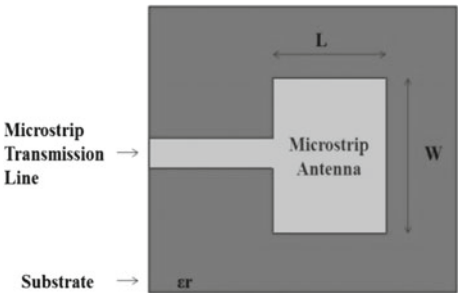


Fig. 1 Microstrip patch antenna

As per the Transmission line theory, the length of the patch antenna (L) is used to find the resonance frequency, the centre frequency is given as,

$$f_c = \frac{C}{2L\sqrt{\in_r}} \tag{1}$$

The substrate material thickness is designed as 1.5 mm, height 70 mm, width 60 mm. The dimensions of the antenna can be calculated by using the following relationship.

It is to be noted from Eq. 1 that the overall length of the patch should be just one half of the wavelength, in the dielectric substrate medium. The Input impedance is controlled by width W of the microstrip antenna.

3 Feed Techniques and Networks

Use of direct or indirect contact to excite the feed is generally employed. There are many different methods of feeding and four most popular methods are,

1. Microstrip line feed (Insert/Quarter Wavelength feed)
2. Coaxial probe
3. Proximity Coupling
4. Aperture Coupling

In a feed network arrangement of the microstrip array, elements can be fed by a single line or multiple lines [8, 9]. Feeding networks are classified as:

- Series feed network
- Corporate feed network
- Corporate-series feed network

A primitive form of a feed network is the Series Feed Network. All along the transmission line, a small amount of energy is coupled to the following radiating element and there is a flow of power from one end to the far end of the transmission network. One such Series feed line is shown in Fig. 2.

Fig. 2 Microstrip patch antenna with series feed network

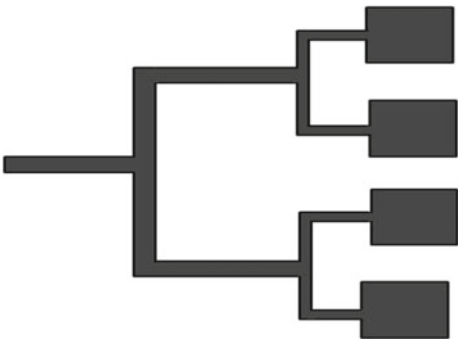


Fig. 3 Microstrip patch antenna with corporate feed network

A different approach, wherein the power can be equally split in between the radiating patch is given by Corporate Feed lines. The power distribution can be controlled by the choice of different power divider topologies. Figure 3 shows the corporate feed line method.

A much more improvised feeding technique is developed by combining both Series feed and Corporate feed to give an advantage of better performance. This feed technique is used in many antenna topologies [10] and various complex antenna networks where high gain requirement persists. References [11–13].

The characteristics, features, feed method and structure and the applications of each type of feed method are listed in the Table 1.

Table 1 Comparison of feed methods of microstrip patch antenna

Characteristics	Microstrip feed line	Coaxial feed line	Aperture couple feed	Proximity couple feed
Spurious radiation from feed	More	More	Less	Minimum
Reliability	Better	Poor	Good	Good
Ease of fabrication	Easy	Soldering and Drilling Needed	Alignment required	Alignment required
Impedance matching (Z)	Easy	Easy	Easy	Easy
Achieved bandwidth	2–5%	2–5%	21%	13%
Advantages	Simple and Easy to Fabricate, it can be eached on the same substrate	Feed can be placed at any desired direction, Low spurious radiation	Eliminates spurious radiation	Eliminates spurious radiation
Disadvantages	It gives undesired radiation	Narrow bandwidth and difficult to model	Fabrication Process is difficult	Thickness of antenna increases

4 Proposed Antenna Analysis

In this paper we designed single element, two elements and four elements microstrip antenna with line feed and corporate feed. And the analysis is done with different antenna materials such as FR-4, RT-Duroid, and Teflon.

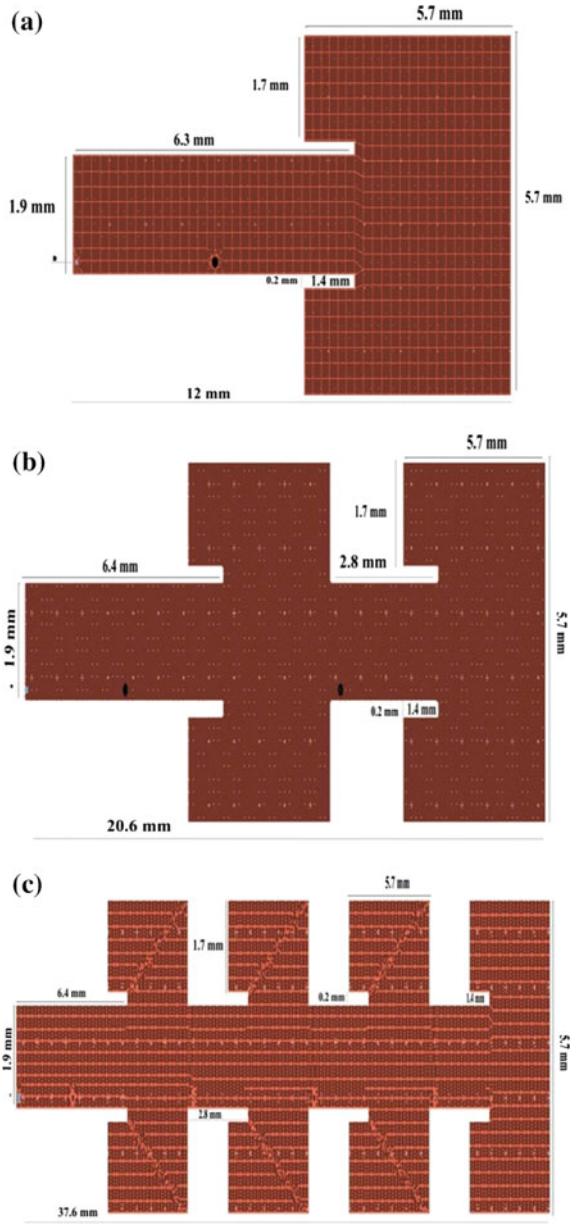


Fig. 4 **a** Single element FR-4 MPA, **b** Two elements FR-4 MPA, **c** Four elements FR-4 MPA

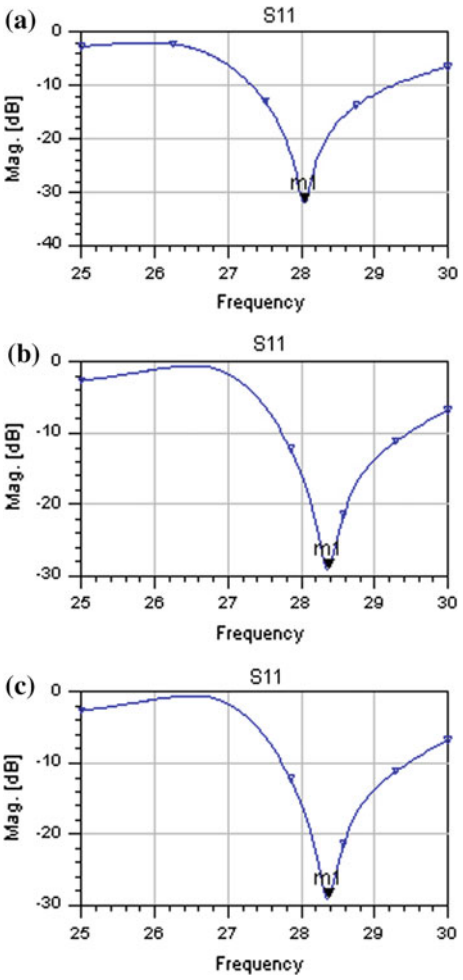


Fig. 5 **a** Return loss of single element FR-4 MPA, **b** Return loss of two elements FR-4 MPA, **c** Return loss of four elements FR-4 MPA

Figure. 4a, b, and c shows the single, two and four element patch antenna with Inset feed and made of FR-4 material using ADS 2009. The Reflection Loss was calculated as shown in Fig. 5a, b and c. The dimensions of the Patch is as follows, Length = 5.7 mm, Width = 5.7 mm, Feed Length = 6.3 mm, Feed Width = 1.9 mm.

Same analysis is done with single, two and four element patch antenna with Inset feed and made of Teflon substrate and given below. The dimensions of the Patch is as follows, Length = 4.7 mm, Width = 5.7 mm, Feed Length = 6.4 mm, Feed Width = 1.9 mm. The Fig. 6a, b and c antennas are designed using ADS and the Reflection Loss was calculated as shown in Fig. 7a, b and c.

Same is repeated for RT-Duroid Substrate with single, two and four element Microstrip patch with Inset Feed is given. Following are the dimensions of the

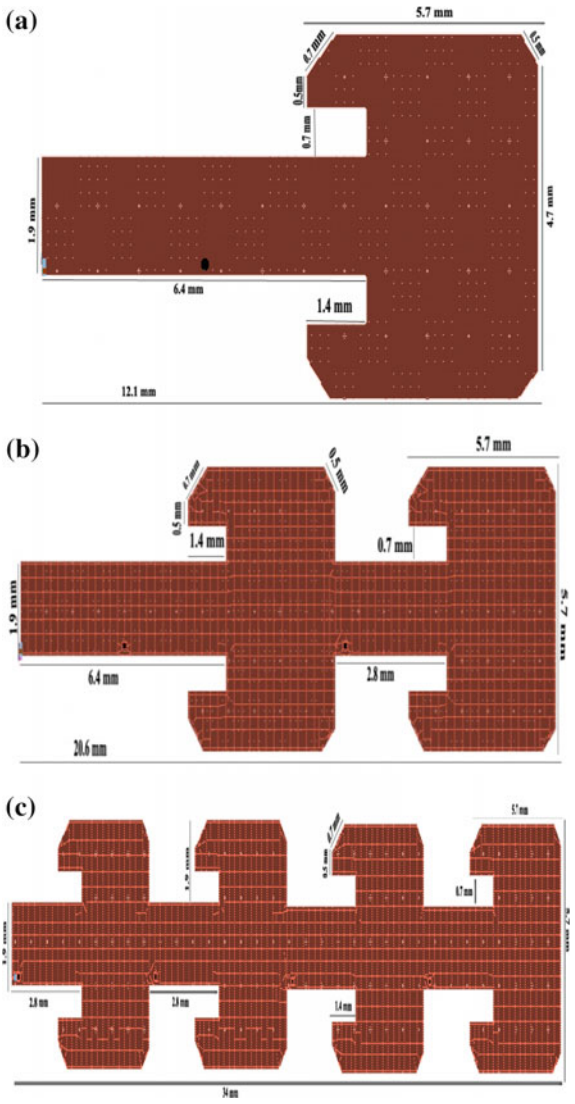


Fig. 6 **a** Single element Teflon MPA, **b** Two elements Teflon MPA, **c** Four elements Teflon MPA

patch, Length = 5.7 mm, Width = 5.7 mm, Feed Length = 6.45 mm, Feed Width = 1.9 mm. The Fig. 8a, b and c antennas are designed using ADS and the Reflection Loss was calculated as shown below in Fig. 9a, b and c.

The above analysis is repeated for Fr-4 Substrate with two four and eight element patch antenna with corporate feed network is given in Fig. 10a, b and c with their corresponding analysed Reflection Loss in Fig. 11a, b and c.

The dimensions of the Patch is as follows, Length = 5.7 mm, Width = 2.85 mm, Feed Length = 2.85 mm, Feed Width = 10.2 mm.

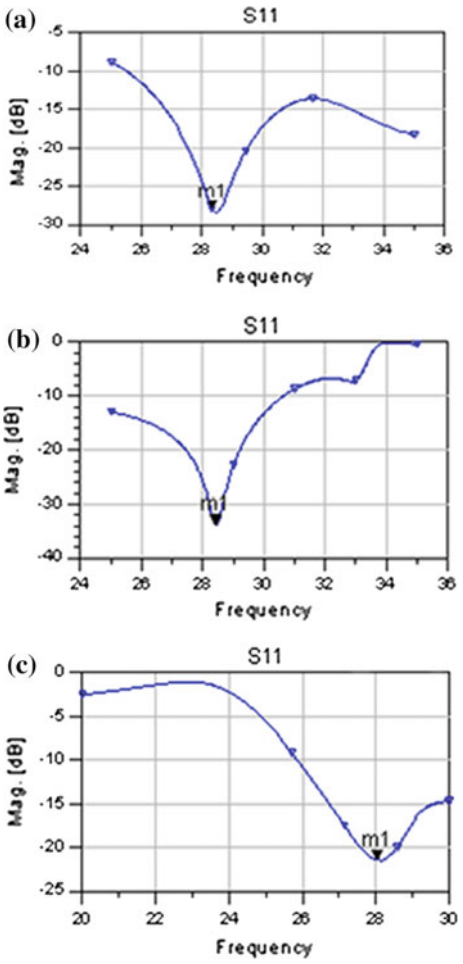


Fig. 7 **a** Return loss of single element Teflon MPA, **b** Return loss of two elements Teflon MPA, **c** Return loss of four elements Teflon MPA

The following tables give the analytical results of the above experiments including frequency of operation, Return Loss and Gain provided by the designed antennas.

From Table 2 we can infirm that, For Line Feed Four elements MPA's providing better gain than the single and two elements antenna. Also, the designed antennas are operating in the mm wave frequency (28 GHz).

From Table 3 it is seen that, In corporate Feed method Eight elements MPA's providing better gain than the two and four elements antenna. Also, the designed antennas are operating in the mm wave frequency (28 GHz). The designed antenna with single, two and four element Line feed (series fees network) using Fr-4, substrate is fabricated and shown in Fig. 12.

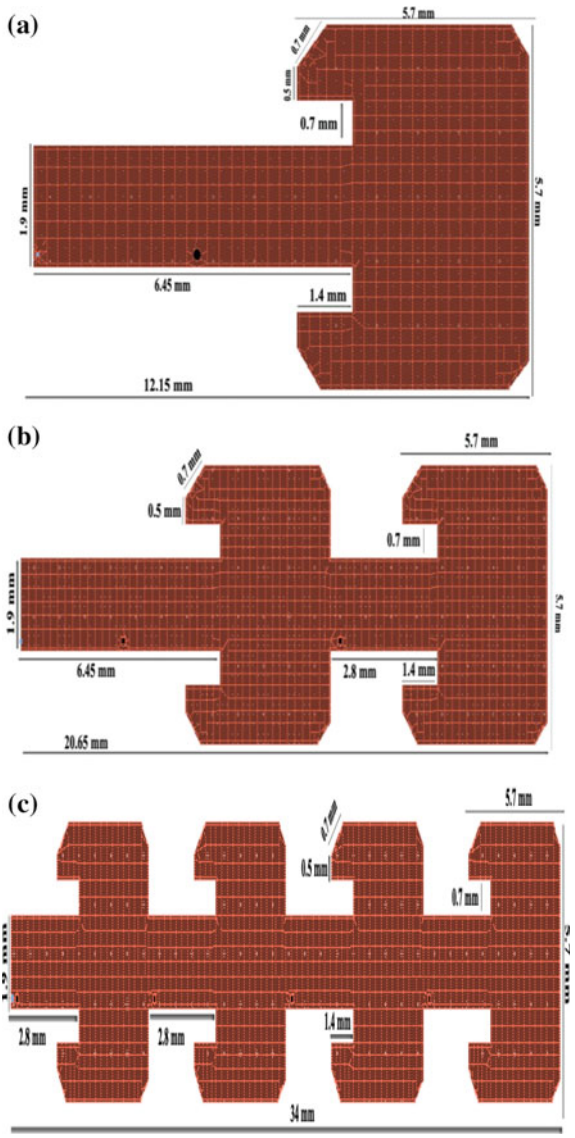


Fig. 8 **a** Single element RT-Duroid MPA, **b** Two elements RT-Duroid MPA, **c** Four elements RT-Duroid MPA

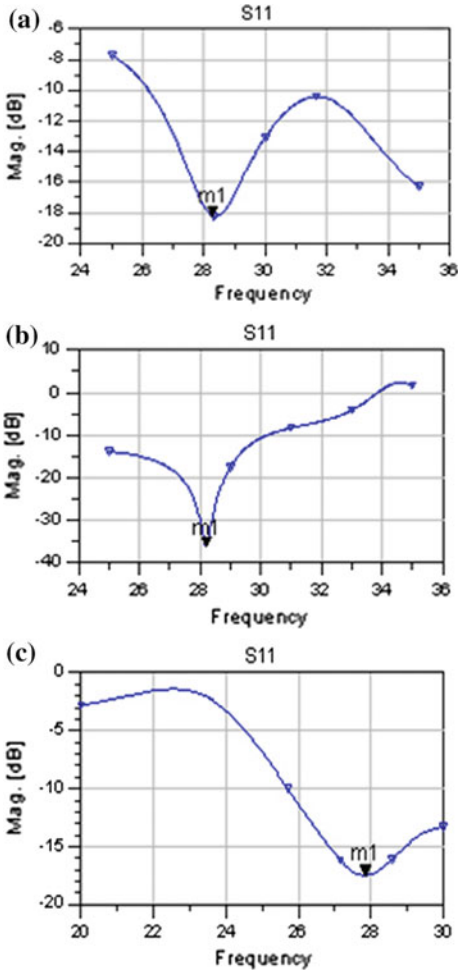


Fig. 9 **a** Return loss of single element RT-Duroid MPA, **b** Return loss of two elements RT-Duroid MPA, **c** Return loss of four elements RT-Duroid MPA

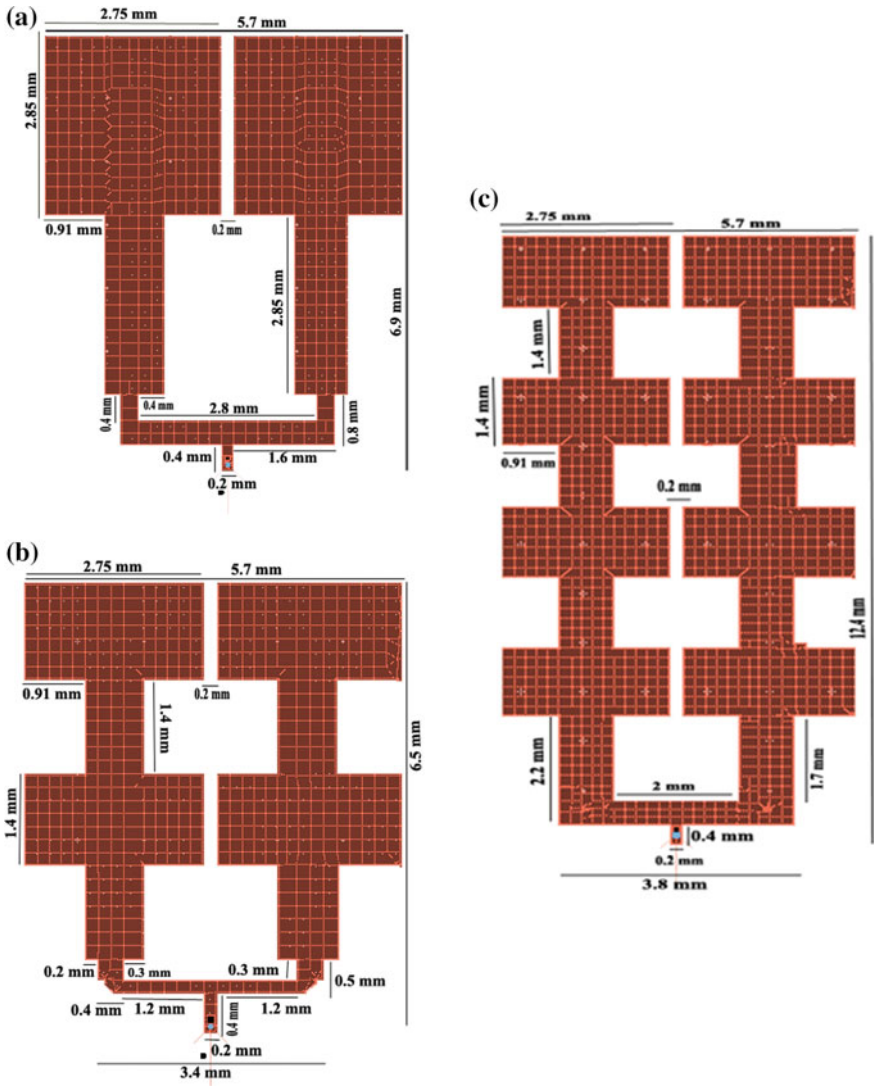


Fig. 10 **a** Two Elements FR-4 MPA with Corporate Feed, **b** Four elements FR-4 MPA with corporate feed, **c** Eight elements FR-4 MPA with corporate feed

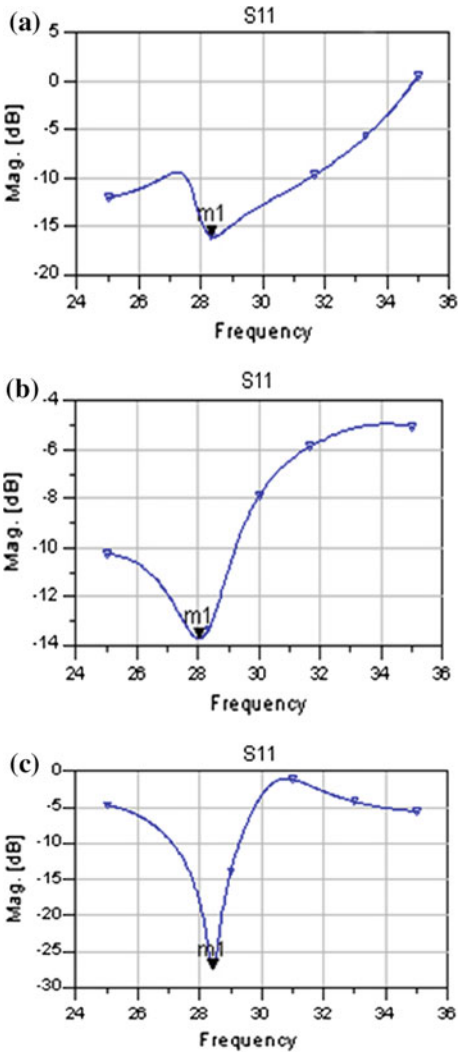


Fig. 11 **a** Return loss of two elements FR-4 MPA with corporate feed, **b** Return loss of four elements FR-4 MPA with Corporate Feed, **c** Return loss of eight elements FR-4 MPA with corporate feed

Table 2 Designed antenna parameters for Line feed (series feed network)

Antenna elements	Substrate	ε_r value	Frequency in Ghz	Return loss	Gain
1	FR-4	4.4	28.05	−32.04	8.67
2	FR-4	4.4	28.37	−28.99	8.71
4	FR-4	4.4	28.19	−22.452	9.33
1	Teflon	2.1	28.33	−22.97	6.87
2	Teflon	2.1	28.44	−33.78	6.80
4	Teflon	2.1	28.04	−21.39	9.04
1	RT-Duroid	2.2	28.28	−18.22	6.80
2	RT-Duroid	2.2	28.22	−36.43	7.16
4	RT-Duroid	2.2	27.86	−17.47	8.5

Table 3 Designed antenna parameters for corporate feed network

Antenna elements	Substrate	Value	Frequency in Ghz	Return loss	Gain
2	FR-4	4.4	28.33	−16.04	3.17
4	FR-4	4.4	28.05	−13.72	3.01
8	FR-4	4.4	28.44	−27.48	9.89

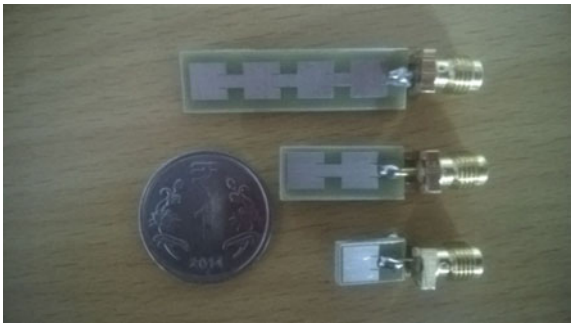


Fig. 12 Fabricated antenna using Fr-4 substrate

5 Conclusion

In this paper we focused on 28 GHz for 5G application with three different materials (Fr-4, RT-Duroid, and Teflon). Further research must take place in mm wave band and the characteristics of other frequencies needs to be studied, the penetration power and the range for communication needs to be further improved. The proposed antenna gives return loss values for the various substrate and hardware implemented for line feed method using FR-4 substrate. Assembly and antenna dimensions are important constraint in the fabrication process.

References

1. Thandaiah Prabu, R., Benisha, M., Thulasi Bai, V., Yokesh, V.: Millimeter Wave for 5G Mobile Communication Apllication. AEEICB, IEEE Explore Digital Library, pp. 236–241 (2016). ISBN: 978-1-4673-9745-2
2. Benisha, M., Thandaiah Prabu, T., Thulasi Bai, V.: Requirements and Challenges of 5G Cellular Systems. AEEICB, IEEE Explore Digital Library, pp. 252–255 (2016). ISBN: 978-1-4673-9745-2
3. Thandaiah Prabu, R., Ranjeetha, R., Thulasi Bai, V., Jayanandan, T.: Design of Alpha/ Numeric Antenna for WIFI Applications. ICT 2016, Springer AISC Series, vol. 542, pp 21–39 (2017). ISBN 978-981-10-3223-3
4. Kumar, P., Thakur, N., Sanghi, A.: Micro strip patch antenna for 2. 4 GHZ Wireless. Int. J. Eng. Trends Technol. **4**(8), 3544–3547 (2013)
5. Khraisat, Y.S.H.: Design of 4 element rectangular microstrip patch antenna with high gain for 2.4 GHz applications. Mod. Appl. Sci. **6**(1), 68–74 (2011)
6. Hyok Song, J., Marek Bailkowski, E.: Ku-Band 16x16 planar array with apperture-coupled microstrip patch elements. IEEE Antenna Propag. Mag. (USA) **40**, 25 (1998)
7. Abbaspour-Tamijani, A., Sarabandi, K.: An affordable millimeter wave bean steerable antenna using interleaved planar subarrays. IEEE Trans. Antennas Propag. **51**, 2193–2202 (2003)
8. Balanis, C.A.: Antenna Theory: Analysis and Design, 3rd edn. (2005)
9. Errif, H., Baghdad, A., Badri, A., Sabel, A.: Design and Simulation of Microstrip Patch antenna with high directivity for 10 GHz applications. In: International Symposium on Signal Image Video and Communications, Isive-2014, 19–21 November 2014, Marrakech, Morroco
10. Errrif, H., Baghdad, A., Badri, A.: Design and analysis of microstrip patch array antenna with serie, coporate and series-corporate feed network. Int. J. Electron. Electr. Eng. (IJEEE)
11. Skolnik, M.I.: Introduction to Radar Systems, 3rd edn. McGraw-Hill, New York (2000)
12. Mailoux, R., et al.: Microstrip Array Technology. Antenna and Propagation, IEEE Transaction **29**, 25–37 (1981)
13. Huque, Md.T.I., Hosain, Md.K., Islam, Md.S., Chowdhury, Md.A.: Design and performance analysis of microstrip array antennas with optimum parameters for X band applications. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **2**(4) (2011)

Multiple Output Off-Line Flyback Converter with a Single Switch

R. Rashmi and M. D. Uplane

Abstract This paper presents a multiple output flyback converter in high-voltage and low-power application with a single switch. Different voltage levels are required in many applications, and as the number of voltage levels increases, the number of regulators increases, which increases the cost and complexity. To overcome this cost and complexity multiple output flyback converter is designed using the UC3844 integrated pulse width modulation IC which drives the MOSFET as well as provides the load and line regulation in current control mode. Control characteristics and performance of a 28-W flyback converter are simulated in PSIM software. Steady-state response of the current loop control with variable load is reported. The experimental result demonstrates that the model can regulate the multi-output flyback with a single switch for the specified voltage and current load.

1 Introduction

With advancement in technology the demand of multiple voltage levels in low- to medium-power applications [1] such as television, computers, video cassette recorder, power over Ethernet (POE), motor drives, control circuit has increased. Multiple number of power supplies can fulfill this demand, but with increased cost, complexity and size. To decrease the size and cost, research on multiple output power flyback converter with multiple control loops is carried out, which suffers several drawbacks such as the requirement of an additional loop for each output which increases the number of components, cost and complexity [2]. In order to overcome the increased cost, high component count and complexity multiple output off-line flyback converter with a single switch is presented in this paper. Multiple output off-line flyback converter with a single switch also decreases failure rate with

R. Rashmi (✉) · M. D. Uplane
Department of Instrumentation Science, Savitribai Phule Pune University, Pune, India
e-mail: ruchirashmi@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_38

Table 1 PWM IC selection and recommended usage

PWM IC	UVLO start	Maximum duty cycle (%)	Power supply input	Application (circuit)
UC3842	16 V	<100	High (off-line)	Forward, buck, boost
UC3843	8.3 V	<100	Low (DC/DC)	Forward, buck, boost
UC3844	16 V	<50	High (off-line)	Flyback, forward, buck, boost
UC3845	8.3 V	<50	Low (DC/DC)	Flyback, forward, buck, boost

lower component count as each additional output requires only an additional transformer winding, rectifier diode and filter capacitor [3, 4]. A two-MOSFET control may give better efficiency and less voltage stress, but at higher cost and circuit complexity.

Generally, PWM ICs are used in power supplies due to their small size, high performance and simplicity. There are a number of ICs are available in the market with UC3843/UC3844/UC3845 series. But for this application of off-line flyback converter with maximum 50% duty cycle and undervoltage lockout (UVLO), UC3844 is appropriate [5]. Turn on and turn off of UC3844 are fixed at 16 V and 10 V, respectively, which is self-biased through the transformer. Different ICs and their applications [6] in different topologies with UVLO are shown in Table 1.

Flyback regulator consists of the basic blocks shown in Fig. 1. AC source is rectified and filtered, and then with the switching action the power is transferred from primary winding to the secondary winding of the transformer and stored in the output capacitor after getting filtered. Switch current is sensed and compared with the reference current, and PWM proportional to the current difference is generated to regulate the voltage. Output 1 is well regulated via feedback, while the other

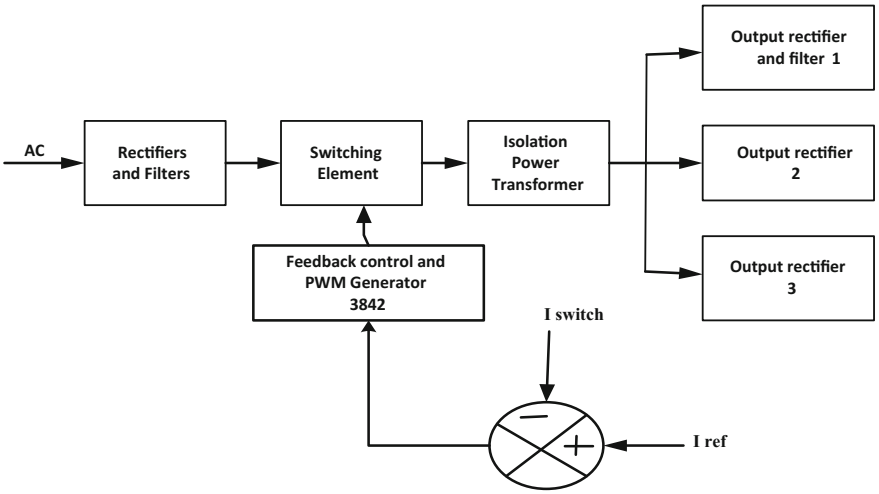


Fig. 1 Block diagram of multiple output flyback converter

outputs are controlled via the coupling of the secondary winding of the transformers [7]. Transformer coupling provides better line regulation as the control circuit instantaneously regulates the input voltage variation without using the error amplifier [8].

This paper is organized as follows: In Sect. 2 operation principal of the flyback converter with two modes is described. The multiple output flyback converter design is shown in Sect. 3. Experimental results are discussed in Sect. 4, and the conclusion is given in Sect. 5.

2 Operation of Multiple Output Flyback Converter

The functional equivalent circuit of the flyback converter is shown in Fig. 2, where Vin is AC source, R1 is rectifier, T1 is a coupled transformer, S1 is a MOSFET switch, D1 is Schottky diode, C1 is electrolytic capacitor, and R1 is load.

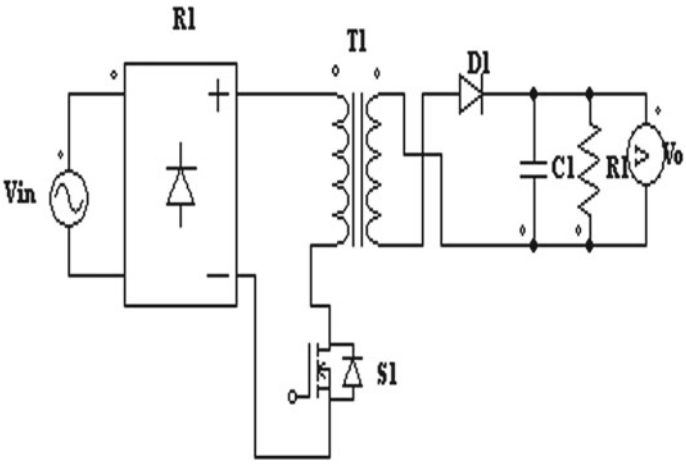


Fig. 2 Flyback converter

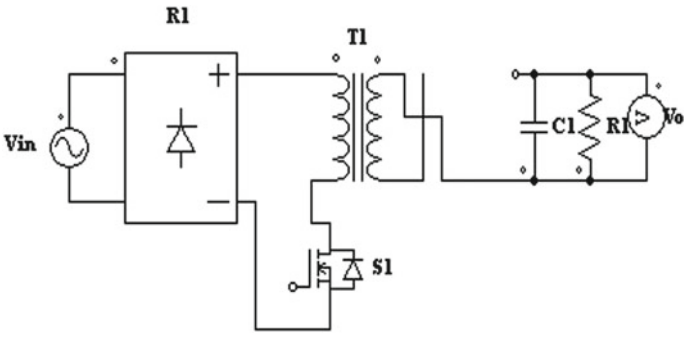


Fig. 3 Operational circuit in Mode I

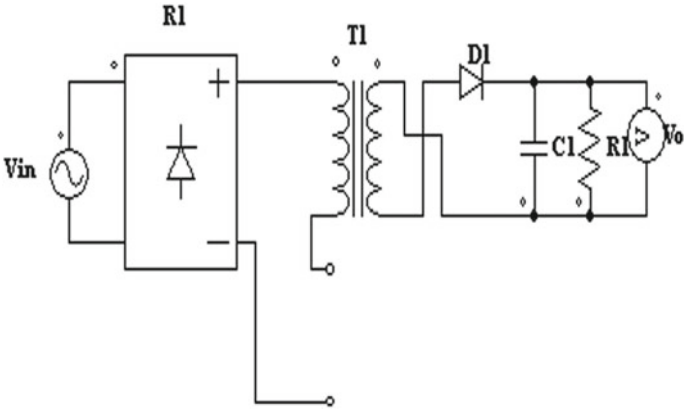


Fig. 4 Operational circuit in Mode II

The flyback converter normally operates in two modes [9]. The functional equivalent circuits of the two modes are shown in Figs. 3 and 4.

2.1 Operation During Switch ON Period of Switch 'S1' (Mode I)

When the switch 'S1' is ON, the source current flows from dotted end to non-dotted end in the primary winding of the transformer. But current does not flow in the secondary winding of the transformer because the diode 'D1' at the secondary winding gets reverse-biased as dotted end is on the higher potential. The output currents are supplied from storage capacitor C1.

2.2 Operation During OFF Period of Switch 'S1' (Mode II)

When the switch 'S1' is OFF, the voltage polarities across the winding reverse in primary as well as secondary windings. The diode on the secondary side of transformer is forward-biased. The current starts to flow in the secondary winding.

The two windings of the flyback transformer do not conduct simultaneously, but are magnetically coupled together so the induced voltage across the winding is proportional to the number of turns.

3 Design of Multiple Output Flyback Converter

3.1 Steady-State Model

Steady state of the converter is analyzed with ideal assumption of switches, inductor and capacitor. Volt-second balance implies that the inductor voltage in one period under steady-state condition should be zero as defined by (1)

$$V_{dc}DT_s + \frac{V0}{n}(1-D)T_s = 0, \tag{1}$$

where

$$V_{dc} = \frac{2Vin}{\pi}. \tag{2}$$

In (2), Vin is the peak value of the input voltage. On simplifying (1) and (2) equations we get output voltage as follows:

$$V_0 = \frac{-nV_{dc}D}{1-D}. \tag{3}$$

In (3), n is the ratio of secondary turns to primary turns of transformer and D is duty cycle. The slope of the inductor current waveform during the period DTs and (1-D) Ts under ideal condition is determined as follows:

$$\frac{di_L}{dt} = \frac{V_{dc}}{L_p} \tag{4}$$

$$\frac{di_L}{dt} = \frac{V_o}{L_s}. \tag{5}$$

In (4) and (5), L_p *and* L_s are the primary and secondary magnetizing inductances, respectively. Here D is chosen to be in the range from 0 to 0.5, which results in buck operation.

3.2 Design of Transformer

The flyback converter transformer performs dual function as both transformer and inductor. The regulation of the secondary voltages is designed through the single-phased five-winding transformer as shown in Table 2, to get characteristics such as

Table 2 Design parameters and their values

Design parameters values		Specification of transformer winding with the number of turns and diameter width		
		Windings	Turns	Diameter (SWG)
Vin	110 V AC	NP_1	74	23
Vo primary	5 V, 4 A	NP_2	13	30
Vo Aux1	−12 V, 0.3 A	NS_1	5	19
Vo Aux2	+12 V, 0.3A	NS_2	12	23
Switching frequency	40 kHz	NS_3	12	23
Po(max)	28 W	SWG—Standard Wire Gauge; core geometry—EE core		
Vin	220 V	Specification of snubber circuit		
Isolated/non-isolated	Isolated	R		4.7 kΩ, ½ W
Topology	Flyback-DCM	C		300 pF, 600 V

1. Lower peak switching transistor collector current
2. Lower winding loss
3. Lower ripple current of input capacitor.

Copper wire width is based on British Standard Wire Gauge (SWG) where with increasing gauge the weight per unit length diminishes by approximately 20% and the diameter diminishes by approximately 10.6%.

Transformer is designed in discontinuous mode where not all the energy stored in the transformer is transferred to the secondary before the switch on of the MOSFET. But in this mode some parameters are enhanced than in continuous conduction mode such as higher primary inductance of transformer, higher collector current, higher dissipation and larger volume of transformer.

The transformer drives in one direction of B-H curve. Since the current and flux never go negative, it may drive the core into saturation. To overcome saturation an air gap is introduced to flatten the hysteresis loop. In multiple output flyback converter only primary winding is regulated through PWM IC and the regulation of auxiliary winding depends tightly on the construction of the transformer.

3.3 Design of Snubber

It is designed to clamp the magnitude of the ringing voltage associated with the transformer leakage inductance within the peak voltage rating 300 V of the MOSFET. The RC time constant is equal to the current spike duration of 1.55 μs.

3.4 Regulation of Flyback Converter

A three-output off-line flyback converter is shown in Fig. 5 where Output 1 is primary and Output 2 and Output 3 are secondary as only reference current for

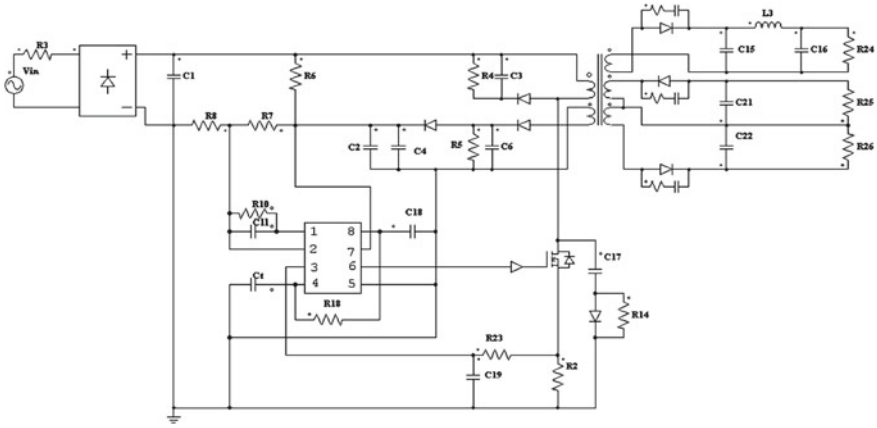


Fig. 5 Circuit diagram of multiple output flyback converter

Output 1 is connected to the feedback circuit, while Outputs 2 and 3 would only be controlled through Output 1 [10]. Outputs 2 and 3 are using same turns/volts as the Output 1 of the secondary winding. Outputs 1 and 3 are positive as the direction of winding for primary side and secondary side of the transformer is the same, while Output 2 is negative due to reverse winding of the secondary with respect to primary. Here the AC source Vin is rectified and filtered in first stage. Snubber circuit is used to provide the protection against large voltage spike during the start of turn off of switch 'S1'. A low-impedance path is provided through the snubber diode for the leakage inductance current of the primary winding. Biasing of the PWM controller IC is provided through R6 resister. The switching current is sensed on Pin 3.

4 Experimental Results and Discussion

Load regulation measurements are taken with one output swept from minimum load to maximum output current, while all other outputs delivered maximum output current. Cross-regulation measurements are also taken under the same conditions. In Fig. 5 only Output 1 is controlled via feedback, while the remaining outputs are set via transformer action.

Here Output 1 is designed to be regulated at 5 ± 5% V for 1–4 A, Output 2 at 12 ± 3%V for 0.1–0.3 A, and Output 3 at +12 ± 3% V for 0.1–0.3 A load with the maximum ripple voltage of 100 mV peak to peak. The regulator is providing a constant load current, and regulated voltage in underload condition is shown in Table 3.

Table 3 Variable parameters of Output 1, Output 2 and Output 3

Output 1		Output 2		Output 3	
Load voltage (volt)	Load current (amp)	Load voltage (volt)	Load current (amp)	Load voltage (volt)	Load current (amp)
Unregulated	≥4.00	Unregulated	≥0.3	Unregulated	≥0.3
4.96	4.00	−10.90	0.27	10.90	0.24
5.04	2.13	−11.82	0.19	11.83	0.16
5.08	1.45	−11.98	0.15	11.98	0.11
5.10	1.10	−12.06	0.12	12.07	0.09

Variation of the load voltage of Output 1 with the load is shown in Fig. 6, where Vo5_1, Vo5_2, Vo5_3, Vo5_4, Vo5_5 and Vo5_6 denote the load voltages at different load currents IL5_1, IL5_2, IL5_3, IL5_4 IL5_5 and IL5_6 as shown in Fig. 7. For the load current lower than 4A the regulator 1 is able to regulate the load voltage at 5 ± 5% V for primary Output 1 as shown in Table 1. Here in the overload condition Vo5_1 and IL_1 show a pulsed voltage and pulse current. In overload condition, the current sense comparator threshold of UC3844 is internally clamped to 1 V, which generates a constant PWM of small ON time. But this small ON time is not large enough to store the energy as required by the load. Even the PWM IC does not get proper bias voltage through the transformer, so PWM IC switches between the ON and OFF conditions and gives rise to pulsed voltage and current.

Variation of the load voltage of Output 2 with the load is shown in Fig. 8, where Vo12n_1, Vo12n_2, Vo12n_3, Vo12n_4 and Vo12n_5 denote the load voltage. At maximum load current of 0.3A the regulator is providing constant -12 V output. The variation of respective load current for Output 2 is denoted by IL12n_1, IL12n_2, IL12n_3, IL12n_4 and IL12n_5 shown in Fig. 9. The regulator is able to drive the load currents below 0.3A while maintaining the desired output voltage. But for the overload condition, Vo12n_1 and IL12n_1 show negative spikes and positive spikes, respectively, due to switching between the ON and OFF condition of the PWM IC to achieve −12 V and 0.4 A load.

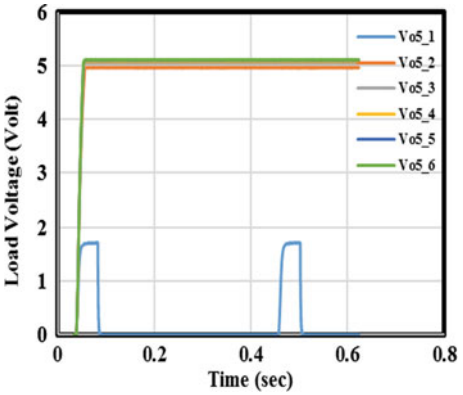


Fig. 6 Variation of load voltage of Output 1 with load

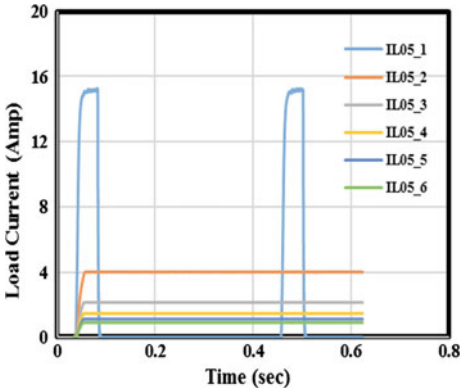


Fig. 7 Variation of load current of Output 1 with load

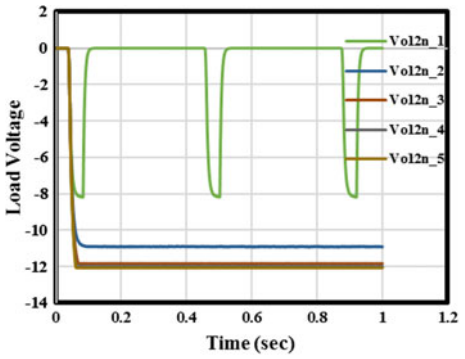


Fig. 8 Variation of load voltage of Output 2 with load

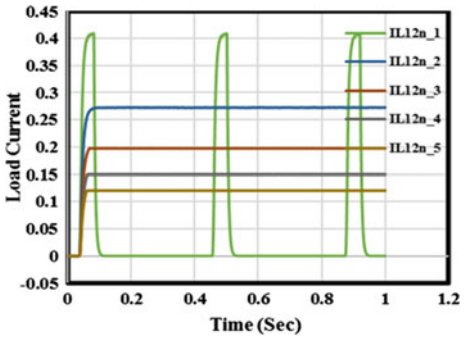


Fig. 9 Variation of load current of Output 2 with load

Similarly, for Output 3, we can see that the regulator is able to drive to load current below 0.3 A as shown in Fig. 10, where Vo12_1, Vo12_2, Vo12_3, Vo12_4 and Vo12_5 denote the load voltage. The regulator is giving constant $+12 \pm 3\%$ output voltage. The variation of load currents ILp12_1, ILp12_2, ILp12_3, ILp12_4 and ILp12_5 with the respective load is shown in Fig. 11. But for the overload condition, Vo12_1 and IL12_1 show positive spikes as transformer does not have enough power to drive the load.

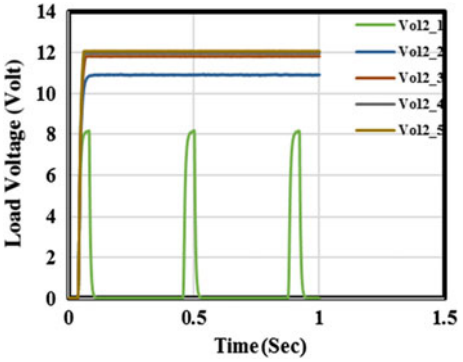


Fig. 10 Variation of load voltage of Output 3 with load

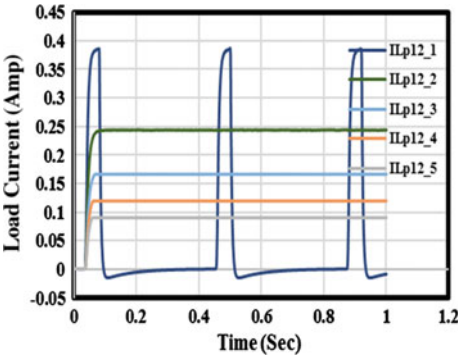


Fig. 11 Variation of load current of Output 3 with load

5 Conclusion

This paper has analyzed and presented a detailed design of simple and cost-effective multiple output flyback converters. Cross-regulation via the transformer winding provides the regulation of three different voltage levels with the elimination of additional switching supplies or post-regulators. In the future output capacitors can also be coupled to eliminate the ringing on the unloaded output.

Acknowledgements This work has been supported by DST-PURSE, Savitribai Phule Pune University. R. Rashmi is thankful to University Grants Commission for SRF.

References

1. Chryssis, C.: High frequency Switching Power Supplies: Theory and Design, 2nd edn. Mcgraw-Hill Publishing (1984)
2. Pressman, A.I.: Switching Power supply Design, 2nd edn. (2009)
3. Billing, K., Morey, T.: Switchmode Power Supply Handbook, 3rd edn. (2009)

4. Barbosa, P.M., Barbi, I.: A single-switch flyback-current-fed DC-DC converter. IEEE Trans. Power Electron. **13**(3), 466–475 (1998). https://doi.org/10.1109/63.668108
5. UNITRODE Application Note U-100A UC3842/3/4/5 "Provides low-cost current-mode" pp. 3-53–3-66
6. UC3844, UC3845, UC2844, UC2845: High performance current mode controllers Publication order number UC3844/D, ON Semiconductor
7. Ayachit, A., Reatti, A., Kazimierczuk, M.K.: Magnetising inductance of multiple-output flyback dc–dc convertor for discontinuous-conduction mode. IET Power Electron. **10**(4), 451–461 (2017). https://doi.org/10.1049/iet-pel.2016.0390
8. Unitrode application note U-100A "UC3842/4/4/5 provides low-cost current –mode control
9. Umanand, L.: Power Electronics Essentials and Application. Wiley India Pvt. Ltd., pp. 549–560 (2013)
10. Chalermyanont, K., Sangampai, P., Prasertsit, A., Thenmontri, S.: High frequency transformer designs for improving cross regulation in multiple-output Flyback converter, PEDS 2007, pp. 53–56

Predictive Analytic as a Service on Tax Evasion Using Feature Engineering Strategies

S. Kishore Babu and S. Vasavi

Abstract Predictive analytics can forecast trends and determines statistical probabilities and to act upon fraud and security threats for big data applications such as business trading, fraud detection, crime investigation, banking, insurance, enterprise security, government, healthcare, e-commerce, and telecommunications. Predictive analytics as a service (PAaaS) framework is proposed in our earlier work. One solution based upon ensemble model that uses Gaussian process with varying hyper-parameters is also given in our earlier works. Test results proved that the third hyper-parameter values yielded a good result with less error rate and more variance which is reliable for a predictive model. This paper presents solution based upon ensemble model that uses best out of prediction algorithms such as artificial neural networks (ANN), auto-regression algorithm (ARX) and Gaussian process (GP). Feature engineering methods such as recursive feature elimination that uses random forest algorithm is used for attribute selection. Performance measures NRMSE and COD are used to analyze the model. Test results proved that neural networks performed well when compared to regression and Gaussian process.

1 Introduction

As explained in [1] predictive models can find relationship between outcome and dependent variables. There are six phases for predictive analytics process. In the initial phase, project is defined with outcomes, objectives, scope, and the deliverables from the project. In the next phase, data is collected from various sources and is analyzed. This analysis requires strategies for preprocessing such as data cleaning, transformation, and data modeling so that useful data is extracted for further processing. Subsequently, validate the initial hypothesis using statistical

S. Kishore Babu (✉)
Andhra Loyola College of Engineering & Technology, Vijayawada, Andhra Pradesh, India
e-mail: skbjpj@gmail.com

S. Vasavi
VR Siddhartha Engineering College, Kanuru, Andhra Pradesh, India

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_39

models. The next phase is predictive modeling for forecasting the future. Results after implementation can be deployed for using it in the day-to-day decision-making. The last phase is monitoring the model in order to ensure that it is providing the expected results. This paper focuses on the performance of computing layer of our framework that is described in [11], i.e., which of the algorithms such as artificial neural networks (ANN), auto-regression algorithm (ARX), and Gaussian process (GP) is better for income tax dataset to identify fraud in projected tax values. The paper is organized as follows: Sect. 2 presents literature survey on various predictive analytics algorithms and various existing Web services. Section 3 outlines the proposed framework. Conclusions and future work are given in Sect. 4.

2 Literature Survey

Work reported in [12] explores the compliance and revenue consequences of the use of predictive analytics in an agent-based model that draws upon the behavioral approach to tax compliance. It is shown that the use of predictive analytics yields a significant increase in revenue over a random audit strategy by affecting the subjective belief and enhancing the social custom. Agent-based modeling for tax evasion is discussed in [13]. Credit scoring and econometric analysis methods from predictive analytics are used. Their analysis compared the outcome of predictive analytics based on tax return data with that of random audits. Predictive analytics (PA) and predictive risk modeling (PRM) models are reviewed such as Deloitte Consulting, the Public Consulting Group, the Case Commons Casebook model, and cutting-edge work in New Zealand in [3]. In an article published by economic times [2] stated that Indian government wants to use big data analytics tools in income tax department so as to differentiate black money holders from genuine taxpayers. In another article [4] reported that predictive analytics can search vast amount of tax data for detecting tax evasion. Authors of [7] explained how predictive data analytics saves lives and taxpayer dollars in New York City. In the work reported by [9] concluded that predictive analytics as a service should be dynamic and self-improving so that it will provide right data to the right person at the right time and place. Authors of [5] discussed feature engineering strategies for credit card fraud detection. Authors did not mention about features creation and their individual impact. A method is proposed in [10] to minimize the investment risk in stock market by predicting the returns of a stock using ensemble learning random forest algorithm. A case study on tax collection per month of the Federal Patrimony Department (SPU) is presented in [14]. But their method works for small samples only. ANN along with genetic algorithm is used to find credit card fraud detection in [15]. Deep learning, advanced machine learning algorithms are used for fraudulent detection in [16]. Authors of [8] have used regression for predicting behavior of stock market analysis.

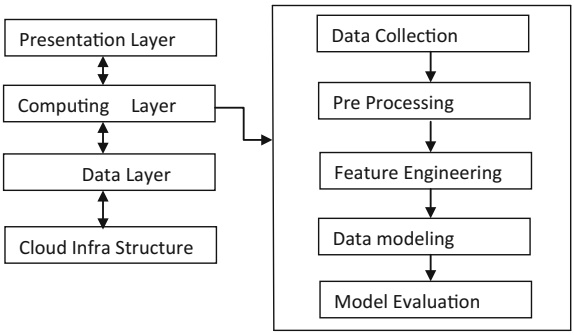


Fig. 1 Flowchart of the proposed system

3 Proposed System

Figure 1 presents the flowchart of the proposed system along with the complete details of computing layer. Presentation layer provides infrastructure as a service (IaaS) for creating virtual machine environment so as to load Hadoop Distributed File System (HDFS). End user queries are submitted to this layer. Computing layer provides predictive analytics algorithms as software as a service (SaaS) via RESTful query services. Data layer stores varieties data such as structured, unstructured, and semi-structured. Cloud infrastructure provides platform as a service (PaaS) for storage and computing services.

3.1 Preprocessing Using Z-Score Normalization

Data collected from variety of sources may have incomplete, noisy, and inconsistent data. Preprocessing helps for data smoothing and normalization. Our framework uses Z-score normalization for normalizing the data. Z-score normalization is calculated using Eqs. (1) and (2)

$$V' = (V - \text{Mean})/\text{StDev} \tag{1}$$

$$Z = (x - \mu)/\sigma) \tag{2}$$

where Z is Z-Score, X is the data element, μ is the mean, and σ is the standard deviation.

3.2 Feature Engineering Strategies

Most important step in fraud detection is to create a predictive model with useful and relevant features. Our framework uses **recursive feature elimination** (RFE) that uses random forest algorithm in each iteration to evaluate the model. For a training set of N tuples, sample selection with replacement is chosen to create training set and further the tree is grown. A number m < M (input variables) is chosen at random, and the best split on this m is used to split the node. This value of m is held constant while we grow the forest. Each tree is grown to the largest extent possible, and there is no pruning. Predict new data by aggregating the predictions of the n trees (i.e., majority votes for classification, average for regression) [17].

3.3 Building Artificial Neural Network

ANN using Feed-Forward propagation is constructed using a set of weights (based on Gaussian distribution) to the input neurons (initial weights will be between 0 and 1) and calculate output neuron value. A**ctivation function** (Sigmoid suits for our data. Tangent function is for complicated data) is applied to the hidden layer neurons to get the final value. Activation function in neural networks will transform the input signal into an output signal and models complex nonlinear patterns. There are many types of activation functions—linear, sigmoid, and hyperbolic tangent, even stepwise.

3.4 Regression

Jackknife Estimation and Cross-Validation methods are used for correcting bias that uses sampling without replacement. Jackknife method is useful in situation where there is no theoretical basis to fall back on. Jackknife is a "leave-one-out" notion in regression, where it follows leaving out each one sample at each iteration and finally calculating the average of these calculations as follows [20]:

Let dataset has samples from $x_1, x_2, \ldots x_n$.

At each iteration, we generate n samples of size $n - 1$ by leaving out one observation at a time say it as $x = \{x_1 \ldots . x_n\}$.

Now compute bias function: $\theta(x)$

Repeat this process n times. From these bias functions, calculate: θ_{-i} by applying the estimation process to the Jackknife sample. Calculate the Jackknife estimate as shown in Eq. (3)

$$\hat{\theta}_* = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i} \tag{3}$$

Jackknife estimate of variance as shown in Eq. (4)

$$\frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{-i}-\hat{\theta}_*\right)^2 \tag{4}$$

Calculate S^2 is an unbiased estimator for σ^2 as per Eq. (5)

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\hat{X})^2 \tag{5}$$

Cross-validation is a method for evaluating the predictive error when we fit a function relating two or more variables. Instead of doing "leave-one-out" cross-validation (LOOCV), it does "leave-several-out." In this framework, we use two cross-validation methods, namely train/test (twofold cross-validation) and Monte Carlo approaches for prediction. Train–validation–test procedure for is used for neural network. The training set (80% of the samples) is the input data used to train the model, and the validation set specifies the estimate model properties such as prediction error for model selection, the test set is used for a calculating the error rate RMSE of the final chosen model. During second time of execution, we got different RMSE value because the samples were split randomly into the train–validation–test sets. Hence, we used Monte Carlo cross-validation approach that performs train–validation–test a number of times (n300 times during comparing three algorithms and 1000 times when comparing regression and neural networks) finally averages RMSE from each of the iteration.

4 Results and Analysis

4.1 Dataset Used

State Government Tax Collections historical data available at https://www.census.gov/govs/statetax/historical_data.html is taken as input dataset. Figure 2 presents attribute list. Data from the Annual Survey of State Government Tax Collections are available online for each year from 1992 to the present. A historical dataset for years 1951 to present is available in Excel format. This paper presents the results of the analysis for the subset of dataset that comprises data of all the states and for the years 1994 to 2014. RStudio software that provides a platform for statistical computing and handling large datasets is used for implementing the algorithm.

- Year
- State
- Name
- FY Ending date
- Total Taxes
- Property Tax
- Tot Sales & Gr Rec
- Tax Total Gen Sales Tax
- Total Select Sales Tax
- Alcoholic Beverage Tax
- Amusement Tax
- Insurance Premium Tax
- Motor Fuels Tax
- Parimutuels Tax
- Public Utility Tax
- Tobacco Tax
- Other Select Sales Tax
- Total License Taxes
- Alcoholic Beverage Lic
- Amusement License
- Corporation License
- Hunt and Fish License
- Motor Veh & Oper Lic
- Motor Vehicle License
- Motor Veh Oper License
- Public Utility License
- Occup and Bus Lic NEC
- Other License Taxes
- Total Income Taxes
- Individual Income Tax
- Corp Net Income Tax
- Total Other Taxes
- Death and Gift Tax
- Docum and Stock Tr Tax
- Severance Tax
- Taxes NEC

Fig. 2 Attribute list

4.2 Performance Measures

Normalized Root Mean Square Error (**NRMSE**) can be calculated using Eq. (6)

$$NRMSE = \sqrt{\frac{1}{N}\frac{\sum_i (d_i - y_i)}{\sum_i (d_i)^2}} \tag{6}$$

Coefficient of determination (COD) can be calculated using Eq. (7).

$$COD = R^2 = 1 - \frac{\sum_i (x_i - y_i)}{\sum_i x_i - \overline{x}_i} \tag{7}$$

After implementing recursive feature elimination algorithm on the input dataset, the best 22 features that are dependent on property tax are obtained out of which the first best attribute is "motor.fuels.tax".

4.3 Jackknife Estimation

These 22 features shown in Fig. 3 are now taken as the input dataset for neural networks and regression algorithms to predict the "Property tax" of all states, and we take the best feature obtained as a result of feature engineering techniques. The predicted value in Jackknife estimation after correcting the bias is 0.041. Root mean square error (RMSE) for c**ross-validation** method of regression is 1.01.

```
The top 5 variables (out of 22):
   motor.fuels.tax, others.select.sales.tax, alcoholic.beverage.tax, public.utility.license, corporation.lice
nse

> predictors(results)
 [1] "motor.fuels.tax"          "others.select.sales.tax" "alcoholic.beverage.tax"
 [4] "public.utility.license"   "corporation.license"     "Total.select.sales.tax"
 [7] "occur.and.bus.lic.NEC"    "total.license.tax"       "public.utility.tax"
[10] "Year"                     "insurance.premium.tax"   "parimutuels.tax"
[13] "corp.net.income.tax"      "alcoholic.beverage.lic"  "Total.gen.sales.Tax"
[16] "motor.vet.Oper.license"   "Tot.sales.Gr.Rec.Tax"    "service.tax"
[19] "tobacco.tax"              "individual.income.tax"   "motor.vehicle.license"
[22] "motor.veh..oper.lic"
> plot(results, type = c("g","o"))
```

Fig. 3 Feature selection result

4.4 Fitting the Neural Network

To predict the "Property tax" of all the states, we took the best feature obtained as a result of feature engineering techniques. Error is calculated using sum of squared error (SSE) as given in Eq. (8):

$$\mathrm{SSE} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 \tag{8}$$

Neural network result for the given test data is −0.99999999.

The predicted value of neural network and its unscaled values is stored for further processing. After performing Monte Carlo **cross-validation** method for the above dataset, the result is 0.172. Table 1 compares NRMSE and COD for all the three algorithms.

4.5 Feature Selection

Second case study: For the comparison between neural networks and regression, we have taken the input dataset, where total taxes are the predicted variable and applied feature selection to find the dependent variables of total taxes. After applying recursive feature elimination, 25 features were selected as the predictors for total taxes column and first best attribute is the "Individual.Income.tax". Now upon performing **cross-validation** method for the above dataset, the result represents the RMSE value of total taxes dataset and best 10 attributes as a result of feature engineering technique. The error value thus obtained is 0.079.

Table 1 NRMSE and COD values for various algorithms

S. no	Process	NRMSE	COD
1.	Gaussian Process	3.8	0.90
2.	Neural Network	0.172	0.98
3.	Regression (ARX)	1.01	0.96

Table 2 NRMSE and COD values for various algorithms

S. no	Process	NRMSE	COD
1.	Neural network with z-score normalization	1.21	0.92
2.	Neural network with Min–Max normalization	0.0113	0.99
3.	Regression (ARX)	0.0792	0.94

4.6 Fitting Neural Network

Now for the total taxes as a prediction value, we considered remaining columns as input and trained our neural network. Authors of [18] stated that convergence is usually faster if the average of each input variable over the training set is close to zero. In our dataset, inputs were all positive, according to the authors of [6], any shift away from zero may cause biasing the updates in a particular direction and may slow down learning. Hence, we followed [6] in shifting the input so that average over the training set is close to zero. Because of this, we used two kinds of normalization approaches Z-score and Min–Max to understand which normalization gives best results for given dataset.

Z-score normalization scales the data between [−1, 1] and Min–Max normalization scales the data between [0, 1]. The trained neural network with 10 neurons in input layer (10, 5, 2) neurons in three hidden layers and 1 at the output layer. For constructing this neural network, we followed the thumb rule mentioned in [19]: The number of hidden layer neurons should be less than twice of the number of neurons in input layer, for determining the number of neurons in the hidden layer.

After applying cross-validation, we found that for given dataset Min–Max normalization is suited best as the error value is less and accuracy is much better as shown in Table 2.

Based and the results obtained and we can conclude that feed-forward back-propagation neural network proved to be better in terms of accuracy than regression because ANN is adaptive systems and is proved to best in constantly changing environments such in our example income tax changes year to year. Also, it is observed that neural network is taking longer time for training than regression.

5 Conclusions and Future Work

This paper presented the work where we build a predictive model that forecasts the tax values using the three algorithms in first case study and ANN, regression in second case study. For predicting the values more accurately, we have applied feature selection method to obtain the best attributes that lay the strong foundation for felicitating the process of production. Later using these, we applied Gaussian process, neural networks, and regression algorithms. We have predicted the values and also applied cross-validation method to find the error rate. The error rate

obtained for Gaussian process 3.8, neural networks is 0.172 and for regression is 1.01. Hence based on the above results, we can conclude that neural network is the best approach among the 3 for prediction which was followed later by regression and Gaussian process, respectively. Using feed-forward helped in reducing the number of false positives and false negatives. Our future work concentrates on exposing our work as a service for finding fraud in tax evasion so as to make it as the immediate framework for fraud detection. This service should be applied on tax return data as and when they are filed instead of waiting and analyzing years later.

References

1. Buytendijk, F., Trepanier, L.: Predictive Analytics: Bringing the Tools to the Data. An Oracle White Paper (2010)
2. http://cio.economictimes.indiatimes.com/news/business-analytics/income-tax-department-to-use-analytics-to-look-for-discrepancies-in-bank-accounts/55849827. Accessed 10 Dec 2016
3. Packard, Dr.T.: SACHS Literature Review: Predictive Analytics in Human Services. Sandiego State University School of Social Work (2016)
4. https://www.vertexinc.com/blog/tax-matters/learning-lab-2-big-data-and-predictive-analytics. Accessed 14 May 2017
5. Bahnsen, A.C., Aouada, D., Stojanovic, A., Ottersten, B.: Feature engineering strategies for credit card fraud detection. Expert Syst. Appl., 134–142 (2016)
6. LeCun, Y.A., Bottou, L., Orr, G.B., Miller, K.-R.: Neural Netowrks: Tricks of the trade: Efficient BackProp, 2nd edn. LNCS, vol. 7700 p. 16. Springer (2012)
7. Howard, A.: https://www.oreilly.com/ideas/predictive-data-analytics-big-data-nyc. 26 June 2012
8. Gharehchopogh, F.S., Bonab, T.H., Khaze, S.R.: A linear regression approach to prediction of stock market trading volume: a case study. IJMVSC **4**(3) (2013)
9. Predictive Analytics as a service for IoT. http://www.opengardensblog.futuretext.com/archives/2014/10/predictive-analytics-as-a-service-for-iot.html. Accessed 21 Oct 2014
10. Khaidem, L.,Saha, S., Roy Dey, S.: Predicting the direction of stock market prices using random forest. Appl. Math. Finan. (2016)
11. Kishore Babu, S., Vasavi, S., Nagarjuna, K.: Framework for Predictive Analytics as a Service Using Ensemble Model. IEEE IACC 2017 (in press)
12. Hashimzade, N., Myles, G.D., Rablen, M.D.: Predictive Analytics and the Targeting of Audits (2014). https://tarc.exeter.ac.uk/media/universityofexeter/businessschool/documents/centres/tarc/publications/discussionpapers/Predictive_28-10-14.pdf
13. Myles, G.D., Hashimzade, N., Page, F., Rablen, M.: Targeting Audits Using Predictive Analytics (2013). https://tarc.exeter.ac.uk/media/universityofexeter/businessschool/documents/centres/tarc/publications/Targeting_Audits_Using_Predictive_Analytics.pdf
14. Serrano, A.M.R., da Costa, J.P.C.L., Cardonha, C.H., Fernandes, A.A., de Sousa, R.T.: Neural Network Predictor for Fraud Detection: A Study Case for the Federal Patrimony Department, pp. 61–66 (2012). icofcs.org
15. Patidar, R., Sharma, L.: Credit Card fraud detection using neural network. Int. J. Soft Comput. Eng. (IJSCE) **1**(NCAI2011), 32–38 (2011). ISSN: 2231-2307
16. Anupriya, K., Kanimozhi, C.: Predicting eshopping data using deep learning middle-East. J. Sci. Res. **24**(S1), 250–256 (2016)
17. Team, A.V.C., Kaushik, S., Jaju, S., Gupta, A.: A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python), Analytics Vidhya, 31 Jan 2017

18. Lecun, Y., Bottou, L., Robert Mullar, K.: Efficient BackProp. Image Processing Research Department AT & T Labs - Research
19. Karsoliya, S.: Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. Int. J. Eng. Trends Technol. **3**(6) (2012)
20. course1.winona.edu/cmalone/stat360/…/Handout18.docx. Accessed March 2017

An Efficient Integrated ERP System Using Multilayer Perceptron

Ch. Rupa, J. Ranga Rao and P. Raveendra Babu

Abstract Maintenance of data flow from one department to another is difficult in any organization currently, because of rapid growth in data. In order to maintain the data and to provide better services and enhance the organization's revenue, enterprise resource planning system has been introduced. As the organization contains massive amount of data and to reduce up-front investment, secure cloud-based ERP systems have been used prominently. Apart from securely maintaining the data, goal of any organization is to increase their revenue and services to the customers. This paper proposes a novel method using artificial neural network multilayer perceptron to forecast the future to increase human resources, machinery and inventory in the organization. It has tested on clinical data repository and analysed the system performance by considering standard measures like RAE, MAPE, etc.

1 Introduction

Currently, utilization of data has been increasing day by day due to raising utilization of technologies by the people. It is directly proportional to the maintenance of data. Data transmission from one department to another department is becoming a difficult task in the enterprise environment. It affects mainly interconnected multidepartmental enterprises. One more issue is that secure data transmission may cause raise in the cost factor in enterprise resource planning (ERP). In cloud-based ERP, distributed departmental system helps to organize and maintain the data well by following appropriate architecture [1].

One of the drawbacks of traditional ERP systems is high up-front investment to maintain the database and security [2]. As all the organizational departments' data

Ch. Rupa (✉) · J. R. Rao · P. Raveendra Babu
Department of Computer Science and Engineering,
V. R. Siddhartha Engineering College (A), Vijayawada, India
e-mail: rupamtech@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_40

should be maintained, the database management becomes a risk factor for the organization. To overcome these issues, many of the enterprises are looking toward cloud-based ERP systems [3], where the service provider hosts the application in the cloud, so it can be accessed by multiple organizations. In cloud-based ERP system, the data maintenance and data security are handled by the cloud service provider itself. By this cloud-based ERP system, organizations have a control over their data that flows from various departments and provides integrated solution to their customers as a single database is used by all the departments of the organization.

By using multilayer perceptron cloud ERP system, any organization can enhance the services to their customers. But in long term, the goal of any organization is to enhance their profits and customer satisfaction. To achieve this, organization should predict their future needs and goals. To do this, multilayer perceptron is useful to forecast the future so that based on these results organizations can increase their human resources, machinery, and inventory to provide better services to their customers and enhance their profits.

In this work, forecasting has performed on clinical data. As healthcare centers are one of the organizations that are mostly using this ERP software [4] to control the flow of data between various departments like pharmacy, finance, and human resources. Software as a service (SaaS) feature can be achieved here by integrating the ERP system with ANN technique. The clinical data has acquired from a healthcare center which implemented cloud ERP in their organization. So, by using this clinical data, results are going to be displayed to analyze number of patients to be admitted in succeeding year based on the previous year data. Due to this analysis, healthcare centers can easily understand the organizational needs to serve the patients and increase their revenue in succeeding year by increasing the human resources, machinery, and inventory. The main objectives of the system are

- To maintain the ERP data in secure cloud environment.
- To implement multilayer perceptron-based data forecasting system.
- Verify the proposed method by comparing the various time series forecasting methods using some standard metrics.

2 Related Work

There are several forecasting methods like linear regression, SM Oreg, Gaussian processes, but they have several disadvantages. Back propagation works faster than earlier existing works for knowing about how the activations in one layer relate to activations in the earlier stage [5]. Back propagation artificial neural network (BP-ANN) is used as a classification tool by Aman Gautam [6] for segregating the data into abnormal or normal.

In particular, the most used forecasting method is linear regression which is proposed by Gupta [4]. This regression modeling technique uses data mining in

order to predict the future. This paper contains of formulation, testing data, and implementation of linear regression on the given data. But there several limitations for linear regression which are as follows:

(1) Linear regression is meant only for linear relationships between variables. (2) It is not accurate to find the relationship between variables if data is nonlinear. (3) Accuracy of the results of linear regression is very low.

Ramchoun proposed multilayer perceptron [3]. Architecture optimization and training gives a clear view on working of multilayer perceptron and back propagation, learning of MLP, and its implementation. There are several advantages of MLP when compared to other forecasting methods which are as follows: (1) Directional accuracy is very high for MLP when compared to linear regression, SM Oreg, and Gaussian processes. (2) Mean absolute error indicates the difference between exact value and predicted value.

As multilayer perceptron comes under artificial neural networks, we should have a clear study of it. Litta et al. [7] proposed artificial neural network in prediction of meteorological parameters during premonsoon thunders which gives a clear view about the working of artificial neural networks in forecasting the future. It is very powerful in machine learning, and application of these neural networks predicts patterns underlying the data better than other forecasting methods and gives accurate results.

In our analysis, we have done comparison on several forecasting methods and found that the accuracy of linear regression is very low when compared to multilayer perceptron, SM Oreg, and Gaussian processes [2, 5, 6]. And even the mean absolute error is very high which explains the difference between exact value and predicted value. Linear regression is having very high mean absolute error and low accuracy rate compared to other forecasting methods.

3 Proposed System

In the proposed system, we have to retrieve clinical data from the healthcare center as cloud-based ERP systems are being implemented, and data can be accessed through secured log-ins only [6]. Clinical data is retrieved from the healthcare centers itself. After retrieving the previous year's data, we need to calculate the patients' count of every year admitted in the care center due to a particular disease. After retrieving these results, multilayer perceptron method is used to forecast the number of patients to be admitted in the succeeding year due to a particular disease based on the previous year's data. Based on these results, the healthcare centers can increase their resources like machinery, inventory, and human resources to provide better services and enhance their profits. The proposed system process is shown in Fig. 1.

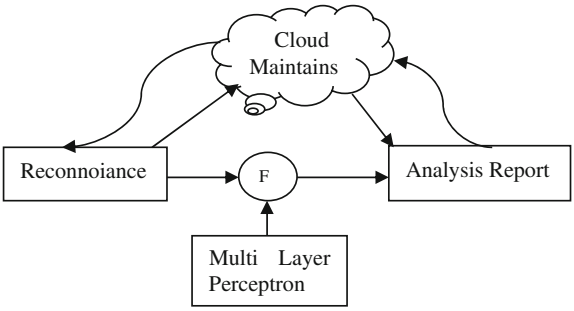


Fig. 1 Process of the proposed system

3.1 Reconnaissance Clinical Data Collection

As the healthcare center is adopting cloud-based ERP system, data retrieval is done from the cloud [7, 8]. As cloud access is restricted only to the authorized users, data is collected only from healthcare centers. After the collection of data, we need to perform some calculations on it in order to retrieve number of patients admitted due to a particular disease in the previous years. Table 1 shows the clinical data collected from cloud enterprise.

3.2 Time Series Forecasting by Multilayer Perceptron

Time series forecasting is used to analyze the previous data to predict the future data [9]. It is very much significant for organizations like healthcare centers to provide better services. To achieve accurate results, it is better to collect previous year's data in massive amount so that patterns can be easily evaluated. Artificial neural network is given importance in the work as it has a capability to find hidden patterns from the given input data and retrieve the output even the input contains noisy data as shown in Fig. 2.

Table 1 Clinical data obtained from healthcare centers

S. No	Year	Malaria	Dengue	Flu
1	2010	250	650	750
2	2011	410	450	468
3	2012	524	120	125
4	2013	530	820	460
5	2014	640	750	840
6	2015	850	640	654
7	2016	120	450	250
8	2017	450	320	480

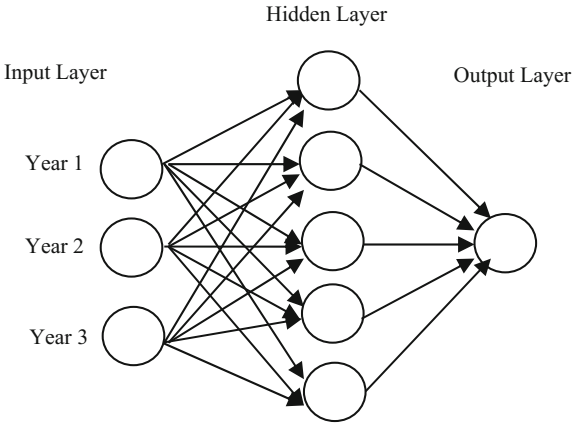


Fig. 2 Data classification by multilayer perceptron

In the above-mentioned figure, input layer contains reporting period of data flow among all the departments of an organization, hidden layer contains of patterns evolved from input data, and output layer contains forecasted data [10, 11]. Multilayer perceptron contains of three layers: One is input layer, second is hidden layer, and third is output layer. Initially, the input data is sent to input layer, and the intermediate output that is generated is given to the output layer which processes and generates the output.

4 Methodology

Multilayer perceptron is used to forecast the future of organization needs. Before implementing this method, work has been done on testing other forecasting methods, and based on comparison of the standard metric results, work has further gone through this method. After comparing several methods, result is acquired from the best method based on the following metrics which are calculated as follows.

4.1 Algorithm

To forecast the future by the past data, here used is artificial neural network (ANN)-related multilayer perceptron approach [12]. In this approach, first the input is divided into certain neurons which consist of some hidden neurons. These hidden neurons are intermediate results which are generated between input and output, and prediction function also can be performed on this until applied on all the neurons of the network. At the end, a single output neuron has been generated as an output which can be considered as predicted value by the organization for maintaining

their resources. The below steps are used to predict the value by using multilayer perceptron.

Requirements: pattern ∼x, MLP, enumeration of all neurons in topological order.
Result: Calculate output of MLP
Step 1: for all input neurons 'i' in network do

1.1: set $a_i \leftarrow z_i$
1.2: end the procedure if all the neurons in network are completed

Step 2: for all the hidden, output neurons 'i' in topological order do

2.1: set net $i \leftarrow w_i 0 + p_j \in Pred(i) w_{ij} a_j$
2.2: set $a_i \leftarrow flag(neti)$
2.3: end for

Step 3: for all output neurons i in network do

3.1: assemble a_i in output vector ∼y
3.2: end for

Step 4: return ∼y.

4.2 Standard Metrics

By calculating the standard metrics like mean absolute error, relative absolute error, mean absolute percentage error, root-mean-squared error, and mean-squared error, we can evaluate the best method for forecasting and get the better results of the forecasting method [6] in order to increase the revenue of the healthcare centers and to provide better services to the patients by increasing the resources in various departments of the enterprise like inventory, human resources, and machinery. References [13, 14].

1. Mean absolute error
   It is a measure to predict the accuracy of forecasting method in statistics like trend estimation. These are like trend estimate in statistics. It should be used to know the difference between two consecutive values.

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t|$$

2. Relative absolute error
   It is simple predictor which averages the actual values.

$$RAE = \frac{\sum_{i=1}^{n} |p_{ji} - a_i|}{\sum_{i=1}^{n} |a_i - (\frac{1}{n}\sum_{i=1}^{n} a_i)|}$$

where 'P_{ji}' is the predicted value to the individual case 'j', and 'a_i' is the target value.

3. Mean absolute percentage error
   It is used to predict accuracy of the forecasting method. It is a measure which can be used in statistics.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| e_{t/y_t} \right|$$

4. Root-mean-squared error
   It is a measure used to find the differences of values predicted and observed.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} e_t^2}$$

5. Mean-squared error
   It is used to measure the average of the squares of the deviations.

$$MSE = \frac{1}{n} \sum_{t=1}^{n} e_t^2$$

Table 2 shows the comparison results of various forecasting methods by including multilayer perceptron (MLP) method.

The above table shows the comparison of various forecasting methods like linear regression, Gaussian processes, and SM Oreg. From the above table, we can observe that multilayer perceptron gives best results as direction accuracy is more

Table 2 Comparison of various forecasting methods

Metrics	Multilayer perceptron	Gaussian processes	SM Oreg	Linear regression
Mean absolute error	0.0009	193.3017	122.1547	230
Root relative squared error	0.0002	57.1637	50.6747	62.4235
Direction accuracy	100	66.6667	66.6667	0
Relative absolute error	0.0002	52.8633	38.4491	59.1929
Mean absolute percentage error	0.0003	79.0042	27.824	92.1868
Root-mean-squared error	0.001	242.1275	09.8416	271.8374
Mean-squared error	0	58625.730	44033.486	73895.5625

than the others and difference between the predicted and given values is very less for multilayer perceptron.

5 Results and Analysis

Figure 3 shows the number of patients to be admitted in the succeeding year based on the previous year's data by using multilayer perceptron forecasting method [13].

After obtaining the above results, we need to cross-check whether the result is accurate or not. Our work also includes calculating the metrics in order to analyze the best forecasting method. Based on the analysis of standard metrics, we can choose the best forecasting method as the result retrieved to be supplied to the organization for maintaining the resources and enhancing the profits as shown in Fig. 4. By performing analysis on different forecasting methods, we can observe that multilayer perceptron is best suitable, as direction accuracy is high for it when compared to others as shown in Fig. 4. And difference between the predicted value and actual value is very low.

The above graph shows that MLP is best when compared to other forecasting methods [8, 15]. Here, we can see that accuracy of MLP is high and mean absolute error of MLP is low that is the difference between predicted value and actual value. This graph gives a clear view that multilayer perceptron has a directional accuracy and low mean absolute error.

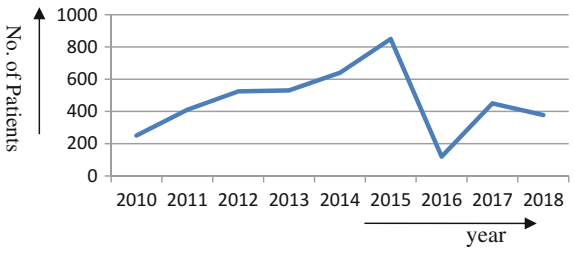


Fig. 3 Multilayer perceptron-based yearly patient details

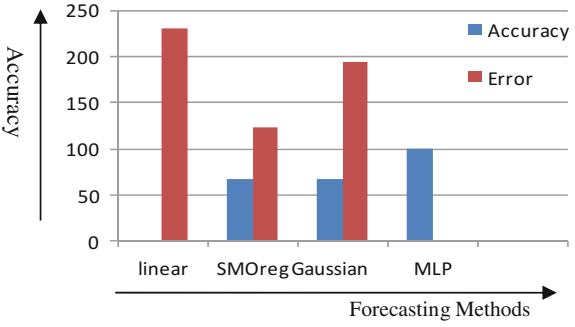


Fig. 4 Accuracy comparison of forecasting methods

6 Conclusion

Nowadays, data maintenance led a key position in enterprise resource system. This research utilizes back propagation which is the most widely used tool in ANN. By this work, we can conclude that multilayer perceptron gives better results when compared to other forecasting methods based on the standard metrics analysis, and by using these results, healthcare centers can predict number of patients to be admitted in the succeeding year due to particular disease and can maintain resources within the organization like inventory, human resources, and machinery. And the comparisons with other forecasting methods show multilayer perceptron gives accurate results than other forecasting methods based on the standard metrics obtained. By using the multilayer perceptron which is proved to better than the other methods, good results are obtained which are fed to the organization in order to manage the resources in the succeeding year.

Authors have taken the consent from the concerned person/authority to use the materials, etc. in the paper. Authors will be solely responsible if any issues arise in future with regard to this.

References

1. Meena, M., Bharadi, V.A.: Performance analysis of cloud based software as a service (SaaS) model on public and hybrid cloud. In: 2016 Symposium on Colossal Data Analysis and Networking (CDAN) (2016)
2. Al-Ghofaili, A.A., Al-Mashari, M.A.: ERP system adoption traditional ERP systems vs. cloud-based ERP systems. In: Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014) (2014)
3. Ramchoun, H., Amine, M., Idrissi, J., Ghanou, Y., Ettaouil, M.: Multilayer perceptron: architecture optimization and training. Intl. J. Interact. Multimed. Artif. Intell. **4**(1), 26 (2016)
4. Gupta, S.: A regression modelling technique on data mining. Intl. J. Comput. Appl. **116**(9), 27–29 (2015)
5. Tiwari, A. et al.: ANN-based classification of mammograms using nonlinear preprocessing. In: Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications. Springer, Singapore (2018)
6. Gautam, Aman, et al.: An Improved Mammogram Classification Approach Using Back Propagation Neural Network, Data Engineering and Intelligent Computing, pp. 369–376. Springer, Singapore (2018)
7. Litta, A.J., Idicula, S.M., Mohanty, U.C.: Artificial neural network model in prediction of meteorological parameters during premonsoon thunderstorms. In: Intl. J. Atmos. Sci. **2013**, 1 − 14 (2013)
8. Saini, I., Khanna, A., Peddoju, S.K.: Cloud and traditional ERP systems in small and medium enterprises. In: 2014 International Conference on Information Systems and Computer Networks (ISCON) (2014)
9. Advanced Time Series Forecasting Methods. Advances in Time Series Forecasting, pp. 3 − 10. Bentham Science Publisher, **Sharjah** (2012)
10. Setiawan, A.S., Elysia, Wesley, J., Purnama. Y.: Mammogram classification using law's texture energy measure and neural networks. In: International Conference on Computer Science and Computational Intelligence (ICCSCI), vol. 59, pp. 92–97 (2015)

11. Bhateja, V., Misra, M., Urooj, S.: Human visual system based unsharp masking for enhancement of mammographic images. In: Journal of Computational Science (2016)
12. Boedecker, J. Dr.: Multilayer Perceptron, Machine Learning Lab, University of Freiburg (2015)
13. Forecasting performance evaluation and reporting. Business Forecasting, pp. 143–279 (2016)
14. Molnár, B., Máriás, Z.: Design and implementation of a workflow oriented ERP system. In: Proceedings of the 12th International Conference on e-Business (2015)
15. Bhateja, V., Urooj, S., Misra, M.: Technical advancements to mobile mammography using non-linear polynomial filters and IEEE 21451-1 information model. IEEE Sens. J. **15**(5), 2559–2566 (2015)

A Comprehensive Survey of Services Provided by Prevalent Cloud Computing Environments

N. Joshi and S. Shah

Abstract Cloud computing has emerged as a successful model for enterprises as well as for academicians, promising on-demand, scalable and fault-tolerant computing facilities. Different cloud service providers offer various types of services and support wide range of software frameworks. Additionally, clients have different in-house system capabilities and application requirements. Thus, one cloud environment may not fit for all. In this paper, we elaborate on various services provided by cloud service providers under infrastructure as a service and platform as a service model. We also present comparative study of services provided by proprietary and open-source-based prevalent cloud computing environments. At the end, the paper gives recommendations and relevant justification about which cloud environment is suitable to a client based on client's requirements and available resources.

1 Introduction

Cloud computing is a ubiquitous service provisioning model providing different hardware and software resources over the network. It ensures on-demand resource provisioning and pay per use billing model [1], thus amalgamating principles of grid computing and utility computing. Essential characteristics of cloud computing additionally include elasticity, scalability of infrastructure, and multitenant environment where applications of multiple clients can be cohosted.

Cloud computing services are mainly categorized as infrastructure as a service (IaaS), provisioning hardware components like processor, memory, network and platform as a service (PaaS), provisioning server instances prebooted with operating

N. Joshi (✉) · S. Shah
Mukesh Patel School of Technology Management and Engineering, Mumbai, India
e-mail: nilambarijoshi@gmail.com

S. Shah
e-mail: seema.shah@nmims.edu

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_41

system and installed application framework for application hosting and software as a service (SaaS) that provides ready-to-use network-based user-centric applications.

There is a difference in cloud deployment approach as well. Public clouds are set up and owned by cloud service provider (CSP), and resources are provided to different clients. Private clouds are solely controlled and used by a single client, whereas hybrid cloud allows integration of public cloud and private cloud services. Many small and medium enterprises (SME) can integrate old legacy applications with new cloud-based applications.

Comparison of cloud computing environments is a widely discussed topic. In paper [2], cloud computing is defined as fifth utility and its prospects are discussed. Papers [3–6] compare different components of open-source cloud computing environments, whereas [7] has done performance evaluation of private cloud based on Eucalyptus and CloudStack. Papers [8, 9] discuss framework and parameters to evaluate cloud environments. PaaS CSPs are studied in [10]. In paper [11], comparison of open-source clouds is made with respect to different services provided. But none of these studies have given insight into the suitability of cloud environments as per client's requirements and available resources. Further, most of the earlier studies focus on IaaS. In our paper, we have incorporated the study of PaaS service providers as well.

Various factors such as availability of in-house infrastructure, legacy applications, budget, IT expertise, data and application sensitivity need to be considered by the clients while choosing appropriate cloud services, cloud deployment model, and cloud service provider.

This paper aims at studying different prevalent cloud computing environments and suggests suitability of a cloud environment as per client's requirements and available resources. The paper discusses services provided under IaaS and PaaS service models by cloud computing environments in Sects. 2 and 3, respectively. It does a comprehensive study of different proprietary and open-source environments. Section 4 provides recommendations about which cloud environment should be chosen as per client's requirements and available resources, giving justification based on the study done.

2 IaaS Cloud Computing Environment

IaaS provides hardware resources encompassing CPU, memory, network bandwidth, and storage. The resources are owned and maintained by cloud service provider (CSP) and rented out to the clients. The client gets the flexibility to choose an operating system, application framework, etc. By subscribing to IaaS, clients can invest in in-house human and financial resources in business development instead of infrastructure management. IaaS services can be categorized as below.

2.1 Computational Services

Computational services provision the resources required for program execution like CPU and memory. They are provisioned as logical units in many ways.

Server virtualization is the traditional way which abstracts underlying hardware by introducing hypervisor layer. It divides host hardware into logical units; each unit is called as virtual machine (VM). Each virtual machine can be independently booted irrespective of the underlying host operating system and acts as an independent machine. In reality, virtual machines share hardware of host machine but are isolated from one another by the hypervisor. Clients with monolithic applications can prefer virtualization-based instances. Some CSPs use proprietary hypervisors like VMware ESXi and Microsoft's Hyper-V, and some use open-source hypervisor like Kernel Virtual Machine is used for Google Cloud Platform (GCP) [12]. Amazon Web Services (AWS) [13] supports XenServer with unmodified guest operating system as well as Xen with para virtualized images.

The recent practice of containerization-based approach is a lightweight alternative to VM. It uses host operating system kernel but logically divides user space into different execution environments called as containers. Each container contains OS image tailored to a particular application, so it is much lighter than VM. Almost all prevalent proprietary as well as open-source cloud environments support docker [14] runtime-based container service. Clients developing highly granulated microservices-based applications can opt for this.

Bare metal-based IaaS is another important way but provided by very few CSPs. Service providers like IBM provide bare metal servers which can be virtualized by the clients with the hypervisor of their choice. This gives flexibility to the client but at the same time puts more responsibility for IT administration on client's shoulder.

An option available with few providers is of preemptive virtual machine. These VMs are offered at a very low cost but at the same time, when resource crunch is experienced and instances with stringent SLAs need resources, these VMs are preempted. So the user may lose his work. This can be used for nontransactional and noncritical applications.

Many CSPs give good support to migrate legacy applications on cloud. Thus, consumer can make his applications more scalable. Open-source frameworks like Eucalyptus [15] provide good API compatibility with AWS services, so hybrid cloud development is easy.

2.2 Network Services

Network resources include network bandwidth and network interface plus middleware components to deal with network traffic. Service providers facilitate virtual

networking among VMs over physical networking. Network traffic isolation among virtual machines of different organizations is achieved by virtual private network (VPN) which creates a private network of virtual machines used by a client's organization. VMs from one VPN can access each other but not virtual machines in another VPN. This ensures isolation of applications and data owned by different organizations in a multitenant environment. Clients with workflow-based applications will need VPN for secure interapplication communication.

Another important service provided by CSPs is content delivery network (CDN) that works on the principle of caching web content on network edge point near client. For the first time, the client request is forwarded to actual web server. The web server mentions in response whether the response is cacheable or not. The network edge node caches the response if it is cacheable. If the same request is sent again to the edge node, then the response is given from cache instead of the actual web server, thus reducing network traffic and improving response time. Web-based applications with static content will be benefited by CDN.

2.3 Storage Services

IaaS mainly provides two types of storage. One is block storage facility which stores data in raw format without considering structure of the data. Data is transferred considering data block as unit. With block storage, only block which contains the requested part of the file is transferred instead of the whole file. AWS's elastic block storage, GCP's persistent storage, or OpenStack's cinder service are block type storages. Cold storage service is peculiar service based on block data storage used to archive data like images and videos which are not used frequently. They are low price storage but have more latency compared to normal storages.

Another storage type is object storage which stores data as of particular type like a file and database table. They are used to store massive unstructured data. AWS's S3, GCP's cloud storage, or OpenStack's [16] swift are some of the examples of object storage. Some CSPs like Microsoft, IBM, and VMware [17] provide SAN-based block storage as well as NAS-based file system-based storage capabilities.

Table 1 shows a comparison of different IaaS cloud computing environments with respect to the services mentioned above. It is evident that GCP, AWS, Microsoft Azure provide a wide range of services. Whereas, in open-source alternatives, OpenStack provides comprehensive services.

Table 1 Comparison of IaaS cloud computing environments

Resource type	GCP [12]	AWS [13]	Azure [18]	Bluemix [19]	vSphere [17]	Open Stack [16]	Cloud Stack [20]	Open Nebula [21]	Eucalyptus [15]
Virtualization	KVM	XenServer	Hyper-V	As per clients choice	ESXi	KVM, vCenter XenServer	XenServer, KVM, vCenter, Hyper-V	KVM, vCenter, Xen	KVM, vCenter, Xen
Containerization	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bare metal	No	No	No	Yes	No	Yes	No	No	No
Preemptive VM	Yes	Yes	No	No	No	No	No	No	No
Migration support	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Block storage	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cold storage	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Object storage	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VPN	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CDN	Yes	Yes	Yes	Yes	No	Yes	No	No	No

3 PaaS Cloud Computing Environment

PaaS provides software frameworks preinstalled on hardware as an instance on which applications can be easily deployed. The CSP takes care of installation and configuration of application execution environment, and client developer can truly concentrate on application development. The hardware used for PaaS services can be owned by PaaS CSP like App Engine of Google or rented out from other IaaS providers as provided by Cloud Foundry on IBM [19] datacenters.

Most of the CSPs support different linux distributions and windows versions. A wide variety of programming languages and frameworks are also supported inclusive of Java, .net, PHP, Python, ruby, etc. In addition to web-based application development, many providers support mobile applications which leverage massiveness and scalability of cloud computing. Firebase framework provided by Google Cloud Platform [12] is one of the examples.

Database services by CSPs provide structured as well as unstructured database. Relational or structured databases like MySQL, PostgreSQL, and SQLServer are supported to handle transactional data. With the advent of big data, unstructured NoSQL databases like MongoDB, HBase, DynamoDB, and Cassandra are used to store images, videos, big data collected from various sources. In-memory databases like Redis are also widely supported to reduce the latency of I/O operations.

In the past one decade, demand of application development in areas of big data, Internet of Things (IoT) and machine learning (ML) is increasing. Keeping in mind future development in these domains, many CSPs provide platforms to develop and deploy related applications. PaaS providers provide special frameworks like cloud-based Hadoop cluster to facilitate map reduce-based parallel and distributed processing [22], as well as Spark integrated with Hadoop to process stream data.

Machine learning-based APIs for voice recognition, image classification, text translation, natural language processing for text analysis are being provided by many service providers to develop ML-based applications.

Integration of IoT and cloud computing is giving rise to the concept of IoT cloud [23]. Many CSP facilitate this by providing software development kits of different languages like embedded C, Java, .net, Arduino yum for developers to develop IoT-oriented applications. To ensure reliability and fault tolerance, message queuing protocols like AMQP or MQTT are supported.

Different proprietary as well as open-source platforms-based PaaS services are compared in Table 2. All PaaS providers support web application platforms and databases. Most of them support mobile application platforms as well except Cloud Foundry. Support for big data, ML, IoT is largely available with GCP, AWS, and Azure.

Table 2 Comparison of PaaS cloud computing environments

Resource type		GCP [12]	AWS [13]	Azure [18]	Bluemix [19]	Cloud Foundry [24]	Open Shift [25]	Sales Force [26]
Application platform	Web application	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Mobile application	Yes	Yes	Yes	Yes	No	Yes	Yes
Databases	Structured	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Unstructured	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	In-memory	No	Yes	Yes	Yes	Yes	Yes	Yes
BD support	Hadoop cluster	Yes	Yes	Yes	Yes	Yes	No	No
	Stream data handling	Yes	Yes	Yes	Yes	Yes	No	No
	Data analytics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Data warehouse	Yes	Yes	No	No	No	No	No
ML support	Speech recognition	Yes	Yes	No	Yes	Yes*	Yes*	Yes*
	Language translation	Yes	Yes	Yes	Yes	Yes*	Yes*	Yes*
	Image classification	Yes	Yes	Yes	Yes	Yes*	Yes*	Yes*
	Video surveillance	Yes	Yes	Yes	No	Yes*	Yes*	Yes*
IoT support	Programming, messaging support	Yes	Yes	Yes	Yes	Yes*	Yes*	Yes*

* Using the third party like GCP /AWS services

4 SaaS Cloud Computing Environment

Software as a service provides ready to use applications over the network, which are user-centric. It gives importance to convenient use, look and feel, user-based customization, etc. SaaS providers are responsible for application development, deployment, and publication. SaaS may use proprietary infrastructure and frameworks. Social networking giants like Facebook, Twitter, LinkedIn, banking industry, government offices have their own datacenters to support services. Small- and medium-size SaaS providers mainly leverage public IaaS and PaaS to ensure scalability, availability, and robustness without upfront investment in these technologies. Websites for online courses and travel bookings are hosted on public cloud service providers. The range of services provided under SaaS service model is huge. Most of the applications are web based. SaaS provider should take care of handling a sudden increase in application requests, platform upgrading, network-based security threats.

5 Key Findings—IaaS and PaaS Cloud Computing Environments

The comparative study presented in Tables 1 and 2 can guide a client to take a decision to choose service provider as per requirements. It will also suggest whether to move to public cloud or deploy a private/hybrid cloud using open-source technologies and to choose service provider as per requirements.

As a client of IaaS, following points can be noted with respect to proprietary CSPs. AWS, GCP, and Azure provide wide range of services provisioning computational, network, and storage resources. IBM provides bare metal servers as part of IaaS which gives client flexibility to choose hypervisor as well as other middleware components. This service can be used if client doesn't have adequate infrastructure but has skilled-IT staff to set up cloud and need dedicated resources. VMware provides IaaS based on its proprietary hypervisor and can be recommended to deploy private or dedicated cloud. Being licensed, it is not cost-effective solution.

With respect to open-source IaaS service providers, OpenStack provides almost all IaaS services at par with the proprietary CSPs. It is highly flexible and gives full control to the client. At the same time, it demands highly skilled technical IT staff to configure and maintain the cloud. Eucalyptus is highly compatible with AWS. This can be preferred option if hybrid cloud with AWS is envisaged in future. Except for OpenStack, other open-source IaaS frameworks do not support cold storage and CDN services.

As a client of Platform as a Service, following points can be considered.

GCP and AWS go hand in hand as far as PaaS service offerings are concerned. Both provide a rich set of programming language and database support. Big data

analysis, machine learning, and IoT-related application building frameworks are provided at par with each other. Microsoft Azure provides PaaS services supporting a majority of languages and databases. But its support for machine learning and IoT domain-related applications is limited. IBM provides PaaS service on its own infrastructure but based on third-party Cloud Foundry framework. Cloud Foundry and OpenShift online provide support for most of the languages and databases and thus can be considered promising PaaS frameworks to build PaaS cloud on third-party infrastructure. With respect to big data, machine learning, IoT domain application development, they can provide solution integrated with GCP or AWS services or using other third-party software.

Suitability of different cloud computing environments based on client's system parameters, viz. infrastructure, budget, and technical expertise are as given in Table 3. Other requirements considered are legacy applications, sensitivity of data, and support for special applications like big data.

Table 3 Summary of suitability of cloud computing environments based on client's requirements and resource availability

Client system parameters				Examples	Recommendations	Justification
Is	B	T	Other			
A	I	I	–	Academic projects	CloudStack/ Eucalyptus-based in-house cloud	Open-source and less complex installation
I	I	A	–	Academic projects, research projects	Open-source-based in-house cloud, GCP, AWS free trial offerings, preemptive virtual machine	Free trial offerings are economical and in-house cloud setup can be done on commodity hardware machines
I	A	I		SME, startups, research (noncomputer)	GCP, AWS, Microsoft, Cloud Foundry-based PaaS services	Less IT overhead due to ready to use instances on public cloud
I	A	A	–	SME, startup	AWS/GCP IaaS services	Scalable infrastructure on public cloud
A	I	A	–	Research Projects	OpenStack/ OpenNebula-based in-house cloud,	OpenStack/ OpenNebula provides robust flexible framework but complex

(continued)

Table 3 (continued)

Client system parameters				Examples	Recommendations	Justification
Is	B	T	Other			
						installation process
I	A	A	Special domain—big data, ML, IoT	Research projects SMEs	GCP, AWS cloud-based PaaS services	Rich APIs and Libraries-based service offerings
A	A	A	Sensitive data, Critical applications	Research projects SMEs	OpenStack/ Eucalyptus-based hybrid cloud	Private cloud is within client's administrative controls
I	A	A	Multiple legacy Applications	Old corporates or organizations	OpenStack/ Eucalyptus-based hybrid cloud or AWS/GCP using migration tools	Hybrid cloud is cost effective. Seamless migration of legacy applications to the cloud gives better performance

Is—Infrastructure, B—Budget T—Technical Expertise; A—Adequate, I—Inadequate

6 Conclusion and Future Plan

Cloud computing is a promising framework ensuring scalability, availability, and on-demand network-based resource provisioning. Prevalent proprietary, as well as open-source framework-based cloud service providers, offer various services under IaaS, PaaS, and SaaS service models. IaaS CSPs provide services offering computational, storage, and network resources, whereas PaaS CSPs provide services offering software development and deployment environments for web-based as well as mobile-based applications. PaaS also provide frameworks for new technologies in big data, machine learning, IoT domains.

A wide range of services provided by CSPs gives many options to the users to choose from. Considering existing infrastructure, legacy applications, available budget, and IT expertise, clients can make choice of appropriate CSP. Based on our comprehensive study, we have recommended and justified suitability of different cloud as per client's system and application requirements. Open-source framework-based cloud environments are economical compared to proprietary and hence can be widely used for academic and research projects. They have a drawback of having limited support for special purpose applications like big data, machine learning, and IoT. But they can integrate with third-party tools. Among public proprietary cloud providers, Google Cloud Platform and Amazon Web Services provide highly rich, comprehensive range of services. They can be opted

more at an enterprise level. Academic and research projects demanding scalability and with good funding support can opt proprietary public clouds. Proprietary CSP has come up with economical services like preemptive instances which can be used for noncritical application hosting.

Considering wide use of cloud-based services especially in big data, ML, and IoT domain, as a future plan, we intend to study in detail geographical distribution of data centers. This is to decide suitability from a performance point of view. We also plan to study technical strategies of resource provisioning, load balancing, auto-scaling to ensure scalability, availability, and reliability of cloud environments.

References

1. National Institute of Standards and Technology (NIST). http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800–145.pdf
2. Buyya, R. et al.: Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. Future Gener. Comput. Syst. **25**(6), 599–616 (2009)
3. Kumar, V., Sharma, S.: Comparative study on open source cloud platform solutions available for deploying on private, public or hybrid cloud. Intl. J. Innov. Res. Comput. Commun. Eng. **5** (2), 109–113 (2017)
4. Kumar, R. et al.: A comparative study for cloud computing platform on open source software. Intl. J. Eng. Technol. (AIJET) **1**(2), 28–35 (2014)
5. Manjaly, J.S., Jisha, S.: A Comparative Study On Open Source Cloud Computing Frameworks. Intl. J. Eng. Comput. Sci. **2**(6), 2025–2029 (2013)
6. Limbole, R.P., Bhole, G.P.: Infrastructure-as-a-service and comparative study on open-source cloud infrastructure solutions. In: International Journal of Current Engineering and Technology, pp. 2082–2488 (2014)
7. Ali, M.M., Abdulrazak, A.: Performance evaluation of private clouds eucalyptus versus cloudStack. Intl. J. Adv. Comput. Sci. Appl. **5**(5), 108–117 (2014)
8. Schmidt, R.: A framework for comparing cloud-environments. In: Proceedings of the Federated Conference on Computer Science and Information System, pp. 553–556 (2011)
9. Dhivya, R. et al.: Parameters and methods used to evaluate cloud service providers: a survey. In: International Conference on Computer Communication and Informatics (ICCCI -2016), pp. 1–5 (2016)
10. Varma, M.K., Choi, E.: Comparative Study of various platform as a service frameworks. Intl. J. Cloud Comput.: Ser. Arch. (IJCCSA) **6**(1), 23–34 (2016)
11. von Laszewski, G. et al.: Comparison of multiple cloud frameworks. In: IEEE 5th International Conference on Cloud Computing (CLOUD), pp. 734–741 (2012)
12. Google Cloud Platform. https://cloud.google.com/products/
13. Amazon Web Services. https://aws.amazon.com/
14. Docker. https://www.docker.com/
15. Eucalyptus. https://en.wikipedia.org/wiki/Eucalyptus_(software)
16. OpenStack. https://www.openstack.org/software/
17. Vmware. https://www.vmware.com/cloud-services/infrastructure.html
18. Microsoft Azure. https://azure.microsoft.com/en-in/
19. IBM Bluemix. https://www.ibm.com/cloud-computing/bluemix/cloud-servers
20. OpenNebula. https://www.opennebula.org/
21. CloudStack, https://cloudstack.apache.org/
22. Bajaber, F. et al.: Big data 2.0 processing systems: taxonomy and open challenges. In: Journal of Grid Computing, pp. 379–405 (2016)

23. Celesti A. et al.: A watchdog service making container-based micro-services reliable in IoT clouds. In: IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud), pp. 352–358 (2017)
24. Cloud Foundry. https://www.Cloud Foundry.org
25. OpenShift. https://www.openshift.com
26. SalesForce. http://www.salesforce.com/in

Conditional Generative Recurrent Adversarial Networks

Siddharth Seth and Mukesh A. Zaveri

Abstract Generative adversarial networks (GANs) introduced by Goodfellow et al. since their advent have had a number of improvements and applications in image generation tasks and unsupervised learning. Recurrent model and the conditional models are two derivations of GANs. In this paper, conditional recurrent GAN is proposed. By using conditional settings in recurrent GANs, they can be used to generate state-of-the-art images. The conditional and recurrent models are compared with the proposed model using the generative adversarial metric proposed by Im et al. where the discriminator of one model competes against the generator of another. The results show that the proposed model outperforms the other two models.

1 Introduction

Deep learning, a sub-field of machine learning, has been inspired by the structure and function of a human brain, thus moving us closer toward achieving general artificial intelligence. It is said to be deep as the learning takes place at several layers of abstraction [1]. The learning is based on a hierarchical model [1]; i.e., the lower layers learn the low-level features and higher layers the high-level features. Some deep learning-based architectures are deep belief networks (DBNs) [2], convolutional neural networks (CNNs) [3], recurrent neural networks (RNNs). Deep learning has given state-of-the-art results, beating in some cases human experts, in tasks such as image classification, audio processing and natural language processing to name a few.

S. Seth (✉) · M. A. Zaveri
Computer Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India
e-mail: siddharthseth1993@gmail.com

M. A. Zaveri
e-mail: mazaveri@coed.svnit.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_42

Image generation was previously used to understand how the neural network model perceives the visual world, its scope thus being limited as a diagnostic tool [4]. Deep learning however has brought immense improvement in the quality of images generated. In [5], Dosovitskiy et al. use CNNs for generating chairs, and in [6], Gregor et al. use recurrent neural networks (RNNs) for image generation. Deep convolutional GANs (DCGANs) [7] which were proposed by Radford et al. employed deep convolutional neural networks, and [8] proposed Laplacian pyramid of adversarial networks (LAPGAN). References [9–12] were some other papers that discussed the image generation models. Even though there were numerous proposals on image generation suggesting different methods for generating images, there was not much success until the introduction of GANs by Goodfellow et al. GANs avoid the difficulties posed due to approximation of probabilistic computations by the deep generative models and have recently become popular due to the immense success they have achieved in this difficult task of image modeling. The DCGANs specifically were one of the most successful models generating realistic images from different datasets such as LSUN and Imagenet-1 k [13].

This paper focuses on the sequential model of GANs which have shown to produce good results in image generation. One of the sequential models proposed is the LAPGAN which generates images in a coarse-to-fine fashion sequentially in multiple steps. Another sequential model generative recurrent adversarial networks (GRANs) [4] took motivation from model proposed by Gatys et al. [14], the DRAW network proposed by Gregor et al. [6] and the LAPGAN. The DRAW network uses variational autoencoders (VAEs) [15]. VAEs are a class of probabilistic generative models that have shown promising results in generating images. VAEs use latent variables which encode the images in a vector form in a high dimensional space. The problem with VAEs is that they often produce blurry images. LAPGANs are also based on a sequential model but are generative adversarial models.

Conditional generative adversarial networks take a condition as input at both discriminator and generator side, thereby directing the output to as desired. This condition may be given as an image or a set of features represented as vectors along with the input noise vector or may be given as input to the output layer which is responsible for making the final decision of classification in case of a discriminator and a generated image in case of a generator. The proposed GAN model in this paper is based on the GRAN and conditional GAN models. The contributions of this paper are as follows:

- By using condition setting in GRAN, the proposed model generates state-of-the-art and valid images.
- Generator of conditional GRAN requires less training time compared to GRAN model in matching the model's data distribution to the real data distribution.
- The discriminator of conditional GRAN yields lower error rate as compared to GRAN and conditional GAN.

Identifying and filling in missing data in images by masking the generated image over the corresponding damaged image, extending a small dataset by generating images from the same probability distribution and using them in training data, using the images from generators to verify discriminator's learnt features are some of the applications of generating images within close reach.

2 Related Work

Prior to generative adversarial networks, some of the methods proposed for image generation were the DRAW model [6] using VAEs, [16] using an iterative forward diffusion process and various others using CNNs. GANs have however outperformed any of the previously known methods for generating images. The following sections describe the models appropriate to the proposed model for the image generation task.

2.1 Generative Adversarial Networks

GANs framework can be seen as a pair of networks engaged in a two-player minimax game where each player tries to minimize its loss and maximize its profit [17]. Of the two players, first is the discriminator that tries to discriminate between real data distribution and the model's data distribution and second is the generator that tries to match the model's probability distribution to that of the real data's probability distribution. For images, the generator takes random noise values from a probability distribution, z such as Gaussian or normal, and maps them to output values x through a series of transpose convolutions and nonlinear transformations. The discriminator outputs high values wherever the density of the real data is greater than the density of generated data. The generator changes the samples that it produces to move uphill along the function learned by the discriminator. In other words, the generator moves its samples into areas where the model distribution is not yet dense enough. Eventually, the generator's distribution matches the real distribution and the discriminator has to output a probability of one-half everywhere, because every point is equally likely to be generated by the real data set as to be generated by the model. The minimax equation we thus get is [17]:

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))]. \quad (1)$$

$V(D,G)$ is the cost function of two networks: the discriminator (D) and the generator (G). E denotes the expectations over real data x and model's data z. $D(x)$ gives the output as a single scalar of whether the given sample x is from data or model distribution. Similarly, $G(z)$ represents the image output by the generator,

G and $D(G(z))$ thus gives the probability that the sampled image $G(z)$ came from original data distribution. However, maximizing $\log D(G(z))$ instead of minimizing $\log(1-D(G(z)))$ solves the problem of vanishing gradients and leads to stronger gradients in learning. The following two equations better represent the cost function of a GAN model:

For discriminator [4],

$$\max_{D} E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))]. \tag{2}$$

For generator [4],

$$\max_{G}[logD(G(z))]. \tag{3}$$

GANs model introduced by Goodfellow et al. is a simple model and only works well enough on small datasets like MNIST. Since then a number of modifications, [4, 8, 18, 19] have been put forward, one of them being the DCGANs [7] architecture which has been used in the proposed conditional GRAN model too.

Apart from discriminating between fake and real images, the discriminator can also be used in image classification. Features extracted by a discriminative model for image classification can be visualized using a generative model giving us an understanding of how the discriminative model perceives the images.

2.2 Conditional Generative Adversarial Networks

With simple GANs, one cannot control the data being generated as the generator takes a noise vector z and transforms it randomly into a data point from its learned probability distribution. By giving a condition, we can control the mode of the output and direct the output as desired. The conditional information may be class labels, images themselves or data from other modality. Here in the model, class labels are used as a conditional setting.

With conditional settings applied on GANs, the objective function can be modeled from that of GANs as [18]:

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x|y)] + E_{z\sim p_z(z)}[\log(1-D(G(z|y)|y))] \tag{4}$$

This comes from the conditional probability formula, i.e., probability of x given y, where y is the conditional setting. As the conditional variable, y is used in both discriminator and generator, the cost function determines the output of $D(x)$ given y for discriminator and the output of $G(z)$ given y for generator. Since $G(z|y)$ is then fed into discriminator, the cost function again calculates the probability of $G(z|y)$ being from real data distribution given condition y to discriminator. The cost

function used in the conditional GRAN for discriminator and generator comes after breaking Eq. 4 similar to Eq. 1.

For discriminator [18],

$$\max_D E_{x\sim p_{data}(x)}[\log D(x|y)] + E_{z\sim p_z(z)}[\log(1 - D(G(z|y)|y))]. \tag{5}$$

For generator [18],

$$\max_G [logD(G(z|y))] \tag{6}$$

The model suggested by Mirza in [18] is a simple proof of concept, not using any complex architecture for image generation task. This conditional approach applied to the GRAN model gives impressive results.

Inspired by [14], the DRAW model in [6] and LAPGAN model in [8], Im et al. in [4] proposed the GRAN network which generates images sequentially by adding the consecutive images produced by the generator after each time step onto an image canvas. The number of steps can be varied as suitable with one step being equivalent to a simple GAN.

The model of conditional GRAN described in the next section is built on top of GRAN and produces better visual quality images with less training required as shown in Sect. 4. It functions the same way a GRAN does, processing the image from generator at multiple time steps and then adding all the intermediate updates to get the final image. The architecture however is based on DCGAN model using its suggested parameters and an additional input in the form of a conditional setting is provided at both generator and discriminator. Section 3 describes the model in detail along with the parameters chosen for training.

3 Proposed Model

The model is based on generative recurrent adversarial networks, wherein a conditional setting is applied at input to every layer of both the generator and the discriminator and is called conditional GRAN. The cost function for the model is same as that of a conditional model as the input and final output remain the same. Only the processing part of image by the generator changes as the image sampled from noise vector is processed multiple times at the generator's end. The objective remains the same—maximizing the discriminator's probability of assigning the correct label to both the sampled data from generator and the training data.

Generator of the conditional GRAN takes three inputs—a noise vector z sampled from uniform distribution, a conditional vector y, and the encoded vector $H\_C(t)$. $H\_C(t)$ is the encoded vector of the image from generator and is given out by the discriminator at time t. Conditional vector y may be a vector of class labels, an image or data from any other modality. Initially, $H\_C(t)$ is assigned a vector of zeros. After subsequent processing of the image through the generator, at different

time stamps, it consists of the encoded values of the images given as input to the discriminator by the generator. In encoding, similar vectors represent similar images. The concept is taken from VAEs [15]. These three vectors are concatenated before giving them as input to the generator. The generator then through consecutive upsampling tries to generate an image $\Delta C(t)$ at time t from the current probability distribution it knows. This image is then given as input to the discriminator which gives an encoded vector $H\_C(t+1)$ as output. This encoded vector along with a randomly sampled noise vector z and the conditional vector y is given as input to the generator after concatenation. The generator outputs an image $\Delta C(t+1)$. This recurrence takes place till T time steps from $0\,to\,T-1$ and all the intermediate updates that is the images $\Delta C(0)$ till $\Delta C(T-1)$ are accumulated to give the final image on the canvas C. This final image is then fed to the discriminator for classification. The discriminator takes two inputs—the final sampled image, C, and the same conditional vector y as given in generator which lets the discriminator know which class of image it should expect. Therefore, if the condition y is a label '4' from MNIST dataset then the discriminator knows that the input image should look like the digit '4'. If it does not, it classifies the image as fake and the corrective updates are backpropagated to the generator for adjusting its weights to better understand how the digit '4' looks like.

This whole process in a simple GAN takes place in a single step where the generator takes in a randomly sampled noise vector z and produces an image which it then gives as input to the discriminator for discerning between real and fake data. Figure 1 gives an overview of how the model works where y is the conditional vector, z is the sampled noise vector, and $H\_C(t)$ is the encoded output from the discriminator at time step t. $\Delta C(t)$ denotes the image generated at time step t. By adding all the subsequent images $\Delta C(0) \ldots \Delta C(T-1)$ produced at time steps $0 \ldots T-1$, we get the final image as output on a canvas, C.

DCGAN [7] architecture is used in conditional GRAN:

- Strided convolutions instead of pooling layers—all convolutional networks is used with no average, min or max pooling;
- Using batch normalization in both generator and discriminator stabilizes learning by normalizing the input at each layer, each unit to have zero mean and unit variance except at the output layer of generator and input layer of discriminator;
- Using convolution layers instead of fully connected layers;
- Using LeakyReLU [20] activation in discriminator and generator (which is an exception to what is proposed in DCGAN about using ReLU in generator) fixes the dying ReLU problem.

Adam optimizer [21] has been used, with the momentum term β_1 set to 0.5 and the slope leak set to 0.2 as suggested in DCGAN. Learning rate used is 0.0001. We also use one-sided label smoothing. We execute generator function twice instead of just one time along with one call to discriminator to avoid fast convergence of the discriminator. Throughout our model, we use the bounded range [–1, 1] and thus use the *tanh* function wherever necessary.

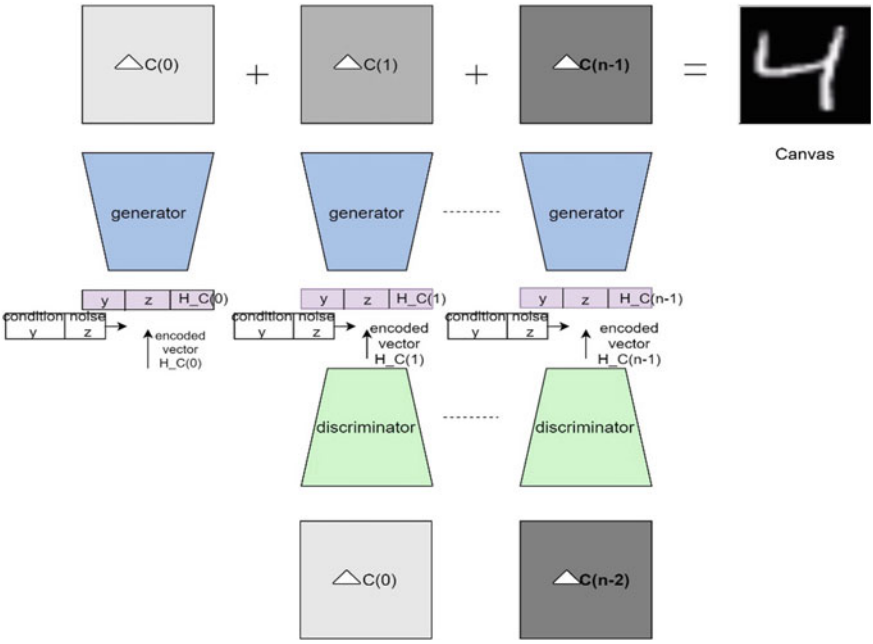


Fig. 1 Conditional GRAN architecture

The images produced on Canvas C are the final state-of-the-art images that have been shown in the results and compared with the images sampled from conditional GAN and recurrent GAN models. The following algorithm represents the model succinctly:

Algorithm CGRAN:

$z_t \sim p(Z)$
$H\_C(0) = vector(zeros)$
$for\ \ t = 1\ in\ number\ of\ steps\ T$:
    $C(t) = \ generator(z_t, H\_C(t-1), y)$
    $H\_C(t) = discriminator(C(t), y)$
$C = \ \tanh\left(\sum_{t=0}^{T-1} C(t)\right)$

$d_{logits_{fake}} = discriminator(C, y)$
$d_{logits_{real}} = discriminator\left(real_{image}, y\right)$

$d_{loss_{real}} = cross\_entropy\_loss(d_{logits_{real}}, ones)$
$d_{loss_{fake}} = cross\_entropy\_loss(d_{logits_{fake}}, zeros)$
$g_{loss} = cross\_entropy\_loss(d_{logits_{fake}}, ones)$

$d_{loss} = d_{loss_{real}} + d\_loss_{fake}$
$update\ Discriminator$
$update\ Generator$

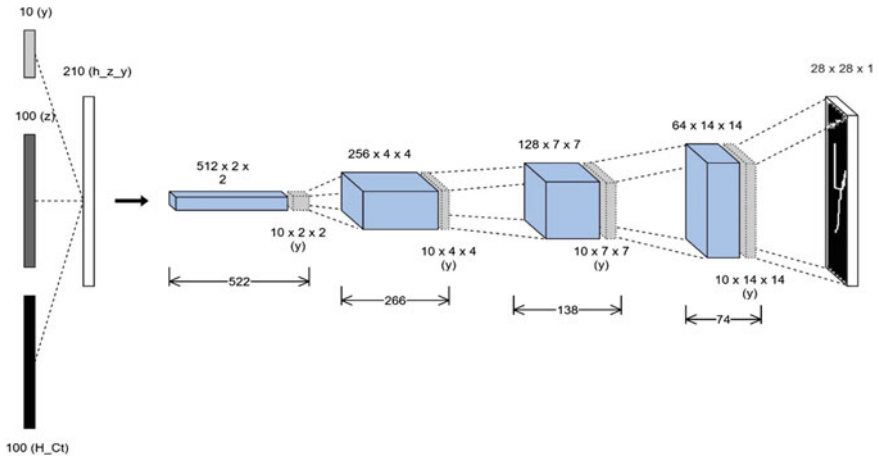


Fig. 2 Generator architecture for conditional GRAN

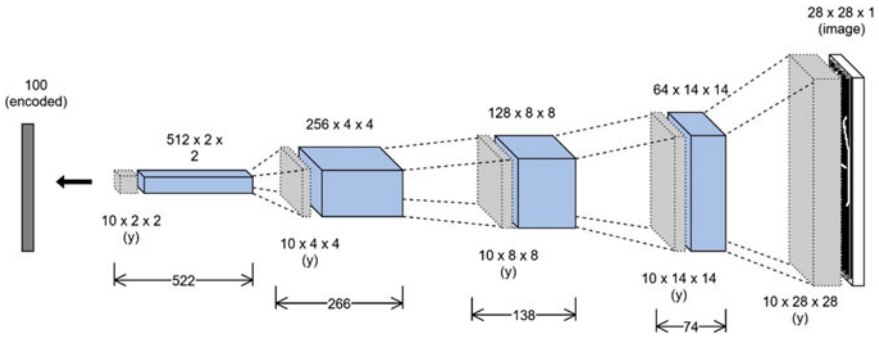


Fig. 3 Discriminator architecture for conditional GRAN

Figures 2 and 3 depict the generator and discriminator architecture used in the conditional GRAN.

4 Simulation and Results

Experiments were conducted on MNIST dataset for 1000 images generated by each model with 100 images per class. After training the networks for 30 epochs, it was observed that conditional GRAN model performs better than GRANs as it generates all valid MNIST images, whereas GRANs occasionally generate images that cannot be classified to either of the ten classes and thus seem invalid. Throughout the experiments, all the models' architectures were maintained to have the properties of

Table 1 Model parameters used while experimentation

S. No.	# Filters	Learning rate	Comments
1	512, 256, 128, 64	0.00005	Performed same as #3
2	**512, 256, 128, 64**	**0.0001**	**Best architecture**
3	80, 40	0.0001	Performed same as #1
4	80, 40	0.00001	Worst performing among the 4

the DCGAN architecture and the models evaluated with different layers' sizes. Table 1 shows some of the parameters chosen while experimentation. We found that the models with more number of layers fared better than others. The architecture chosen for the final model is highlighted in the table. This model gave the best visual quality images.

Some sample images generated by the three models—GRAN, CGRAN, and CGAN, along with the actual MNIST dataset are shown in Fig. 4 for comparison. Also, Fig. 5 shows some non-classifiable images generated by the GRAN network.

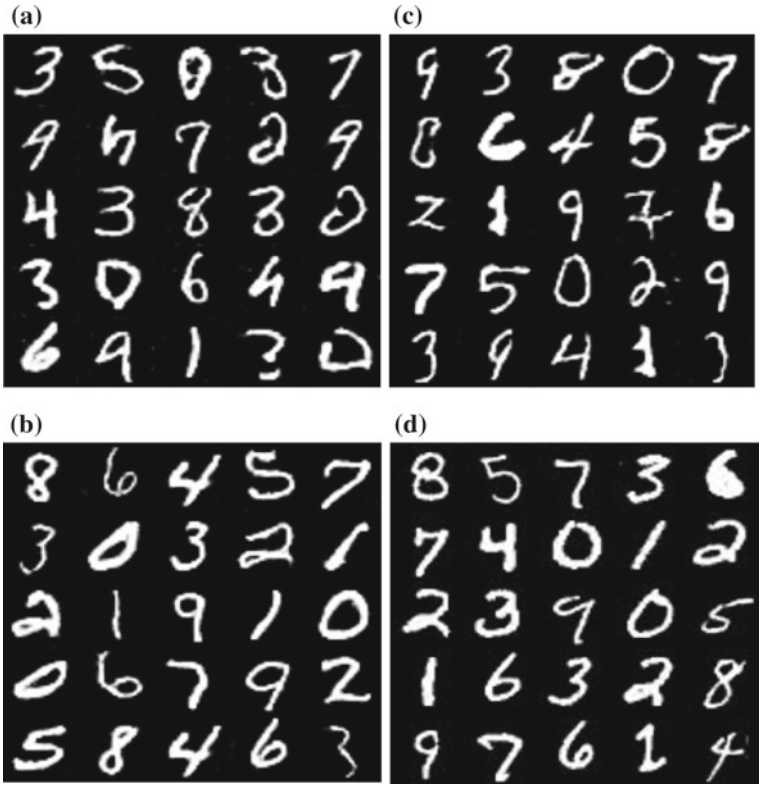


Fig. 4 Images generated by **a** GRAN, **b** CGRAN, **c** CGAN models, and **d** actual MNIST

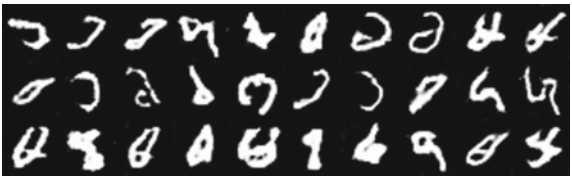


Fig. 5 Non-classifiable images generated by GRAN model

Table 2 Error ratio for image classification by the discriminators of different models on 1000 images

	CGAN versus CGRAN	Winner	CGRAN versus GRAN	Winner
CGAN data	7.2	CGRAN	0.8	CGRAN
GRAN data	7.4	CGRAN	0.5	CGRAN
CGRAN data	2.1	CGRAN	2.7	GRAN

We follow the GAM metric [4] for evaluating the GAN models as we find it a reasonable method for comparison between different GAN models' performance. Table 2 shows the error ratio comparison between three models: conditional GRAN, GRAN, and conditional GAN which calculates the ratio of number of images falsely classified as real for generated sample data by model 1 to that of model 2 (r_{sample}). Also, the test ratio r_{test} depicts the ratio of number of images from real data falsely classified as fake. This should be $\simeq 1$ as it tells the discriminators are at most equally biased.

In Table 2, second column represents r_{sample} of CGAN/CGRAN and fourth column shows r_{sample} of CGRAN/GRAN. As it is evident from the table, CGRAN model's discriminator outperforms others by classifying more images correctly as either fake or real. The only place where it falters is on its own dataset. Similar is the case with the GRAN model's discriminator which fails on its own dataset. This is due to the biasing problem where the network's discriminator is fooled by its own generator. The evaluation metric for two competing models, Model 1 and Model 2, is as follows [4]:

$$winner = \begin{cases} Model\,1, & if\ r_{sample} < 1\ and\ r_{test} \simeq 1 \\ Model\,2, & if\ r_{sample} > 1\ and\ r_{test} \simeq 1 \\ Tie, & otherwise \end{cases}. \tag{7}$$

5 Conclusion

We showed that applying conditional setting to a GRAN model can yield better results from both generator and discriminator where the generator generates better images with less training and discriminator better identifies the images as fake or real.

We conducted the experiment on MNIST dataset due to computation limitations and in future plan to extend our results to bigger datasets like CIFAR-10 which we assume will give more pronounced results due to their complexity compared to MNIST.

References

1. LeCun, Yann, Bengio, Yoshua, Hinton, Geoffrey: Deep learning. Nature **521**(7553), 436–444 (2015)
2. Hinton, G.E.: Deep belief networks. Scholarpedia **4**(5), 5947 (2009)
3. LeCun, Y., Haffner, P., Bottou, L., Bengio, Y.: Object recognition with gradient-based learning. In: Shape, Contour and Grouping in Computer Vision. Lecture Notes in Computer Science, vol 1681. Springer, Berlin, Heidelberg (1999)
4. Im, D.J., Kim, C.D., Jiang, H., Memisevic R.: Generating images with recurrent adversarial networks. (2016). arXiv:1602.05110
5. Dosovitskiy, A., Springenberg, J.T., Brox, T.: Learning to generate chairs with convolutional neural networks. In: CVPR, pp. 1538−1546 (2015)
6. Gregor, K., Danihelka, I., Graves, A., Rezende, D.J., Wierstra D.: Draw: a recurrent neural network for image generation. In: Proceedings of the International Conference on Machine Learning (ICML) (2015)
7. Radford, A., Metz, L., Chintala. S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of the International Conference on Learning Representations (ICLR) (2015)
8. Denton, E.L., Chintala, S., Fergus, R.: Deep generative image models using a Laplacian pyramid of adversarial networks. In: Advances in Neural Information Processing Systems, pp. 1486−1494 (2015)
9. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Advances in Neural Information Processing Systems, pp. 2234−2242 (2016)
10. Ranzato, M., Mnih, V., Susskind, J.M., Hinton, G.E.: Modeling natural images using gated mrfs. IEEE Trans. Pattern Anal. Mach. Intell. **35**(9), 2206–2222 (2013)
11. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
12. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**(5786), 504–507 (2006)
13. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009, pp. 248−255. IEEE (2009)
14. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. (2015). arXiv:1508.06576
15. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: Proceedings of the Neural Information Processing Systems (NIPS) (2013)
16. Sohl-Dickstein, J., Weiss, E.A., Maheswaranathan, N., Ganguli S.: Deep unsupervised learning using nonequilibrium thermodynamics In: Proceedings of The 32nd International Conference on Machine Learning (2015)
17. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio Y.: Generative Adversarial Networks. In: Advances in Neural Information Processing Systems, pp. 2672− 680 (2014)
18. Mirza, M., Osindero, S.: Conditional generative adversarial nets. In: Proceedings of the Neural Information Processing Systems Deep learning Workshop (NIPS) (2014)

19. Li, C., Wand, M.: Precomputed real-time texture synthesis with markovian generative adversarial networks. In: European Conference on Computer Vision, pp. 702−716. Springer International Publishing (2016)
20. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of ICML, vol. 30 (2013)
21. Kingma, D.P., Ba, J.L.: Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR) (2015)

A Multiphase Pre-copy Strategy for the Virtual Machine Migration in Cloud

Ruchi Shukla, Rajeev Kumar Gupta and Ramgopal Kashyap

Abstract Cloud Computing is a fastest growing era of computer science. It provides computing services through the internet. With the huge demand of the cloud resources over past few decades, load balancing is the major issue for the cloud provider. To achieve the proper load balancing virtual machine (VM) migration is used. Over past few decades, several VM migration approach have been proposed, but main challenging issue in these approach is to diminish the downtime (time for which VM is unavailable) and total migration time (time required to finish the migration process). To handle mentioned issues, this paper introduced Multi phase pre-copy based live migration approach. It works in multi phases such as: In first phase, it transferring all memory pages. In second phase, it maintain the history of each pages and according to this it takes a decision whether to send this page or not. After the second phase AR forecasting approach is use to predict the page behavior in the next interval which helps to determine either to send this page or not. CloudSim simulation tools is utilized to estimate the overall performance of the process. Experiment result says that proposed approach minimize Total migration Time (TMT), Downtime (DT).

1 Introduction

Cloud computing is an internet based model that offer computing services like CPU, memory, storage area, software services and other devices on demand. Services that can be distributed by several users through using of high speed

R. Shukla · R. K. Gupta (✉) · R. Kashyap
Department of CSE, Sagar Institute of Science and Technology,
Bhopal 462036, Madhya Pradesh, India
e-mail: rajeevmanit12276@gmail.com

R. Shukla
e-mail: shuklaruchi39@gmail.com

R. Kashyap
e-mail: ramgopalkashyap@sistec.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_43

networks. It is an utility computing in which customer pays only for used resources [1].Cloud computing can be deployedin three forms: Private, Public and Hybrid cloud.Private cloud provide secure cloud based environment and it delivered services only for specified client. Public cloud provides the services through the internet for all users, Google App Engine is the example of public cloud. Public cloud services are sold on demand. Hybrid cloud is a mixture of one or more cloud.

Virtualization is an important concept of cloud computing. Virtualization is the procedure of partitioning a one physical machine into number of physical machine and run on concurrently [2]. It shares the same physical resources. In modern computing installations virtualization has become a main facility and it interpret the opportunities for hardware utilization and application isolation and also simplifying resource allocation and management [3]. VM migration is main property of virtualization [4]. It permits to moves VM between the server. Three types of methods are available for migrating VM:

(i) **Pre copy**: It is iterative live migration approach which transfer all VM data (VM pages) to the target VM. Only those pages which are modified during last round of iterations are send in the next iteration [5]. This process is repeated until it not reached the extreme number of iterations, then close the virtual machine and send entire memory pages and processor state to the target physical machine, then begin the virtual machine at the target physical machine.

(ii) **Post copy**: In this process firstly it prevents the VM running on source physical machine and it transfer only minimum memory pages and processors states that necessitate for starting the virtual machine on destination physical machine.Transferring remaining memory pages through demand paging.

(iii) **Stop and copy**: It is the part of the Non live migration method. In this process, source VM is stop at source host and all memory pages are moved to the target host. When overall pages are reach to the target host, VM is resumed.

One of the prime feature that creates virtualization attractive is that live migration. Live migration is a procedure, it achieve a running VM and migrate it from one PM to another. It provides seamless migration for online services. This authorize the separation between user and data center operator. The functionality of live migration is provided by vendors like as Open VZ, Xen, VMware, KVM, and Hypervisor [6]. The impact of live virtual machine migration are measured in the following metrics as:

i. **Downtime**: It consists of the time when the VM on source host is suspended.
ii. **Total Migration Time (TMT)**: Total time acquired by the system to migrate the VM.
iii. **Total Execution Time**: It is define as the sumation of down time and TMT.

2 Literature Survey

W. Cui, M. Song [7], proposed a Matrix bitmap algorithm approach. This algorithm uses several bitmaps. The bitmap formation allocate the statistics for pages to decide either this page sent or not for one iteration. Many bitmaps are collected in iteration. MAP_LEN is a consistent amount. It defines the number of pages these are collected at one iteration. MAP_LEN measures the length of the matrix bitmap. It is used to gather the dirty bitmap. This dirty bitmap is used for MAP_LEN times and it takes a decision either this page sent or not for iteration. This bitmap to show the improve states of the equivalent pages. Success of algorithm relay on variable MAP_LEN.

S. Sharma and M. Chawla [8], proposed the three phase optimization method (TPO), it migrate the various memory pages in every iterations and also reduces it. This method is performed in three phases such as:

First Phase: Send entire data (pages) to the target VM.

Second Phase: In this phase consider the 2 to n-1 iterations and in this phase all modified pages are migrate iteratively. As a result the similar pages are transferred many times is minimized by using historical statistics structure of modified pages and it also reduces the updated pages and classify in two groups: frequently updated pages and non-frequently updated pages.

Third Phase: It contain stop and copy approach and migrate all remaining pages and also reduces those data that are transferred in previous iterations during migration by applying Run Length Encoding (RLE) compression method. Equations 1 and 2 are used to find the TMT and TPT respectively.

Total migration time: It is define as:

$$T_{mig} = \sum_{i=0}^{n} T_i, \tag{1}$$

where, T_i represents time required in ith iteration.

Total Pages Transferred: It is define as:

$$V_{mig} = \sum_{i=0}^{n} V_i \tag{2}$$

where, V_i represents total bits transferred in ith iteration and n is the total number of iterations. In this method, each phase efficiently minimizes the TPT, downtime and TMT appropriately.

Bolin Hu et al. [9], present a pre-copy approach that is based on time series. It uses an historical bitmap array which detect the frequently modified pages and stores the last historical statistics pages.

Haikun Liu et al. [10] proposed a migration strategy named full system trace and replay process. In this approach, it shows the creation, implementation and

estimation of checkpoint during the VM migration. Check pointing/recovery and trace/replay both methods are used for recapturing the system state. Firstly it start from checkpoint of previous state moving ahead to pass the exact state by using the log. Replaying is a system that continually logs the non-deterministic events of the system and impact the system data processing. These log files advisor the system that it rerun from the checkpoint. Most events are deterministic and does not be logged on the system that automatically re execute the system as it that replay has done. The trace file inserted at the source host repeatedly and migrate it to the destination host. This approach uses following formula to find TMT and total data transmitted (TDT). This approach decreases the TDT and TMT.

$$\text{TMT} = \sum_{i=0}^{n} t_i \frac{V_{ckpt}(1-\Phi^n)}{R_{trans}(1-\Phi)} \tag{3}$$

where, V_{ckpt} specify checkpoint file size and R_{trans} denotes log transfer rate.

$$TDT = V_{ckpt} + \sum_{i=1}^{n} V_{logi} = TMT * R_{trans} = \frac{V_{ckpt}(1-\Phi^n)}{1-\Phi} \tag{4}$$

where,V_{logi} define the log file of source VM.

3 Proposed Methodology

Live migration process is broadly utilized for server maintenance, fault tolerance, resource consolidation and load balancing. It is also transfer the memory, storage and connectivity. This paper proposed a multiphase Pre copy approach. It works in phases as show in Figure 1.

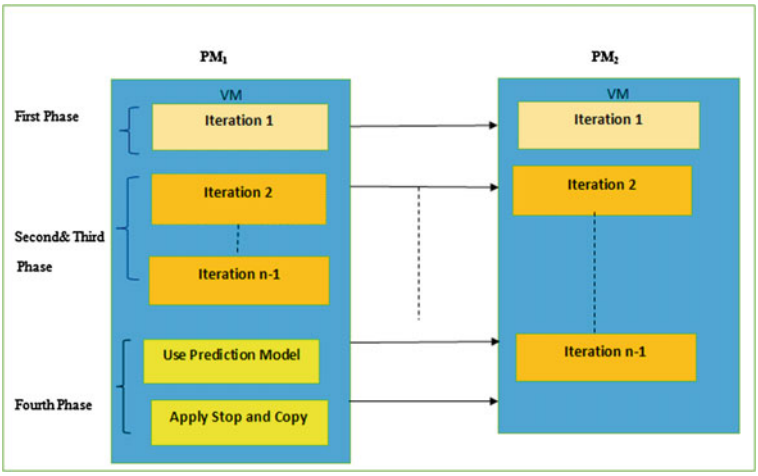


Fig. 1 Iterations in Pre-copy Method

3.1 First Phase

In this phase, firstly we will create the same VM (as candidate VM) on the destination PM then transferred all pages of the candidate VM from source PM to the VM running on target PM. Since, in pre-copy approach VM is remain active on the source PM so number of pages will be modified during the transmission. During the page transmission VM is still, it transfers the full image of VM from source host to target host. It contains historical bitmap structure. It utilize a formation of historical bitmap structure which recognize the regularly modified pages and also capture the several previous historical pages based on this, it will be taken a decision this page is sent or not in current iterations.

3.2 Second Phase

Since, several pages are modified in the 1st iteration due to write intensive workload so if all dirty pages are transferred then it will lead to increase the TMT and down time. To diminish the total pages transferred in the subsequent iteration our approach maintain history of every page in send_to and skip_to bitmap. Page is send to the next phase only when when send_to = 1 and skip_to = 0, otherwise page is not send in the current iteration. The certain bit 1 specify that the corresponding page is modified. As shows in Table 1.

We utilize historical dirty pages send_to_h array to identify this page is send or not in this iterations. send_to_h bitmap array saves the values of past statistics of send_to bitmap. Pages are categorized in two category: highly modified page and least modified page. If the equation is satisfied then we recognize p is to be high dirty page. When number of one's is more than number of zeros then page is declare as a highly modified page and should not be transmitted in this iteration and sent in last iteration. Otherwise we suppose that the page is to be category of low dirty page and sent it to the destination host.

$$K = \sum_{i=1}^{N} p \in \text{send\_to\_}h[i] \tag{5}$$

where N is the size of array, K indicate number of times page is updated in n iterations.

i. send_to: It include the pages updated in last iteration.
ii. skip_to: It include updated pages in present iterations marked those pages that can be skipped in this iteration.
iii. fix_to: It contains those pages which require to be migrated in the previous iterations.

Table 1 Page Transfer Strategy

send_to	**0**	**0**	**1**	**1**
skip_to	**0**	**1**	**0**	**1**
Send or Not	**N**	**N**	**Y**	**N**

3.3 Third Phase

In this phase, it contain frequently modified pages. Through using the historical bitmap send_to_h[i] array of modified pages are partitioned in two groups G1 and G2. The less frequently modified pages are conserved in G1 to threshold and highly modified pages are conserved in G2. At every iteration less frequently modified pages of G1 are transmitted, whereas the highly modified pages of G2 contain those pages which are updated more times, due to this, the downtime is increasing. In this phase we minimize the size of G2 (highly modified pages) by applying forecasting technique, which will predict the dirty pages by using autoregressive (AR) approach. AR is a basic linear time series prediction algorithm in which the present value can be indicated by the sum of a linear combination of several prior values and an error *term*. The general expression of AR(p) model can be denoted as:

$$\mathrm{AR}(p)\colon (X_t) = \mathrm{m}_0 + \mathrm{m}_1 x_{t-1} + \mathrm{m}_2 x_{t-2} + \cdots + \mathrm{m}_\mathrm{p} x_{t-p} + e\mathrm{rr}_\mathrm{t} \tag{6}$$

where $\{x_1, x_2 \ldots . x_t\}$ is the time series, p is the order of AR model and err_t is white noise. The linear dependence can be generalized so that the present value of the series, X_t, depends not only on $x_{t-1,}$ but also on the previous p values (lags), $x_{t-2,\ldots\ldots,} x_{t-\mathrm{p}.}$ Thus an autoregressive process of order p is obtained. Migration method performance is estimated through TMT, downtime and total page transferred (TPT). These parameter are calculated as followed:

$$\mathrm{TPT} = \frac{\text{Total number of pages}}{\text{Time required to send single page}} \tag{7}$$

$$\mathrm{TPT} = \sum_{i=1}^{n} T_i \tag{8}$$

where, Ti is the time taken by the ith iteration.

3.4 Forth Phase: Threshold Calculation

In this phase, the threshold is calculate by the modification rate. This threshold contain the dynamic nature. The value of threshold is vary because each page is modify at every iteration. Its value is totally depend on page modification rate. We calculate the threshold by taking the average of maximum page modification rate and minimum modification rate, by using following formula:

$$T_1 = [[(\max[\textit{page modification rate}] + \min[\textit{page modification rate}]) \div 2]] \tag{9}$$

Suppose we have 16 pages, each page will be migrated in every 2 s then per page contain 30 number of iterations. The value of each page is depend on modification rate because each page is modified in any iteration.

Table 2 Modification history of 31 pages

Number of Pages																
Number of Iteration		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	1	1	1	0	0	1	1	1	1	1	1	0	0	0	1	1	1
	2	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1
	3	1	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1
	4	1	1	1	0	0	1	1	1	1	0	1	1	1	1	0	1
	5	1	1	1	1	0	0	1	1	0	1	1	1	0	0	1	1
	6	0	0	0	1	1	1	0	1	0	1	0	0	1	1	1	0
	7	0	1	1	0	1	1	1	0	1	1	1	0	1	0	1	1
	8	1	1	0	1	1	0	1	1	1	0	0	1	1	1	0	1
	9	1	0	0	1	1	1	0	0	1	1	1	0	0	1	0	1
	10	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	0
	11	1	1	1	1	1	1	0	0	1	1	1	0	1	0	1	1
	12	1	0	0	1	0	0	1	1	1	0	1	0	1	1	0	1
	13	0	1	1	1	1	1	0	1	1	0	1	0	1	1	0	1
	14	1	1	0	1	0	1	1	0	1	1	1	0	1	1	1	0
	15	1	1	1	0	1	1	0	1	1	0	0	1	1	1	0	1
	16	1	0	1	1	0	0	0	1	1	1	1	1	0	0	1	0
	17	0	1	1	0	0	1	1	1	0	1	0	1	0	1	1	1
	18	1	0	1	1	0	0	1	1	0	1	0	1	1	1	0	1
	19	1	1	0	1	1	0	1	1	1	1	0	0	1	1	1	1
	20	0	1	0	1	1	1	0	0	1	1	1	0	0	1	1	1
	21	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1
	22	1	1	1	1	1	0	0	0	1	1	1	1	0	1	0	0
	23	1	1	1	0	0	1	1	1	1	0	0	0	1	1	1	0
	24	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0
	25	1	0	1	1	0	1	1	1	0	0	0	1	1	1	0	0
	26	1	1	0	0	0	1	1	1	0	1	0	1	1	1	0	1
	27	0	1	1	1	1	0	1	0	1	1	1	1	0	0	1	1
	28	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1
	29	1	0	1	1	1	1	0	0	1	1	1	0	1	1	1	0
	30	1	1	0	0	1	1	0	1	0	1	0	0	1	1	1	1
	31	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1

Page Modification Rate(Max):14
Page Modification Rate (Min): 8
Threshold $\boldsymbol{T_I}_{+}$ 0.6875

When Eq. 1 is satisfied then we calculate modification rate of pages by threshold. In every iterations T_I is determined by using send_to_h[i] array. Table 2 shows the history of 31 pages which is used to calculate the threshold.

This phase contain the frequently modified pages till last iterations and minimize the unnecessary transference of similar pages frequently and minimize TPT along with migration time.

After this phase, the migrated VM start activities on this, when the migration process will be completed:

$$T_i = \frac{Pi}{Di} \tag{10}$$

Where, T_i represents time taken by iteration i, P_i represents pages dirtied in iteration i, and D_i represents page dirty rate.

4 Results Analysis

CloudSim [11] is the tool which is used for measuring the performance of our approach. For this purpose it is evaluate against existing approach named "Three phase optimization method for precopy based VM live migration [8]" in term of TMT and down time. Both approaches are implemented in CloudSim simulator. Following environment is created during the experiment:

For the proposed two phase pre copy approach

Number of VM created during the experiment = 2, Pages in each VM = 15, 20, 25
Previous Recorded data for each page = 8, Iterat_threshold = 7

For the already exist efficient virtual machine migration approach

Number of VM created during the experiment = 2, Pages in each VM = 15, 20, 25
Previous Recorded data for each page = 8, Iterat_threshold = 7
Dirty_Threshold = 5

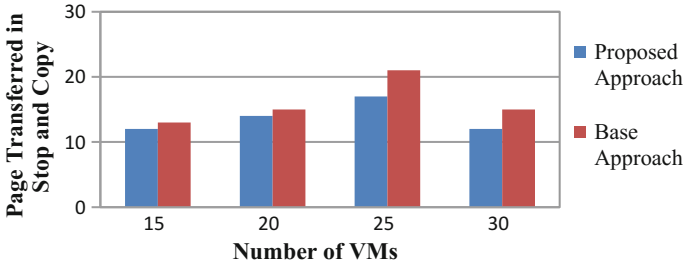


Fig. 2 Total page transferred in stop and copy phase

Figure 2 illustrates that our design approach transferred a smaller amount number of pages in stop and copy phase as compare to competitive approach.

Figure 3 illustrates that proposed approach minimize the downtime as compare to competitive approach.

Figure 4 shows that our designed approach transferred a smaller amountof pages in migration contrast to the base approach.

Figure 5 illustrates that proposed approach required less time to migrate the as compare to competitive approach.

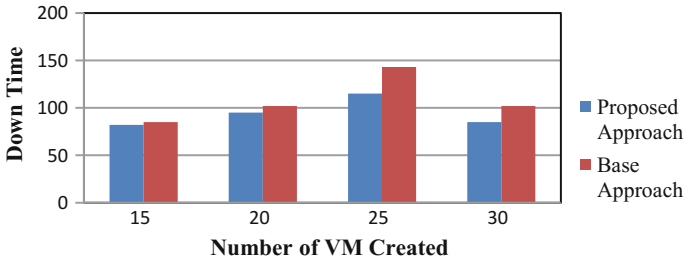


Fig. 3 Total downtime during the VM migration

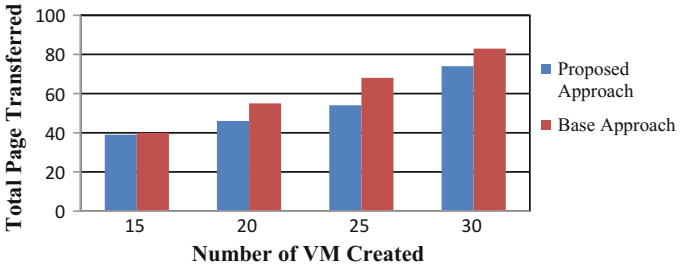


Fig. 4 Total page transferred during the VM migration

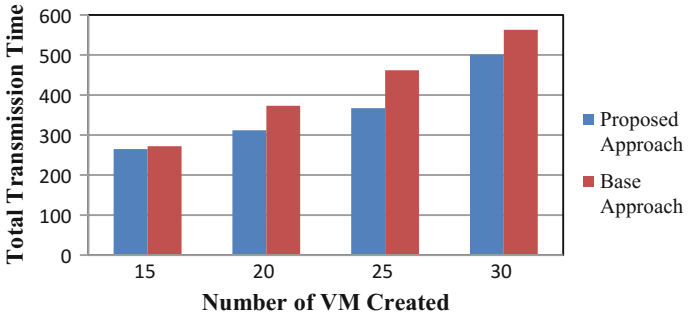


Fig. 5 Total migration during the VM migration

5 Conclusion and Future Work

As cloud provides, services to the multiple user by creating the VM for each user. Hence, number of virtual machines are created into a single PM which may lead a situation where some VM are overloaded and other are unloaded. So to deal with this situation VM migration approach is triggered. This paper proposed and multi-phase pre-copy VM migration approach which not only minimize the migration time also minimize the number of transferred pages. Proposed approach is simulate via CloudSim simulation tool. Experiment result illustrate that proposed approach diminish TMT and down time by minimizing the number of transferred pages. This approach can also be implemented in real cloud.

References

1. SangeetaSharma, Meenu Chawla, "A technical review for efficient virtual machine migration", in IEEE International conference on cloud & ubiquitous computing & emerging technologies (CUBE), pp. 20–25, 2013
2. Kaur, Pankajdeep, Rani, Anita: Virtual Machine Migration in Cloud Computing. International Journal of Grid Distribution Computing **8**(5), 337–342 (2015)
3. Rajeev Kumar Gupta and R.K. Pateriya," Balance Resource Utilization (BRU) Approach for the Dynamic Load Balancing in Cloud Environment by using AR Prediction Model ", Journal of Organizational and End User Computing (JOEUC), Volume 29, Issue 4, Article 2, 2017
4. W. Lio, Tao Fan et.al, " Live migration of virtual machine based on recovering system and CPU scheduling", In Proc. ITAC, China, pp. 1088-1096, 2011
5. C. Clark, K. Fraser et al., "Live Migration of Virtual Machines", Proc. The 2nd conference on Symposium on Networked Systems Design & Implementation, pp. 273-286, 2005
6. F. Ma, F. Liu, and Z. Liu, "Live Virtual Machine Migration Based on Improved Pre-copy Approach", In Proc. IEEE Int'l Conf. on Software Engineering and Service Sciences (ICSESS). IEEE, Vol.7 No.10, pp.230-233, 2010
7. W. Cui and M. Song, "Live memory migration with matrix bitmap algorithm", in Proceedings of the IEEE 2nd Symposium on Web Society (SWS '10), pp. 277–281, August 2010
8. S. Sharma, M. Chawla, "Three phase optimization method for precopy based VM live migration", SpringerPlus journal, 2016
9. B. Hu, Z. Lei, Y. Lei, D. Xu, and J. Li, "A Time-Series Based Precopy Approach for Live Migration of Virtual Machines", IEEE 17th International Conference on Parallel and Distributed Systems, pp. 947-952, 2011
10. H.Liu, H. Jin, X. Liao, L. Hu, and C. Yu, "Live migration of virtual machine based on full system trace and replay," in Proceedings of the 18th International Symposium on High Performance Distributed computing (HPDC'09), pp.101-110, 2009
11. R. Calheiros et al. " CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services", 2011

Fuzzy C-Means-Based JPEG Algorithm for Still Image Compression

Vanitha Kakollu, G. Narsimha and P. Chandrasekhar Reddy

Abstract The magnitude of information that is conveyed via the Internet has escalated more rapidly during the last few decades. Image compression provides an appreciable way to dwindle the image size. For the conservation of bandwidth during still image compression, JPEG is advisable. As one can gather and communicate the images early with reduced size, we planned an innovative JPEG compression algorithm (de Queiroz, Member IEEE, IEEE Trans Image Process 7 (12):1661−1672, 1998, [1] Prakash, IEEE Member, Mitchell, IEEE Fellow, Stepneski, IEEE Int Conf Image Process 3:494−497, 2001, [2], Sreelekha, Sathidevi, An improved JPEG compression scheme using human visual system model, 2007, [3]) with fuzzy c-means-based clustering in this work. The anticipated algorithm produces enhanced results in comparison with standard algorithms in the form of some estimation parameters like MSE, PSNR, and quantity of bits transmitted (Egger, Li, IEEE Int Conf Image Process 3:326−330, 1994, [4]). The speed is increased with the proposed JPEG algorithm, thereby diminishing the memory that is necessary for hoarding the encoded bits. The reassembled image after decompression is as similar as the image which was sent as an input.

V. Kakollu (✉)
Department of CSE, JNTU, Hyderabad, India
e-mail: vanithajntuh@gmail.com

G. Narsimha
Department of CSE, JNTUH College of Engineering, Sultanpur, Sangareddy, India
e-mail: narasimha06@gmail.com

P. Chandrasekhar Reddy
Department of Electronics and Communications Engineering, JNTU, Hyderabad, India
e-mail: drpcsreddy@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_44

1 Introduction

The two-dimensional image f(x, y) is sent to an encoder which generates a group of symbols. We then transmit this through a channel, and the encoded image is now the input to the decoder. The recreated image f'(x,y) is generated by the decoder. The output f'(x, y) is an accurate imitation of f(x, y) in lossless compression. Else it means that there is some misconception present in the reenacted image.

The Joint Photographic Experts Group (JPEG), the team that shaped the JPEG standard, is a noticeable compression proposal. Not just using less memory, and also the content in the regenerated image in a JPEG compression appears very much identical [5–7]. Though the quality is reduced with JPEG compression, the image would look nearly as identical as the unprocessed image.

The JPEG algorithm wipes out the components, having high frequency, which cannot be acknowledged by the human eye.

1.1 JPEG Algorithm

The comparatively straightforward JPEG algorithm can be acknowledged via the steps given below:

1. We can divide the obtained image into 8 by 8 pixel chunks. Now, add zeros in unfilled pixels around the edges if the dimension of the image is not multiplied by 8 exactly.
2. Obtain the image data for each 8 × 8 pixel block in such a way that there are values to discriminate the color at every pixel.
3. For each 8 × 8 block, the DCT is calculated.
4. The DCT of each 8 × 8 pixel block is multiplied with a normalization mask so as to formulate a more number of pixel values as zero from the DCT matrix.
5. Normalization can renounce many of the components having high frequency. By traversing in a zigzag fashion and grouping them in a one-dimensional array, we can assort the important two-dimensional normalized DCT coefficients. In a one-dimensional array, there are two types of DCT coefficients: direct current element and alternating current elements [8]. The AC components are coded with Huffman coding.
6. Decompression is the reversal of compression. The compressed bit stream is decoded using Huffman coding; the normalized values of DCT are calculated. Later, organize the DCT values in a 2-D array in a zigzag fashion. By multiplying the encoded DCT values with normalization coefficients, we get the decoded DCT values. Next, we execute an IDCT on the de-normalized DCT array. The decoding process generates an image block which will not be indistinguishable to its corresponding original image block that was used while encoding.

The number of prolific DCT coefficients will be supplementary if there is an enormous difference in the pixel values. Else only first few DCT coefficients will be remarkable and the others will be zero. The image gets smoothened by applying filters, and as an effect, the variation in the pixel values of a block is reduced.

The designed compression algorithms are executed in the following procedure.

(1) Step 1: We have to segregate the given image into non-overlapped 8 × 8 pixel blocks.
(2) Step 2: Deduct the quantity G/2 (where the gray-level resolution is denoted by G) for the 64 samples of every 8 × 8 pixel block to level shift them.
(3) Step 3: Calculate the DCT of every 8 × 8 block.
(4) Step 4: By making use of standard normalization matrix, the DCT blocks are normalized.
(5) Step 5: By using fuzzy c-means clustering, the regularized blocks are assembled into clusters.
(6) Step 6: The cluster centers and indexes are encoded and sent to the receiver.
(7) Step 7: The decoding procedure of cluster centers and indexes is done by the receiver.
(8) Step 8: The receiver forms the output based on the cluster centers and indexes.
(9) Step 9: Compute the error (in terms of PSNR) between the reenacted image and the original image (Fig. 1).

2 Fuzzy C-Means Algorithm

This method is considered in this work for clustering the images. Clustering formulates the basis for the analysis of any image. A good amount of the clustering algorithms consider the cases of fully damaged and fully undamaged, and thereby, misclassification results into the picture as the partly damaged tissues are being neglected. Therefore, to overcome these impacts, effective methodologies are to be developed. Fuzzy c-means algorithms are one such clustering method which is specifically meant for identifying the above challenges.

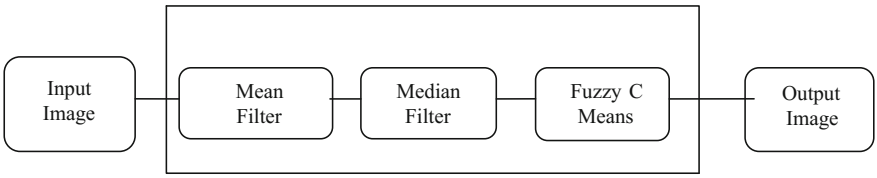


Fig. 1 Fuzzy C-Means-Based JPEG compression

2.1 Fuzzy C-Means Algorithm

Step 1: Select the primary fuzzy pseudo-separation; i.e., assign values to all the W_{ij} (weight assigned to each point x_i in the cluster j).
Step 2: Repeat.
Step 3: Estimate the centroid of each cluster using the fuzzy pseudo-partition, subject to centroid, that minimizes the fuzzy sum of squared error (SSE).

$$SSE(C_1, C_2, \ldots\ldots.., C_k) = \sum_{j=1}^{k} \sum_{i=1}^{m} w_{ij}^{p} dist\left(x_i, c_j\right)^2 \tag{2.1.1}$$

Step 4: Recalculate the fuzzy pseudo-partition, i.e., W_{ij}, by equation,

$$W_{ij} = \frac{1/dist\left(x_i, c_j\right)^2}{\sum_{q=1}^{k} 1/dist\left(x_i, c_j\right)^2} \tag{2.1.2}$$

Step5: Until centroids do not change.

3 Implementation of Planned JPEG Algorithms

The proposed algorithm is implemented on images of various sizes (source of images: waterloo image repository which is open to use). The consideration of results delegates that the estimated compression techniques are extremely a prominent alternate because they are proved to be better in terms of quality metrics like MSE and PSNR. N1 is the scope of information hauling units required to imply uncompressed datasets, and N2 is the number of entities in the encoded dataset.

$$\mathrm{CR} = \mathrm{N1}/\mathrm{N2} \tag{3.1}$$

The original image and the reconstructed image are similar to each other if we use lossless compression algorithms as they not only wipe out redundancy but also remove the redundancy present in the data and they even secure all the information that is present in the input image. In lossy compression algorithms [4], there is a variation between input image and rationalized image. And therefore, higher compression is achieved. We can use the fidelity criteria of objective or subjective type for comparing the original image and reprocessed image.

The measurement of the image quality is an imperative implication in image processing. In many applications, estimation is a compulsion for the excellence of the image. The image quality judgment by human is not sufficient. Therefore, some more metrics like PSNR and MSE are needed. PSNR is the most specialized image quality metric. The variations, in the image which is restructured and the image which we sent as input, will be small when the PSNR value is high.

4 Results

This paper presents the evaluation of the proposed approach with the standard JPEG compression. The intended approach exemplify improved results evaluated to the JPEG (Figs. 2, 3, 4, 5, 6, 7, 8, and 9), (Tables 1, 2, 3, 4, 5, 6, 7, 8, and 9).

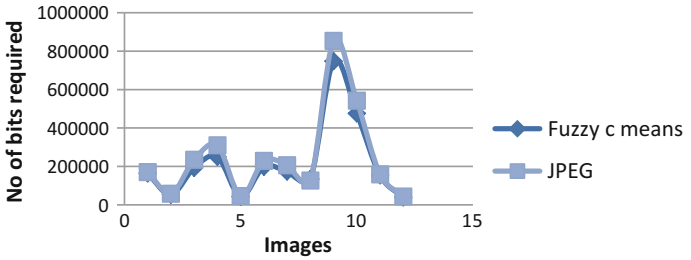


Fig. 2 Comparison between fuzzy c-means and JPEG in terms of number of bits transmitted

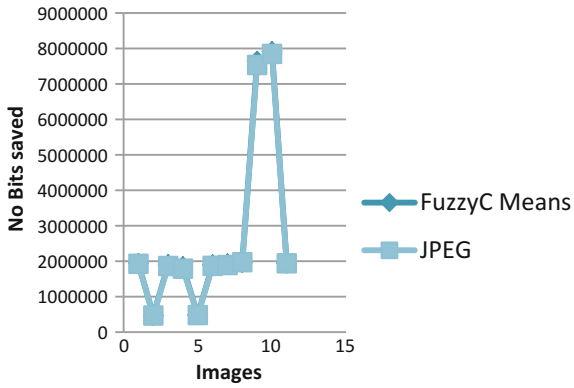


Fig. 3 Comparison between fuzzy c-means and JPEG in terms of number of bits saved

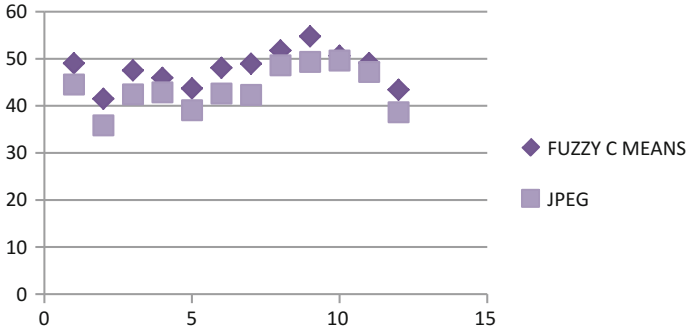


Fig. 4 Comparison between fuzzy c-means and JPEG in terms of PSNR

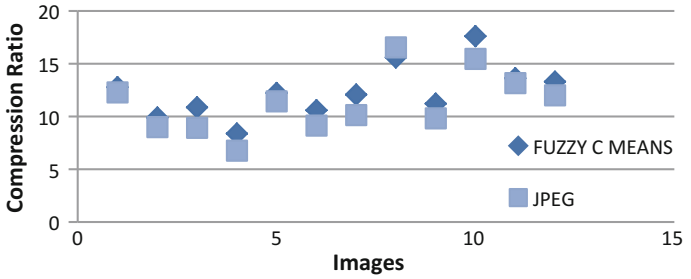
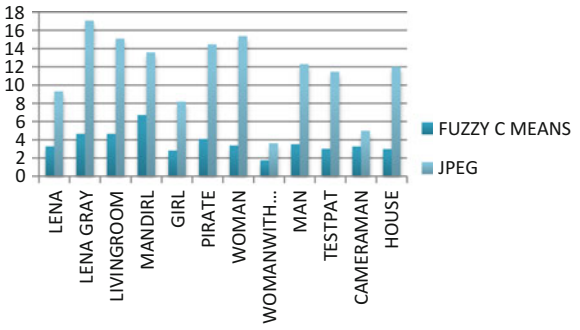


Fig. 5 Comparison between fuzzy c-means and JPEG in terms of compression ratio

Fig. 6 Comparison between fuzzy c-means and JPEG in terms of MSE



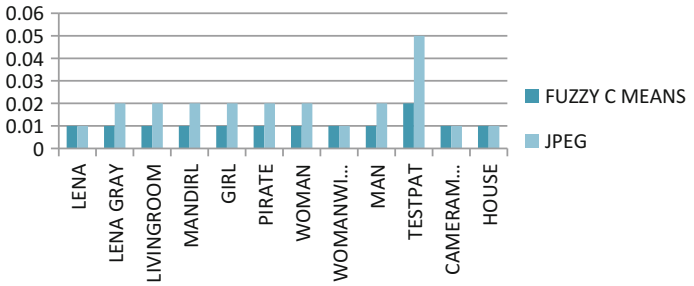


Fig. 7 Comparison between fuzzy c-means and JPEG in terms of NAE

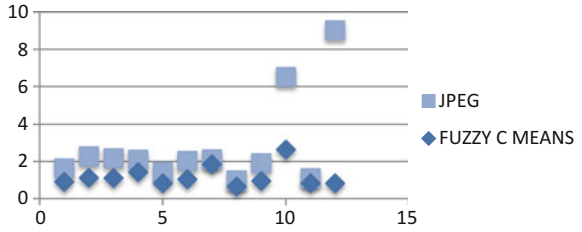


Fig. 8 Comparison between fuzzy c-means and JPEG in terms of AD

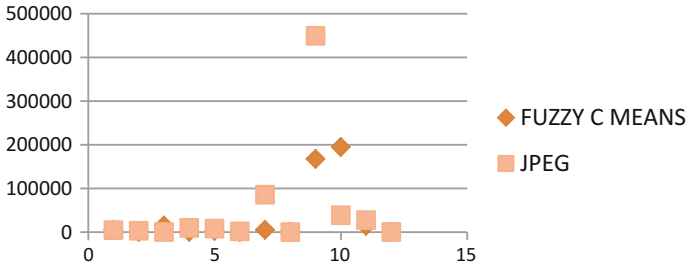


Fig. 9 Comparison between fuzzy c-means and JPEG in terms of SC

Table 1 Comparison between fuzzy c-means and JPEG in terms of number of bits transmitted

Image	Fuzzy c-means	JPEG
LENA	164142	170767
LENA GRAY	52907	58267
LIVINGROOM	192829	234701
MANDIRL	250007	310197
GIRL	42884	45824
PIRATE	198024	229244
WOMAN	173562	206038
WOMANWITHBLACKHAIR	134380	126580
MAN	747462	853840
TESTPAT	476128	542169
CAMERAMAN	153855	159198
HOUSE	39407	43646

Table 2 Comparison between fuzzy c-means and JPEG in terms of number of bits saved

Image	Fuzzy c-means	JPEG
LENA	1933010	1926385
LENA GRAY	471381	466021
LIVINGROOM	1904323	1862451
MANDIRL	1847145	1786955
GIRL	481404	478464
PIRATE	1899128	1867908
WOMAN	1923500	1890314
WOMANWITHBLACKHAIR	1962772	1970572
MAN	7641146	7534768
TESTPAT	7912480	7846439
CAMERAMAN	1943297	1937354
HOUSE	484881	480642

Table 3 Comparison between fuzzy c-means and JPEG in terms of PSNR

Image	Fuzzy c-means	JPEG
LENA	49.04	44.50
LENA GRAY	41.48	35.84
LIVINGROOM	47.50	42.40
MANDIRL	45.92	42.86
GIRL	43.67	39.04
PIRATE	48.07	42.58
WOMAN	48.91	42.32
WOMANWITHBLACKHAIR	51.76	48.60
MAN	54.75	49.31
TESTPAT	50.63	49.62
CAMERAMAN	49.05	47.20
HOUSE	43.40	38.61

Table 4 Comparison between fuzzy c-means and JPEG in terms of compression ratio

Image	Fuzzy c-means	JPEG
LENA	12.77	12.28
LENA GRAY	9.90	8.99
LIVINGROOM	10.87	8.93
MANDIRL	8.38	6.76
GIRL	12.22	11.44
PIRATE	10.59	9.14
WOMAN	12.07	10.13
WOMANWITHBLACKHAIR	15.60	16.56
MAN	11.22	9.82
TESTPAT	17.61	15.47
CAMERAMAN	13.63	13.17
HOUSE	13.30	12.01

Table 5 Comparison between fuzzy c-means and JPEG in terms of MSE

Image	Fuzzy c-means	JPEG
LENA	3.27	9.31
LENA GRAY	4.66	17.07
LIVINGROOM	4.66	15.09
MANDIRL	6.71	13.58
GIRL	2.82	8.18
PIRATE	4.09	14.47
WOMAN	3.37	15.37
WOMANWITHBLACKHAIR	1.75	3.62
MAN	3.51	12.29
TESTPAT	3.01	11.45
CAMERAMAN	3.26	4.99
HOUSE	2.99	12.01

Table 6 Comparison between fuzzy c-means and JPEG in terms of NAE

Image	Fuzzy c-means	JPEG
LENA	0.01	0.01
LENA GRAY	0.01	0.02
LIVINGROOM	0.01	0.02
MANDIRL	0.01	0.02
GIRL	0.01	0.02
PIRATE	0.01	0.02
WOMAN	0.01	0.02
WOMANWITHBLACKHAIR	0.01	0.01
MAN	0.01	0.02
TESTPAT	0.02	0.05
CAMERAMAN	0.01	0.01
HOUSE	0.01	0.01

Table 7 Comparison between fuzzy c-means and JPEG in terms of AD

Image	Fuzzy c-means	JPEG
LENA	0.90	1.63
LENA GRAY	1.12	2.28
LIVINGROOM	1.11	2.18
MANDIRL	1.42	2.10
GIRL	0.82	1.42
PIRATE	1.04	2.03
WOMAN	1.84	2.12
WOMANWITHBLACKHAIR	0.65	0.98
MAN	0.95	1.91
TESTPAT	2.64	6.51
CAMERAMAN	0.83	1.10
HOUSE	0.82	9.02

Table 8 Comparison between fuzzy c-means and JPEG in terms of SC

Image	Fuzzy c-means	JPEG
LENA	5535	4811
LENA GRAY	640	3138
LIVINGROOM	14995	413
MANDIRL	806	9963
GIRL	2994	8006
PIRATE	8	1769
WOMAN	5279	85713
WOMANWITHBLACKHAIR	909	0
MAN	167615	449624
TESTPAT	195085	38703
CAMERAMAN	14619	27657
HOUSE	990	530

Table 9 Comparison between JPEG and fuzzy c-means-based JPEG

Original Image	JPEG	Fuzzy C Means based JPEG
		
		
		
		
		
		

Source of images: waterloo image repository

5 Conclusion

Fuzzy c-means-based JPEG compression algorithm is projected and evaluated with standard JPEG algorithm in this paper. With this algorithm images can compressed with less encoded bits, and as a result, the loading and storing of these images take less time. As MSE of the proposed approach is low compared to regular JPEG and the perfect classification correctness is augmented with the estimated approach because of high PSNR. The projected approach shows better results than conventional JPEG in terms of image quality metrics like Compression ratio, NAE, AD, SC. The encoded bits which are desired to symbolize the compressed image are less compared to JPEG.

References

1. de Queiroz, R.L., Member IEEE: Processing JPEG-compressed images and documents. IEEE Trans. Image Process. **7**(12), 1661−1672 (1998)
2. Prakash, R., Member IEEE, Mitchell, J.L., IEEE Fellow, Stepneski, D.A.: Enhanced JPEG compression of documents. IEEE Intl. Conf. Image Process. **3**, 494−497 (2001)
3. Sreelekha, G., Sathidevi, P.S.: An Improved JPEG Compression Scheme using Human Visual System Model, pp. 98−101. IEEE (2007)
4. Egger, O., Li, W.: Very low bit rate image coding using morphological operators and adaptive decompositions. IEEE Intl. Conf. Image Process. **3**, 326−330 (1994)
5. Ramesh, C., Venkateswarlu, N.B., Murthy, J.V.R.: Filter augmented JPEG compressions. IJCA **60**(17), 1−5 (2012)
6. Krishna, M., Srinivas, G., Prasad Reddy P.V.G.D.: Image smoothening and morphological operators based JPEG compression. JATIT **85**(3), 252−259 (2016)
7. Krishna, M., Prasad Reddy P.V.G.D., Srinivas, G.: A smoothening based JPEG compression for an objective image quality enhancement of regular and noisy images. IJAER **11**(6), 3799 −3804 (2016)
8. Xiangzhi, B., Fugen, Z.: Edge detection based on mathematical morphology and iterative thresholding. IEEE Intl. Conf. Image Process. **2**, 1849−1852 (2006)

Supportive Communication System for the Elderly Disabled People

Pushpa Kotipalli, E. R. Praveen Kumar, M. A. S. Mohan Raju and D. Murali Krishna

Abstract As age progresses, the elderly people experience speech impairment and physical disability due to several health problems like amyotrophic lateral sclerosis (ALS), osteoarthritis, Parkinson's and essential tremor. At some point in the disease progression, 60–80% of elderly people is unable to meet their daily communication needs using natural speech. The plight of large majority of older individuals by and large remains unnoticed. Here, we propose an assistive device for their communication to meet their daily needs. But, the shakiness in hands makes them unable to operate assistive devices. As eyes are not affected by most of the diseases, the tracking of eye movements can be employed as a tool for operating the assistive device. Therefore, we propose an eye movement-based supportive communication system for the elderly disabled people. The proposed method is proved through simulations results.

1 Introduction

By 2050, according to the United Nations, 20% of the world population will be aged over 60 years and around 130 million people will suffer from osteoarthritis, of which 40 million people will be severely disabled and become immobile. Also, as per the National Council on Aging report, at least one of the chronic health conditions like cancer, heart disease, stroke, Parkinson's disease, and other neurolog-

P. Kotipalli (✉) · E. R. Praveen Kumar · M. A. S. Mohan Raju · D. Murali Krishna
ECE Department, Shri Vishnu Engineering College for Women, Vishnupur, Bhimavaram, West Godavari 534202, Andhra Pradesh, India
e-mail: pushpak@svecw.edu.in

E. R. Praveen Kumar
e-mail: emani3815@svecw.edu.in

M. A. S. Mohan Raju
e-mail: surendra@svecw.edu.in

D. Murali Krishna
e-mail: ece_krishnad@svecw.edu.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_45

ical diseases is occurring 92% of the elderly population. Most of these diseases may result in speech impairment with which caretaker experiences frustration in understanding what the patient is saying [1].

The most popular neurological disorder is Parkinson's disease. As it is progressive in nature, the health of person with Parkinson's disease gradually worsen with time. Parkinson's disease foundation reports that there are more than 10 million Parkinson's disease-affected people worldwide. So far no cure was invented for Parkinson's disease. In general, men have 1.5 times more chances of getting Parkinson's disease as compared to women. At present, treatment is available to manage the symptoms only. Most of the elderly people are suffering from another type of movement disorder known as essential tremor. Shakiness in the hands is the first sign of essential tremor. Another problem faced by elders is the dexterity impairment which makes them so weak that they cannot do any sort of work requiring handiness. For example, the person with dexterity impairment does not face speech problem but it may become impossible for him make a phone call because his shaky hands do not support him to press phone buttons. This does not influence speech communication itself but makes it hard to make a phone call or use a wide range of other equipment.

Several such health problems in elderly people make them disable in lifting a small weight or operating even a switch. Therefore, elderly people with speech and body impairments must be provided with the best possible supportive communication system. Supportive communication means communicating with other person using some means other than speech such as sign language, symbols, pictures, or synthesized speech. Supportive communication is also known as "Augmentative and Alternative Communication" (AAC) [2]. The proposed research proposal considers the development of synthesized speech using electronic devices using gaze tracking techniques.

Present-day applications of gaze tracking include consumer products or research systems only. Even though gaze tracking can be used for several other applications which involve human and computer interaction, it is not used because of the two main drawbacks: (*i*) Present available gaze tracking systems are invasive, and (*ii*) they are very expensive. Systems based on non-invasive techniques like remote video-based gaze trackers [3] and head-mounted video-based eye trackers [4] are also available in the market, but their price falls in the range of Rs. 3,00,000−82,00,000 approximately. Main reasons for high price are (*i*) custom software implementation and (*ii*) use of specialized digital processor for achieving high performance. Considering these lapses in present technology, our proposed system aims at development of low-cost AAC system for the elderly people with speech impairment and disability in upper limbs.

Applications such as video conferencing, eye typing, driver alertness, and AAC require human–computer interaction which in turn requires human gaze tracking. Gaze tracking algorithms are broadly classified into three categories; they are (*i*) reflected light-based techniques, (*ii*) electric skin potential-based techniques, and (*iii*) contact lense-based techniques. Of these techniques, electric skin potential- and contact lense-based techniques are intrusive techniques, whereas reflected

light-based techniques are nonintrusive and they include pupil tracker, pupil and cornea reflection tracker, and Purkinje image tracker [5, 6]. Reflected light-based techniques use light reflection from different eye parts such as cornea, iris, and retina for gaze tracking. Main drawback of the reflected light-based techniques is that the tedious user-dependent parameter calibration is required for every user. Moreover, most of the reflected light-based techniques are monocular (2D) vision techniques, and real-time problem is 3D gaze tracking. Stringent constraints are imposed on the conversion of 3D gaze tracking problem into 2D gaze tracking. Stringent constraints like holding user's head quite still or fixing the apparatus relative to the user's head are very difficult to implement when the person is already suffering from health issues. Eye tracking systems are produced mainly by companies which include DynaVox, Prentke Romich, Quick Glance, Tobii Dynavox, and LC Technologies.

Keeping this in mind, we propose a low-cost supportive communication system for the elderly people who have lost their ability to move and speak, either temporarily or permanently through a traumatic accident.

Remaining paper is organized in the following way. Section 2 proposes supportive communication system. Section 3 discusses gaze data estimation methods. Section 4 describes the hardware implementation of the proposed system. Section 5 discusses the simulation results. Section 6 introduces concluding remarks.

2 Proposed Supportive Communication System

The video camera captures the movement of pupil in the eye. Based on the information provided by the video, the eye tracking and gaze estimation algorithms in the DSP estimate the gaze direction and location. Now, based on estimated gaze direction and location, the processor will highlight the image on the display system. The highlighted image will give different options for the blinking of the eye, and at the same time, the information related to the highlighted image will send to the speaker counted at the display system and also to the wireless transmitter. The wireless transmitter is used to send the information of the highlighted image to the electronic gadget available at the caretaker. As shown in Fig. 1, the wireless receiver will receive the transmitted signal and alert the caretaker.

2.1 Algorithm for Proposed System

1. The camera mounted on the screen in front of the user continuously tracks the face and locates his eyes. Then it calibrates the location of eyes of the user for further processing.
2. If the user stares at the screen for 3 s, the MENU with different OPTIONS will be displayed on the screen.

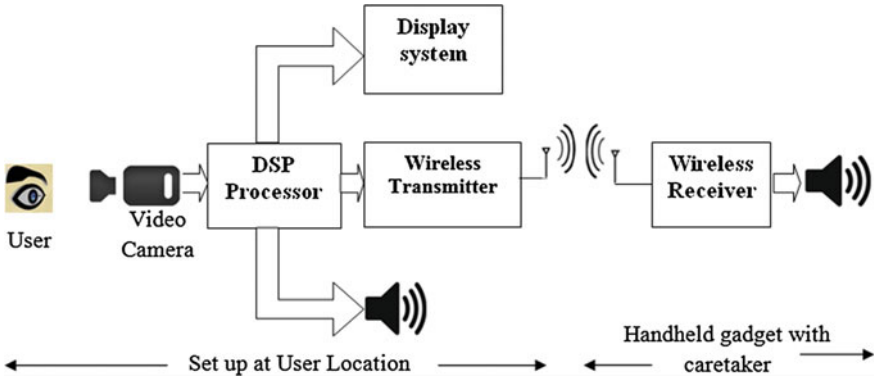


Fig. 1 Block diagram of proposed supportive communication system

3. Then the camera tracks the eye gaze for 3 s.
4. If the eye gaze is fixed for 3 s on an OPTION, it will be highlighted to indicate the selection. If not, go to Step 3.
5. If the highlighted OPTION is appropriate one, the user has to blink to open the OPTION. A new window will appear with sub-OPTIONs.
6. Steps 3−5 are repeated until he reaches the required OPTION.
7. If the highlighted OPTION is wrong, the user has to shift his eye gaze toward the appropriate one. Steps 3−7 are repeated until the required OPTION is selected.
8. If the caretaker is nearby, the audio output of the selected OPTION is obtained by blinking eyes twice.
9. If the caretaker is remotely located, by blinking eyes three times, the message can be sent to the handheld device carried by the caretaker.
10. In case of emergency, a buzzer can be switched on by blinking eyes continuously.

3 Gaze Data Estimation

There are three main ways in which the eye moves, that are usually taken into consideration when analyzing the gaze data: saccades, fixations, and smooth pursuit. A fixation is when the eye holds still, focusing on something. Saccades are "conjugated movements that change the visual axis of the eyes from one point of fixation to another with maximal speed." The smooth pursuit is used by the eye to steadily focus on and follow a small moving object to keep the image stationary. When using eye tracking as an input method, the "dwell time" is the amount of time

it takes after a person fixates on something before it becomes selected. The information gathered during eye tracking is often referred to as the "gaze data."

The point on the display screen at which the user stares for 3 s is called "gaze position." The gaze position is estimated with the following setup.

3.1 Gaze Position Estimation

The coordinates of the gaze position are determined by using the eye tracking and gaze estimator algorithms. As shown in Fig. 2, the camera captures the images of eye at regular intervals of time and sends them to eye tracking and gaze estimator where the coordinates of the gaze position are estimated.

3.2 Gaze Direction Estimation

The image acquired by the camera directed at the eyes of the user is transferred to the face detection and eye tracking estimators. In the eye tracking estimator, eye position and pupil edge are detected. The face detector output and eye tracker output are applied as input to find the eye gaze direction which locates the application (location) on the screen the user is looking at.

The algorithm suggested by Laurence Durnell [6] for estimation of pupil radius and location is shown in block diagram in Fig. 3.

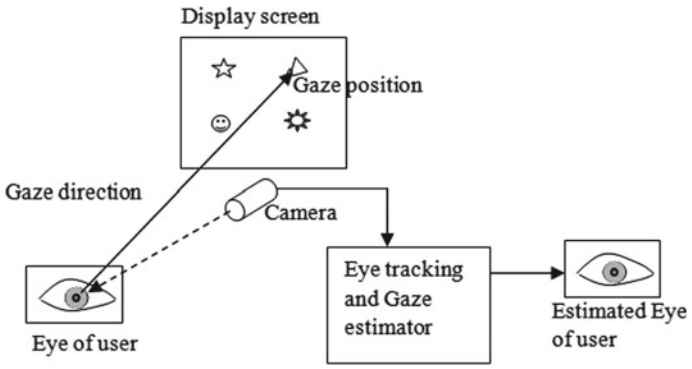


Fig. 2 Gaze position estimation

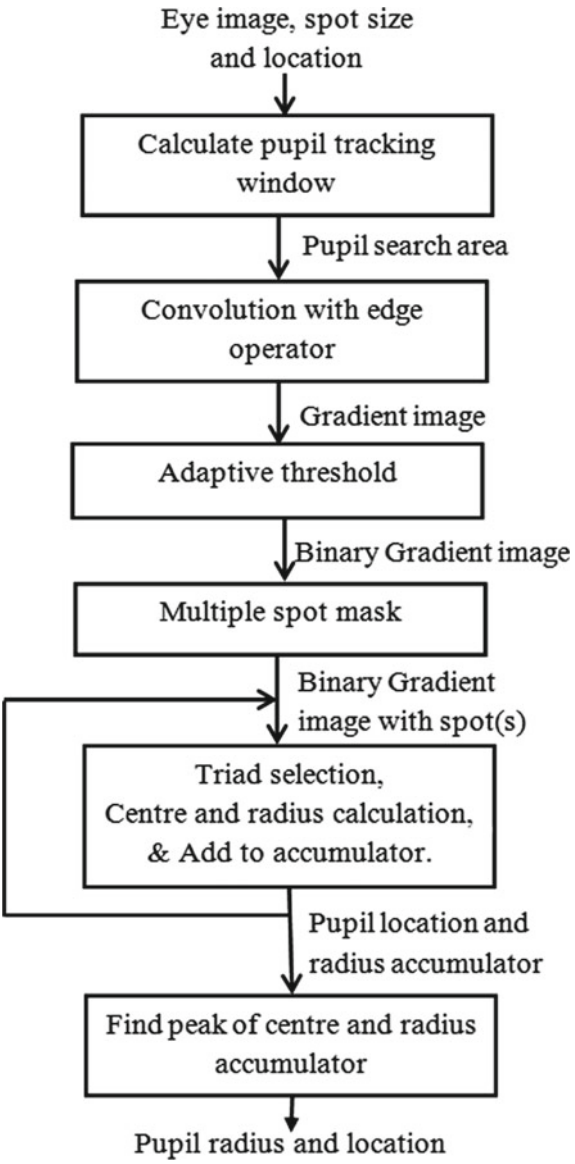


Fig. 3 Methodology for estimation of pupil radius and location

3.3 Iris Localization

In the preprocessing, face rotation is adjusted. From the image of the face, eyes are detected by using gradient directions and intensity information. Iris is detected by detecting circular disk in the eye. Correlation output is determined using weighting kernel W, intensity kernel I, and gradient kernel C_R as

$$Cor = \alpha(I \otimes C_R) + (1 - \alpha)W$$

where $0 \leq \alpha \leq 1$ is used for reduction of spurious detections [7]. The location of the image of the eye with maximum *Cor* indicates the iris center.

4 Hardware Implementation

A video camera is used to capture the eye movement of the user. Video camera is interfaced with Raspberry Pi module. Video camera transmits the captured video data as frames to the ARM processor in the Raspberry Pi module. Existing eye tracking and gaze estimation algorithms are loaded into the ARM processor, and the ARM processor processes the image data received and estimates the gaze direction. Block diagram of eye tracking algorithm is shown in Fig. 4.

The peripheral devices like display screen, loud speaker, wireless transceiver, and buzzer are interfaced to the ARM processor. The algorithm to display several pages of MENU and its OPTIONs to convey messages the user wishes to convey will be developed and implemented. Then the algorithm for producing audio to the selected messages will be developed and tested.

Then the wireless transceiver will be interfaced with the DSP and tested. The mechanical design involves display system with camera placed at the bottom of it

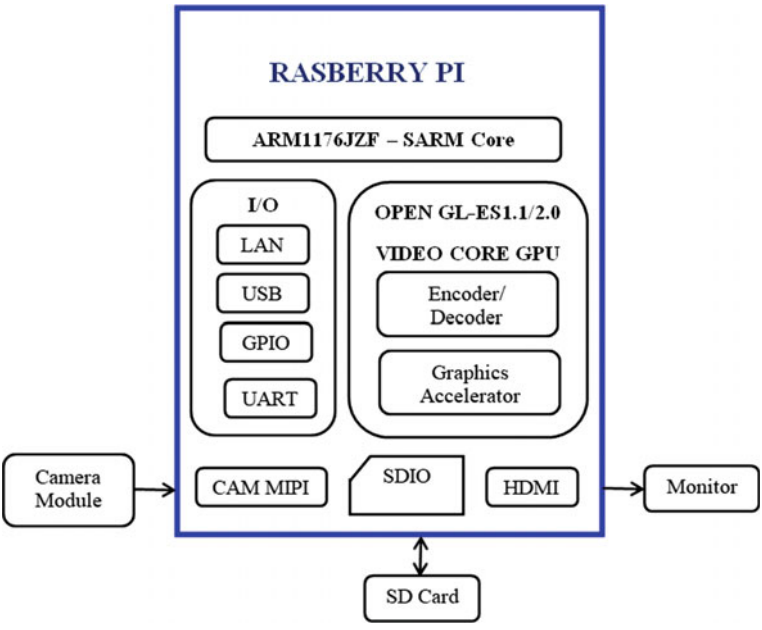


Fig. 4 Block diagram of eye gazing circuit implementation

and facing toward the user. Provisions will be given for buzzer, wireless transmitter, and a loud speaker. Remaining all electronics part should be placed in a compact box attached to the display screen. A handheld device for the caretaker will be designed. It consists of wireless receiver with buzzer and display of short message.

5 Results

In this following Fig. 5a–c, firstly the face is detected and then the eye ball is detected. From the detected eyeball, the iris gets highlighted. Based on the iris detection, the particular direction of eye movement is displayed.

If the user stares at the screen for three seconds, the MENU page (as shown in Fig. 6a) will be displayed. The MENU page consists of four sections—FOOD, MEDICINE, FEELINGS, and DAILY NEEDS. These sections can be changed according to the user requirement.

The user selects one of the items mentioned on the MENU page by staring at it. Proposed system's dwell time could be in milliseconds, but the time taken in selection of the item on the MENU page purely depends on the cognitive skills of

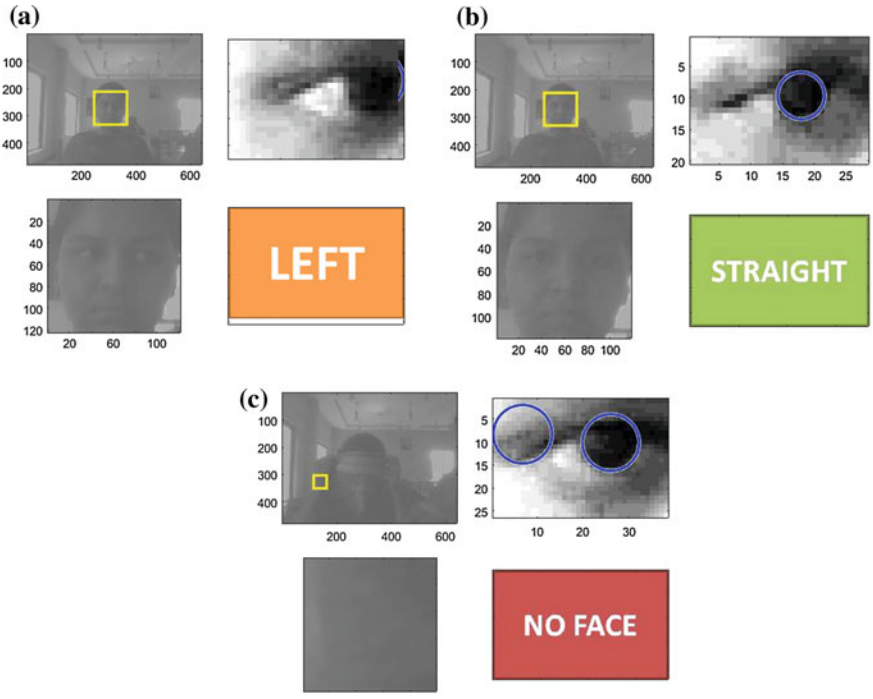


Fig. 5 Detecting eyeball in particular direction **a** left, **b** straight, and **c** no face detected

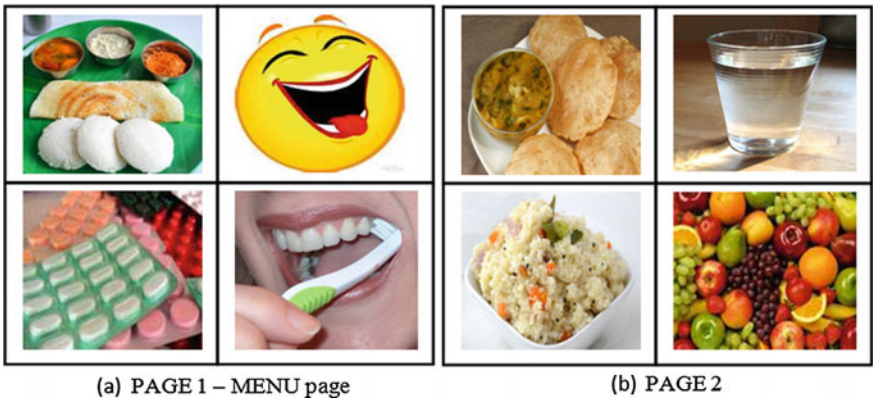


Fig. 6 PAGE 2 consists of items related to FOOD item in the MENU page (PAGE 1)

the user. Therefore, the Pause time will be adjusted for individual user. It may range from 1 to 3 s. After the pause time, the space around the selected item will be highlighted. If the highlighted item is the one the user looked at, then the user blinks to open the page related to that item. Here, the blink of user's eye acts as ENTER key on the computer keyboard. Let us assume that the user selected the item FOOD, the page related to FOOD opens, as shown in Fig. 6b. In this manner, any number of pages can be included as per user's recommendations.

6 Conclusions

Elderly people with disability in speech and body movements are unable to communicate properly with others. Proposed supportive communication system becomes a helping hand to elderly people. Proposed system is based on eye tracking. The accuracy of the iris detection depends on the distance of the user from the monitor. The estimated gaze data locates the position on the screen at which he is staring. The screen interfaced with ARM processor displays different pages designed to fulfill the messages related to the daily needs of the elderly person. With eye movements, staring and blinking, he can communicate comfortably with caretakers.

Declaration: Authors have taken the consent from the concerned authority to use the materials, etc., in the paper. Authors will be solely responsible if any issues arise in future with regard to this.

Acknowledgements Authors would like to acknowledge the encouragement given by management of Shri Vishnu Engineering College for Women.

References

1. Stöppler, M.C.: Difficulty with speech: symptoms & signs. News Lett. (2017). MedicineNet.com
2. Loncke, F.: Augmentative and Alternative Communication: Models and Applications. Plural Publishing Inc, San Diego (2014)
3. Morimoto, C., Amir, A., Flickner, M.: Detecting eye position and gaze from a single camera and 2 light sources. In: Proceedings of 16th International Conference on Pattern Recognition, pp. 314−317 (2002)
4. Babcock, J., Pelz, J.: Building a lightweight eye tracking headgear. In: ACM Eye Tracking Research and Applications Symposium, San Antonio, TX, USA, pp. 109–114 (2004)
5. Chuang, Z., Jian-Nan, C.H.I., Zhao-Hui, Z., Zhi-Liang, W.: A novel eye gaze tracking technique based on pupil center cornea reflection technique. Chin. J. Comput. **7**, 1272−1285 (2010)
6. Durnell, L.: Eye Tracking Systems, Patent No.: US 7,391,887 B2 (2008)
7. George, A., Routray, A.: Fast and accurate algorithm for eye localization for gaze tracking in low resolution image. IET Comput. Vis. **10**, 660−669 (2016)

Real-Time Digital Control of Synchronous Buck Converter for Low-Power Application

R. Rashmi and M. D. Uplane

Abstract This paper presents a real-time digital control of the DC–DC synchronous buck converter. A TMS320F288069 M DSP microcontroller is used for real-time regulation via serial communication and controlled through the MATLAB Simulink. This allows buck regulator power control system without the actual buck regulator main control loop which is normally achieved by the integrated chip (IC). The synchronous buck converter presents the advancement in converter part with synchronous MOSFETs, advancement in control part with programmable DSP controller and in communication part with low-cost GUI and easy implemented RS232. The GUI is designed for desired range of voltage and switching load. The converter's output voltage is dynamically adjustable from 0 to 5.5 V, while capable of supplying a maximum load current of 6 A from an input supply of 9–12 V.

1 Introduction

Power converter design and control is getting more and more challenging with the demand of ever increasing functionality such as variable power supply, real-time control, graphical user interface (GUI), faster transient response, efficiency, thermal performance and portability. Normally, power supply design focuses on one or two parameters either in converter topology [1] or device selection or control IC selection or control algorithm selection or communication protocol. In order to meet all-round requirement of power supply, this paper focuses on the advancement in converter part, control part as well as in communication part as shown in Table 1.

Converter part: With the new MOSFET technologies, Rds ON and total gate charge (QG) is getting lower and lower with the increased size of the die or by paralleling the discrete devices in power converter design while the minimum diode drop with its physical limitation is 0.3 v which becomes significant for the low voltage supply [2]. Hence to design low power converter, a nonsynchronous buck

R. Rashmi (✉) · M. D. Uplane
Department of Instrumentation Science, Savitribai Phule Pune University, Pune, India
e-mail: ruchirashmi@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_46

Table 1 Power supply characteristics and their dependency on the various sections of converter

Required characteristics	Improved section
Variable power supply	Control section
GUI	Communication and control section
Faster transient response	Converter (device selection)
Efficiency	All
Thermal performance	Converter
Small size	Converter

converter is replaced with synchronous buck converter. A synchronous buck converter uses two MOSFET switches in power converter part which increase the efficiency, converter size and thermal performance [3].

Control part: DSP controllers based high-power converters have shown their potential application in past few years [4] due to their optimized processor for fast operation, low cost, energy efficient. DSP controller combines the features of a microcontroller and the performance advantages of a DSP core and delivers the real-time million instructions per second (MIPS) and the tightly integrated peripherals to implement predictive control algorithms [5]. DSP controller has attracted significant attention due to its outstanding control of nonlinear controllers for nonlinear processes [6] with GUI. The fixed controlled integrated chip (IC) cannot provide GUI with variable voltage requirement so to achieve the flexibility in converter design programmable controlled integrated chip (IC) DSP controller is used and is controlled through the MATLAB Simulink.

Communication part: With revolution in the communication protocol number of communication protocol could be used for the communication with the PC. With easily implementation and low cost, RS232 communication protocol is used to communicate between controller and MATLAB [7]. MATLAB-based control and GUI provide full control of functions and setting. Serial communication, voltage and load are settable from the main screen. The larger screen available on PCs provides multiple readouts, and enhanced data displays compared to front panel operation.

This paper is organized as follows. In Sect. 2, operation principal of real-time synchronous buck regulator is described. Design of different sections of synchronous buck regulator converter with current feedback loop is mentioned in Sect. 3. Control of real-time buck regulator with GUI is described in Sect. 4. The result is discussed in Sect. 5 and the conclusion is given in Sect. 6.

2 Operation of Real-Time Synchronous Buck Regulator

In a nonsynchronous buck converter, one MOSFET and one diode is used as shown in Fig. 1 while synchronous buck converter consists of two MOSFET switches, where second mosfet acts as a fast synchronous diode are shown in Fig. 2. An inductor is connected to both mosfet switches, and a capacitor is connected across the output. As the inductor is connected to the both MOSFET in synchronous buck converter switches, there is no discontinuity in the current [8]. The operation of synchronous buck during the DTs when Q1 is on, and Q2 is off when the inductor charges the output capacitor and delivers the power to load. During the (1-D)Ts period, Q1 is off and Q2 is on, and output capacitor supplies the load current [9].

2.1 Operation During DTs Period

In a synchronous buck converter, the MOFET Q1 turns on and MOSFET Q2 turns off during the period DTs as shown in Fig. 3. The inductor current flows from the source to the load. The voltage VL across the inductor during this time is Vi-Vo. The inductor current iL rises linearly with the slope of (Vi-Vo)/L, and the voltage across Q2 is Vi. The energy is delivered to the output inductor, and the load is same as the nonsynchronous buck.

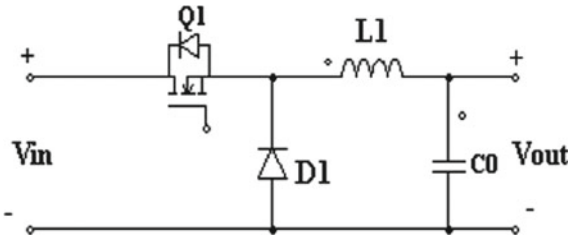


Fig. 1 Nonsynchronous buck converter

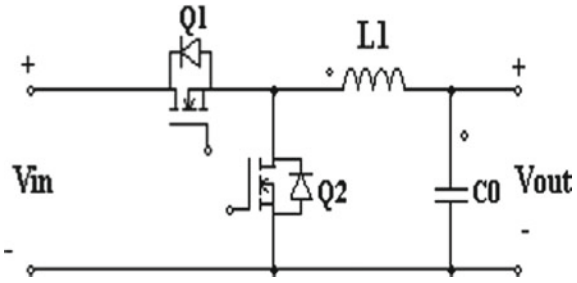


Fig. 2 Synchronous buck converter

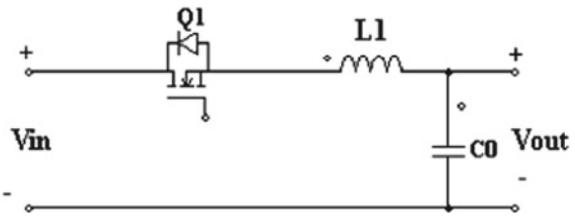


Fig. 3 Operation during the DTs period

2.2 Operation During (1-D)Ts Period

During this time, Q1 is OFF and Q2 is ON. Figure 4 shows the operative circuit during the period (1-D)Ts of the switching period The inductor current cannot change instantaneously. The inductor current starts to decrease. The negative di/dt of the inductor current develops sufficient voltage (Ldi/dt) with a polarity such as to drive the inductor current through the switch S2 for the freewheeling action on the inductor current. MOSFET driver drives two MOSFETs in a complimentary manner in synchronous buck converter with a small dead time between their conduction intervals to avoid occurrence of shoot–through. The synchronous buck converter always operates in continuous conduction because during the dead time inductor current flows through the body diode of lower MOSFET.

3 Design of Real-Time Synchronous Buck Regulator

3.1 Converter Part

LM5109B high voltage gate driver designed to drive both the high-side and the low-side N-channel CSD87588 N MOSFETs in a synchronous buck converter. A shunt resister and op-amp is used for the buffering and real-time monitoring of the inductor current. The buck converter is providing the output from 0 to 5.5 V. One load of 7.5 Ω is permanently connected to the circuit, and the other higher load of 2 Ω can be switched ON or OFF using the MOSFET switch to test the transient response and loop tuning of the circuit.

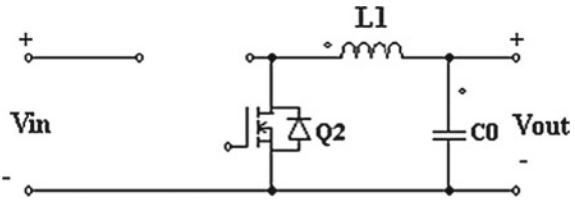


Fig. 4 Operation during the DTs period

3.2 Control Part

The control part is designed around TMS320F288069 M DSP microcontroller with time-critical and non-time-critical task. ADC measurements, loop control and system protection are controlled through the interrupt and taken as time critical task [10]. UART communication is executed in background loop and taken as non-time-critical task.

3.3 Communication Part

To enable programming, debugging and UART application via USB with the electrical isolation to PC is provided by XDS100v2. An MCU receives the signal of inductor current and load voltage from the converter and desired voltage and load on/off condition from the PC, and MCU sends commands to the converter to perform tasks such as output voltage adjustment and load on/off control.

4 Control of Real-Time Buck Regulator

The current through the inductor L1 is actively monitored and buffered using a shunt and op-amp, and the load voltage is sense through the precise resister divider. These analog signals are conditioned and converted to the digital signals using channel 2 and 3 of 12 bit single-ended input ADC and fed back to the *DSP* controller unit with a delay of 12000 μs. ADC uses 14 sample window length which is to properly transfer the charge into its sampling capacitor. MCU also receives the desired load voltage set by the user in GUI from the PC through UART. Then, MCU generates the control signal to MOSFET controlled load with 50% duty cycle. The MCU does a quick and efficient turn ON/OFF of the higher and lower power MOSFETs of buck converter using gate drive by buffering PWM signal of 200 kHz. The block diagram of complete power converter system comprised of buck converter, microcontroller unit (MCU) and the computer with respective signal is shown in Fig. 5 where MCU receives feedback current (Ifb) and voltage (Vfb) from the buck converter and set point of output voltage and coefficients of the PID control algorithm from the user. After execution of the control algorithm, it generates the PWM signal to drive two MOSFETS to regulate the output voltage at the set point. MCU generates 50% duty cycle PWM to drive MOSFET and control the switching of higher load according to the instructions from the user.

The firmware modules mainly used by the microcontroller unit (MCU) are enhanced pulse width modulator (ePWM), analog to digital converter and comparator (ADC), system control and interrupts and universal asynchronous receive

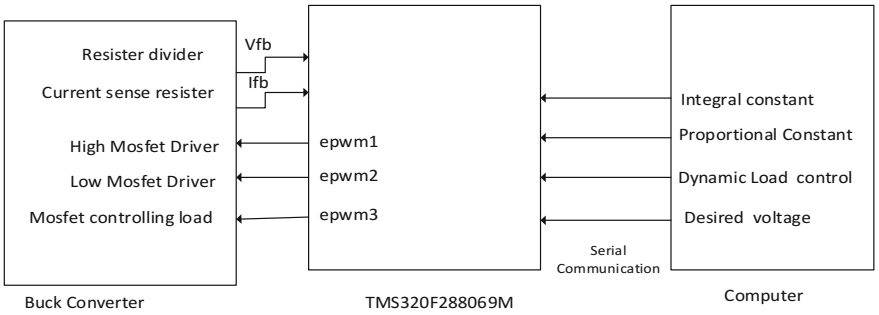
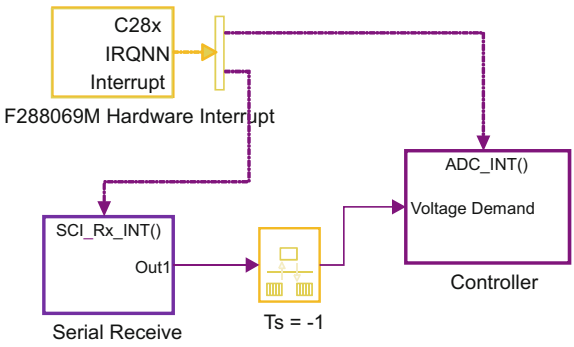


Fig. 5 Block diagram of converter

and transmit (UART). Some firmware modules are shown in the Fig. 6 for software control of the buck converter. In interrupt block, two interrupts ADC interrupt ADCINT1 and serial receive interrupt SCIA_RX is used. The SCI receive block supports asynchronous serial digital communication between the target and other asynchronous peripherals. In SCIA_RX block, *TMS320F288069 M*-supported SCI module A is used for the serial communication using vector data. Additional package header and terminator are used at the front and end of the received data package. Data are 16-bit unsigned integer, with the sample time of 0.06 s. ADC_INT () block denotes an ADC of 12-bit single-ended input with a length of 14 acquisition window. Serial communication blocks as shown in Fig. 7 are chosen with the characteristics such a serial communication port 9, Baud rate 12000000, data bit 8, no parity, 1 stop bit and little endian byte order is selected for the serial configuration blocks and serial receive block.

The user can switch the load and can set the desired set voltage in GUI of the MATLAB as shown in Fig. 8. User can vary the parameters using GUI knob and toggle switch. The parameter is communicated to the MCU using the serial communication.

Fig. 6 Firmware modules of MCU



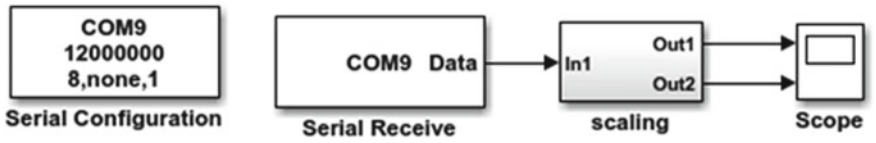


Fig. 7 Signal logging

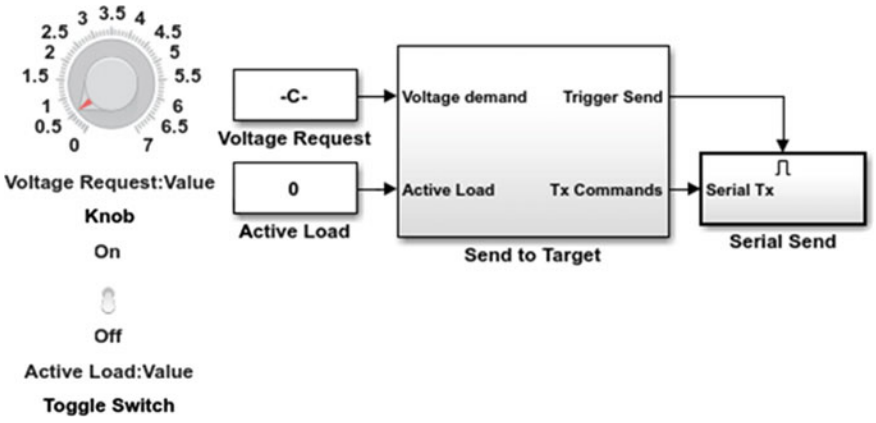


Fig. 8 Parameter tuning

5 Result and Discussion

Steady and Transient State Performance of the Synchronous Buck
Both input currents with or without switching load increase with the increase of output voltage as shown in Fig. 9. But load current with switching load increases sharply as compare to load current without switching load as higher current is required to derive the switching load. Buck converter provides a maximum load current of 5.5 A with 5.5 V as shown in Fig. 10. The load current decreases slightly in both cases low and high load with the increase in the output voltage due to the heating effect of resister at higher voltage and current.

The output voltage is regulated at 2 V by channel 2 as shown in Fig. 11. The load current without switching load is 0.18 mA and with switching load is 6 A. As shown in Fig. 12, the output voltage is regulated at 5.5 V by channel 2 and the load current without switching load is 0.19 mA and with switching load is 5.6 A. The high current spikes due to ringing is shown in load current at switching load in both figures ringing on the switch node is a function of speed of high-side MOSFET's switching, and the stray inductance and capacitance speed depends on the fast-switching MOSFETS.

In steady state of the converter, when low-side MOSFET is on, the power to the load is provided only from the output inductance and capacitance. At this point, the

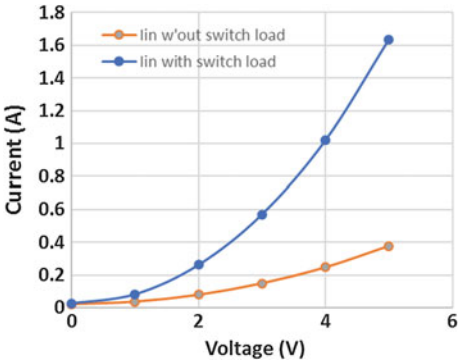


Fig. 9 Variation of input current with load

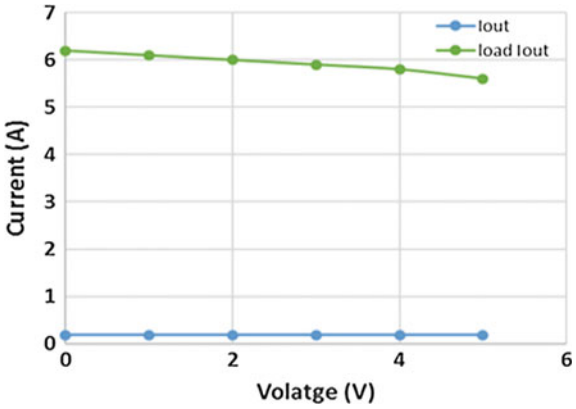


Fig. 10 Variation of load current with load

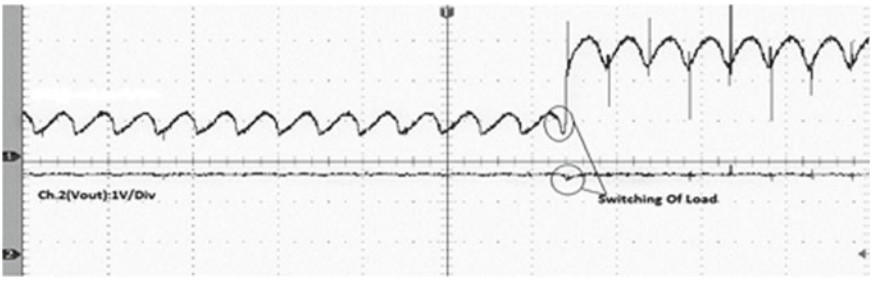


Fig. 11 Load current and voltage response with switching load at 2 V

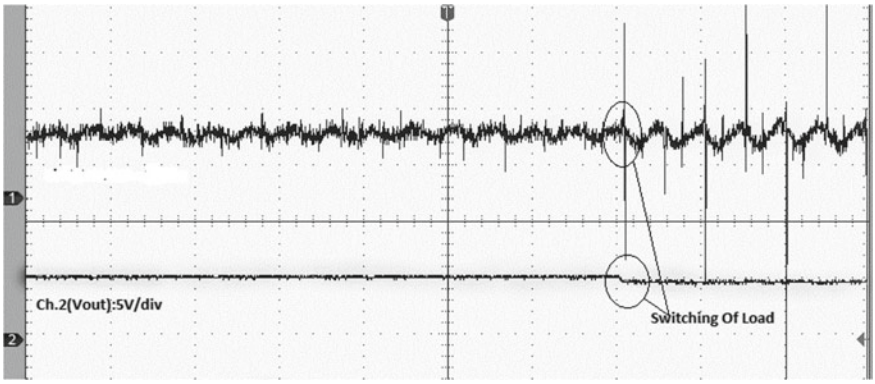


Fig. 12 Load current and voltage response with switching load at 5.5 V

energy is stored in the parasitic inductance of the low side of MOSFET. At the end of switching cycle, the converter prepares to switch the low-side MOSFET off and high-side MOSFET back on to replenish power to the output L. Strong gate drives and a fast-switching MOFETs allow the low-side MOSFET to be turn off quickly. In high load condition, the inductor current keeps flowing to the output and the energy remains in the parasitic drain and source inductance of the low-side MOSFET. After a fixed dead time, the high-side MOSFET turns on, and the energy from the low-side and high-side parasitic inductances appears as an LC ringing wave form on the switch node.

6 Conclusion

This paper shows the real-time synchronous programmable buck converter controlled through the DSP and MATLAB Simulink. Programmable converters parameters can be easily modified by the user using GUI with specified applications. The limitation of this paper is that there is still presence of switch node ringing at higher load which can be removed by using different techniques such as boot resister, high-side gate resister and snubber.

Acknowledgements This work has been supported by DST-PURSE and Department of Technology, Savitribai Phule Pune University. R. Rashmi is thankful to University Grant commission for SRF.

References

1. Hulugappa; V., Kusagur, A.: Comparative analysis of DC−DC converter topologies with soft-Switching techniques In: International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), pp. 106–110 (2016)
2. Sahu; B., Rincon-Mora, G.A.: A low voltage, dynamic, noninverting, synchronous buck-boost converter for portable applications. IEEE Trans. Power Electron. **19**(2), 443–452 (2004)
3. Chen; G., Deng; Y., Dong; J., Hu; Y., Jiang; L., He, X.: Integrated multiple-output synchronous buck converter for electric vehicle power supply. IEEE Trans. Veh. Technol. **66** (7), pp. 5752–5761 (2017)
4. Bist, V., Singh, B.: A unity power factor bridgeless isolated Cuk converter-fed brushless DC motor drive. IEEE Trans. Ind. Electron. **62**(7), pp. 4118–4129 (2015)
5. Srdic; S., Nedeljkovic, M.: Predictive fast DSP-based current controller for thyristor converters. IEEE Trans. Ind. Electron. **58**(8), pp. 3349–3358 (2011)
6. Kalantzopoulos; A., Zigouris, E.: Real-time DSP applications remotely controlled through customized GUIs. In: 2012 5th European DSP Education and Research Conference (EDERC), pp. 71–75 (2012)
7. Song; F., Wang; Y., Cao, H.: Software Design of Brushless DC Motor Control System Based on DSP In: 2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics **2,** 364–367 (2013)
8. Rejwan Uddin; M., Tasneem; Z., Annie; I.S., Salim, K.M.: A high capacity synchronous buck converter for highly efficient and lightweight charger of electric easy bikes. 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 392–395 (2017)
9. Taylor, R., Manack, R.: Controlling switch-node ringingin synchronous buck converters Analog Applications Journal, Texas Instruments Incorporated, 2Q (2012)
10. Kong; L., Zhu, L., Zhang; L., Bao; H.., Rao, C.: Real-time controller based on FPGA and DSP for solar ground layer adaptive optics prototype system at 1-m NVST. IEEE Photon. J. **9**(2) (2017)

Integrated Parallel K-Nearest Neighbor Algorithm

Rashmi Agrawal

Abstract The task of classification in applications of data mining is also known as supervised learning where some specific classes are predefined for the training sample and objects are assigned with appropriate class. This is modeled with any classifier and tested on testing data to find the appropriate class. A simple and efficient algorithm for supervised classification is k-Nearest Neighbor (KNN) in which determining the optimal value of k has been an interesting research problem. In this paper, we develop a new algorithm called "Integrated Parallel k-Nearest Neighbor" using the ensembling technique and obtained better results than any other single classifier based on a suitable distance measure and based on any value of k ranging from 1 to 20.

1 Introduction

K-Nearest Neighbor (KNN) is one of the simple methods in classification for prediction of data using its attributes. This method gives better results over other methods in generalizing even from a reasonably undersized training dataset [1]. KNN method uses a suitable distance measure to classify a new instance. For this purpose, the distance of k-nearest neighbors is computed and the class label of the nearest neighbor is predicted as the class label of new instance. The accuracy of KNN is highly affected by choosing the number of k-nearest neighbors. If k is too small, it is sensitive to noise points and if it is too large, it can lead to a bias of models. Some methods, like cross-validation, are usually used to triumph over such problem, but they are less efficient.

In machine learning, an individual model may behave as an expert outlook for a meticulous dataset and may afford the best results, while it may not turn out

R. Agrawal (✉)
Manav Rachna International University, Faridabad, India
e-mail: drrashmiagrawal78@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_47

apposite results for some other dataset of particular problem. To obtain unswerving results in such cases, we can combine different models. Ensemble methods are learning methods that increase the performance of a classifier by integrating opinions of multiple learners. These methods have the ability of efficient and accurate prediction without demanding the value of k.

This paper is an extended version of [2] in which one algorithm Relevant Feature Selection k-Nearest Neighbor (RFS-KNN) was developed to improve the accuracy of KNN by selecting the relevant feature. The main objective of this paper is to construct an Integrated Parallel k-Nearest Neighbor (IP-KNN) algorithm, using ensembling technique, which increases the accuracy of predictive model greatly.

2 Related Work

In 1999, Baver and Kohavi [3] mentioned boosting as an interesting research line and also noticed that merely standard interpretation of counting of a highly weighted instance would not work. It was shown by Brieman [4] in 1996 that KNN classifiers are not effective in bagging methods due to stability of KNN with respect to variations in the training samples. They also noticed that there is a large overlapping in differently bagged training samples, and hence, the errors of different classifiers are highly correlated causing less efficiency in prediction by voting them.

Grabowski [5] made an ensemble in 2002 using cross-validation to get slightly better results than single classifier. Armores and Sebe [6] proposed a functional approximation to estimate the similarity function be generalization method instead of estimating a parametric or similarity distance. Sufficient information of the literature review is already given by Agrawal [2] For this, they suggested AdaBoost with decision stumps. The proposed distance estimation method was for all kind of classifiers and was not based on parametric approach. The method is effective when training set is small and the estimated similarity uses a small number of dimensions resulting in dimensionality reduction. Agrawal [2, 7–9] has done a lot of work in the k-Nearest Neighbor classification. Athitsos [10] proposed a method to combine boosting with k-Nearest Neighbor classifier. Very few previous works have embarked on the step of using boosting in KNN classifiers, and KNN has been used as part of hybrid ensemble [11].

3 Integrated Parallel k-Nearest Neighbor (IP-KNN) Algorithm

Agrawal [7] showed that the accuracy of a classifier depends on the value of k as well as on the chosen distance measure. She analyzed the integrated effect of various distance measures on different values of k in k-Nearest Neighbor algorithm

on different datasets taken from UCI machine learning repository. Under the assumption that all classifiers do not make the same mistake, we extend this work by developing an Integrated Parallel k-Nearest Neighbor algorithm in which we ensemble various k values in one classifier and various distance measure in another classifier and integrate the results with our previous work Relevant Feature Selection k-Nearest Neighbor Algorithm (RFS-KNN) [2]. We are also motivated by the fact that it is unlikely that all classifiers make the same mistake.

Architectural Framework

According to Rockach [12], in a classifier ensemble framework, we have a set of classifiers CF = {CF1, CF2….CFn} where each classifier performs a mapping of instances with set of class labels. In the design of a classifier using ensembling, the classifier must perform two responsibilities, establishment of individual classifier and developing a rule for combining the classifiers, which predicts the final class label.

In our proposed framework of IP-KNN, individual classifiers are constructed separately and also there is provision for combining the classifiers to predict the final class label. The framework of IP-KNN is shown in Fig. 1. The dataset is passed to the ensemble classifier which has parallel KNN with different values of k. The predicted class labels with these classifiers of different k values are compared, and if we get same result by all classifiers, the instance is stored in a dataset named as accurate dataset otherwise all other instances are stored in a separate dataset namely partially accurate dataset. This partially accurate dataset is passed to the set of parallel KNN classifiers with various distance measures. Further, on the basis of assumption that it is unlikely to repeat the same mistake by all classifiers, we compare the predicted class labels of all classifiers with different class labels and if we find them same, we update the accurate dataset by moving particular instance to accurate dataset from partially accurate dataset. Lastly, we pass the partially

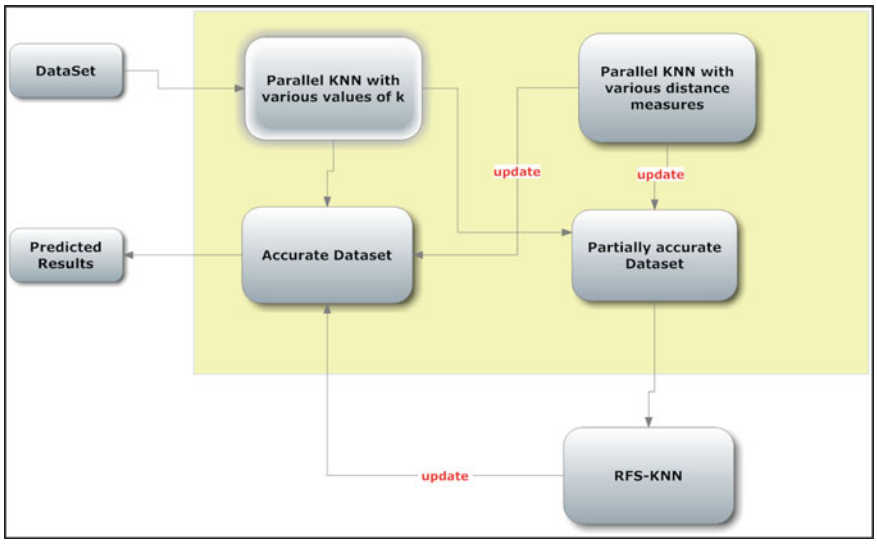


Fig. 1 Framework of Integrated Parallel k-Nearest Neighbor (IP-KNN) algorithm

accurate dataset to our previously developed classifier named Relevant Feature Selection k-Nearest Neighbor (RFS-KNN) which selects the relevant features of a dataset and predicts the class label.

Now the partially accurate dataset is merged with accurate dataset to find the finally predicted outcome, and we calculate the accuracy of the classifier. The algorithm for the same is given below.

Algorithm of IP-KNN

```
Accurate:= Empty,    Partially_Accurate:=Empty
## We declare two datasets Accurate and Partially_Accurate to store the predicted result
kset= [ 1 2 3 4 5 10 15 20]
For i=1 to n
Start loop  Set flag:=1
        Set base_value:= get_class_label(i,1)
## Where get_class_label(i,j) is a user defined function to get the class label of the i record with j number of nearest neighbors.

For each k in kset
Start loop if get_class_label(i,k) != base_value
        Then set flag:=0
        end if
        end loop
       if flag=1 then
       add record i to Accurate dataset
    else
       add record i to Partially_Accurate dataset
end if
end loop
dset= [ d1 d2 d3 d4]
## Where dset is an array which contains the various distance measures.
For i=1 to n
Start loop
Set flag:=1 and Set base_value:= get_class_label(i,1)
For each d in dset
Start loop
        if get_class_label(i,d) != base_value
        then    set flag:=0
        end if
end loop
if flag=1
        then    add record i to Accurate dataset
else    add record i to Partially_Accurate dataset
end loop
```

Algorithm for RFS-KNN

Algorithm for RFS-KNN is already given by Agrawal [2].

4 Results and Discussion

We applied the algorithm Integrated Parallel KNN on glass dataset. The first module of the algorithm "Parallel KNN with various values of K" gives the following table.

From Table 1, we compared the predicted class labels with various values of k. If they all are same, we put that record into the accurate dataset and all other records into the partially accurate dataset. The accurate dataset and partially accurate datasets are shown below (Table 4).

Then partially accurate dataset is passed to RFS-KNN, and we achieved 85.89% accuracy as compared to the individual classifiers using various K and distance measures. We applied IP-KNN on two real datasets Glass and Ecoli taken from UCI machine learning repository and achieved more accurate predicted results. The results are summarized in Table 5 (Figs. 2, 3, 4, 5) (Table 6).

Table 1 Sample records of glass dataset

Record No.	k = 1	k = 2	k = 3	k = 4	k = 5	k = 10	k = 15	k = 20
1	1	1	1	1	1	1	1	1
2	1	1	**2**	**2**	**2**	**2**	1	1
3	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1
6	**2**	**2**	**2**	**2**	**2**	1	1	1
7	1	1	1	1	1	1	**2**	1
8	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1
10	**3**	1	1	**2**	1	**2**	1	1

Table 2 Accurate dataset after applying Parallel KNN with various values of (sample)

Record No.	k = 1	k = 2	k = 3	k = 4	k = 5	k = 10	k = 15	k = 20
1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1

Table 3 Partially accurate dataset after applying Parallel KNN with various values of k (sample). After applying parallel KNN with various distance measures as d1, d2, d3 and d4 (Euclidean, Cityblock, Cosine and Mahalanobis), we update the accurate dataset by moving following records from Table 3 to Table 2

Record No.	k = 1	k = 2	k = 3	k = 4	k = 5	k = 10	k = 15	k = 20
2	1	1	**2**	**2**	**2**	**2**	1	1
6	**2**	**2**	**2**	**2**	**2**	1	1	1
7	1	1	1	1	1	1	**2**	1
10	**3**	1	1	**2**	1	**2**	1	1
11	**3**	1	**3**	**3**	1	1	1	1
17	1	1	1	1	1	**2**	**2**	1
18	1	1	1	1	**2**	1	**2**	1

Table 4 Additional accurate records after applying Parallel KNN with various distance measures (sample)

Record No.	d1	d2	d3	d4
11	1	1	1	1
17	1	1	1	1
19	1	1	1	1
28	2	2	2	2
37	**1**	**1**	**1**	**1**
39	2	2	2	2
40	2	2	2	2

Table 5 Accuracy(in %) of Glass and Ecoli datasets on various k and IP-KNN

Dataset	k = 1	k = 2	k = 3	k = 4	k = 5	k = 10	k = 15	k = 20	IP-KNN
Glass	70.51	69.23	67.94	71.42	65.38	58.97	53.84	56.41	85.89
Ecoli	76.85	74.38	79.33	78.51	79.33	78.51	80.99	80.16	82.64

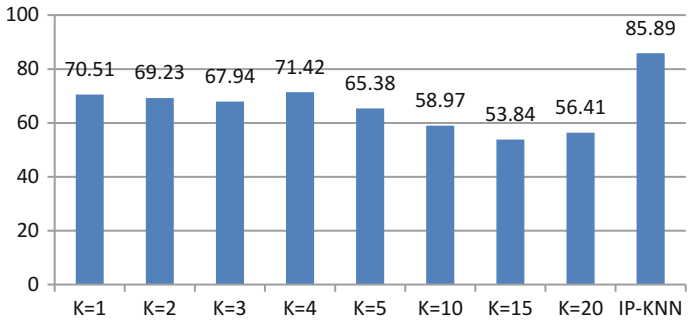


Fig. 2 Comparison of accuracy of IP-KNN with accuracy of various values of k in KNN in Glass dataset

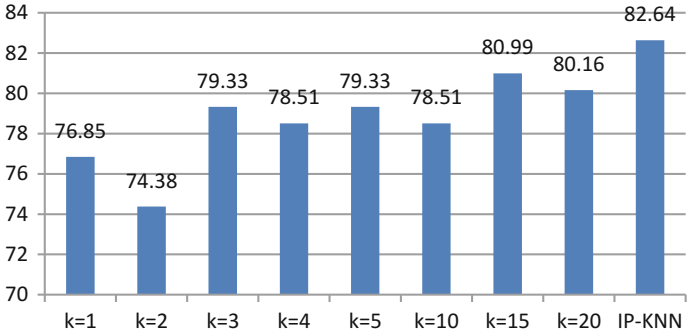


Fig. 3 Comparison of accuracy of IP-KNN with accuracy of various values of k in KNN in Glass dataset

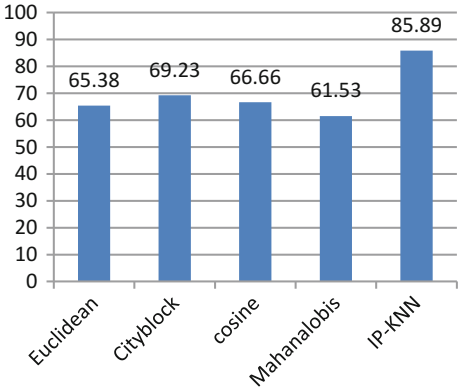


Fig. 4 Comparison of accuracy of IP-KNN with accuracy of various distance measures in KNN in Glass dataset

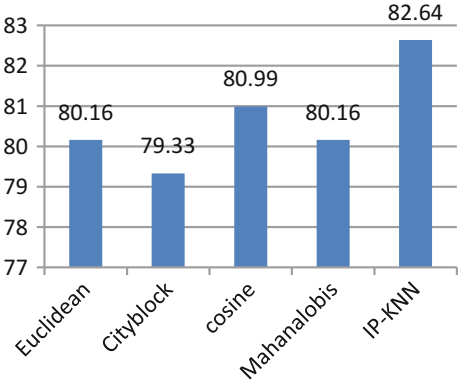


Fig. 5 Comparison of accuracy of IP-KNN with accuracy of various distance measures in KNN in Ecoli dataset

Table 6 Accuracy of Glass and Ecoli datasets won various distance measures and IP-KNN

Dataset	Euclidean	Cityblock	Cosine	Mahalanobis	IP-KNN
Glass	65.38	69.23	66.66	61.53	85.89
Ecoli	80.16	79.33	80.99	80.16	82.64

5 Conclusion

We developed a new algorithm "Integrated Parallel k-Nearest Neighbor" using ensembling technique and applied on two datasets Glass and Ecoli. The results discussed in the above section show that IP-KNN produces better results as compared to any other single classifier neither for a certain distance measure nor for a certain value of k.

Acknowledgements My thanks are due to my supervisor Dr. Babu Ram, Professor, Manav Rachna International University, Faridabad for his constant guidance and support.

References

1. Rokach, L.: Ensemble-based classifiers. Artif. Intell. Rev. **33**(1–2), 1–39 (2010)
2. Agrawal, R.: A modified K-nearest neighbor algorithm using feature optimization. Int. J. Eng. Technol. **8**(1), 28–37 (2016)
3. Bauer, E., Kohavi, R.: An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Mach. Learn. **36**(1–2), 105–139 (1999)
4. Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)
5. Grabowski, S.: Voting over multiple k-NN classifiers. In: 2002 Proceedings of the International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science. IEEE (2002)
6. Amores, J., Sebe, N., Radeva, P.: Boosting the distance estimation: application to the k-nearest neighbor classifier. Pattern Recognit. Lett. **27**(3), 201–209 (2006)
7. Agrawal, R.: Integrated effect of nearest neighbors and distance measures in k-NN algorithm. In: CSI - 2015, 50th Golden Jubilee Annual Convention (2015)
8. Agrawal, R.: Design and development of data classification methodology for uncertain data. Indian J. Sci. Technol. **9**(3) (2016)
9. Agrawal, R.: K-nearest neighbor for uncertain data. Int. J. Comput. Appl. **105**(11) (2014)
10. Athitsos, V., Sclaroff, S.: Boosting nearest neighbor classifiers for multiclass recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 2005 CVPR Workshops. IEEE (2005)
11. Woods, K., Kegelmeyer, W., Bowyer, K.: Combination of multiple classifiers using local accuracy estimates. IEEE Trans. Pattern Anal. Mach. Intell. **19**, 405–410 (1997)
12. Rokach, L.: Pattern Classification Using Ensemble Methods, vol. 75. World Scientific, Singapore (2009)
13. Büchlmann, P., Yu, B.: Analyzing bagging. Ann. Stat. **30**, 927–961 (2002)
14. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: ICML, vol. 96 (1996)

Predicting the Performance of Disability Students Using Assistive Tools with the Role of ICT in Mining Approach

P. Saraswathi and N. Nagadeepa

Abstract With the advancement of technology, ICT and AT interact their function in providing high-quality education for students with special needs. SRM is a vital data mining tasks which strongly used in predicting the behavior of student and determine patterns general to a number of learner solutions. The existing algorithm derived pattern growth concept to introduce the rules. The aim of this paper is to provide great use of e-accessibility and latest technologies for the e-learning process. This paper introduced the SNEMiner Ranking algorithm to improve the achievement and success more effectively for disability students to overcome the problems in existing work. This research work used measures to find valid rules. The study of proposed work shows that SNEMiner Ranking algorithm with interestingness measures provides higher performance evaluation in analyzing the various factors to the special student community.

1 Introduction

The popular approach to analyze student's performance for the future society is data mining. It has been applied in education field to predict student performance. SRM is a significant mining approach to address the problem of prediction. It is applied in e-learning (educational data) to comprehend and foresee the learner's performance. It is said to be prediction rule. If some events A occur then some other events B are liable to track with a specified confidence. In data mining, rules are used to make valid suggestions to analyze student's performance. In the last two decades, ICT in

P. Saraswathi (✉)
Research Scholar (Part-Time), Bharathiar University, Coimbatore, Tamil Nadu, India
e-mail: saisaraswathi.research@gmail.com

N. Nagadeepa
Karur Velalar College of Arts and Science for Woman, Kuppam PO, Karur 639111, India
e-mail: nagadeepa1012@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_48

teaching and learning have been rapidly increased. It gives the optimistic impact of teaching and learning. The classic way of teaching is not helpful for disability students to improve their skills. The innovative use of ICT in their education creates a space for students in a preeminent mode. Students with Cognitive Disability use ICT to cope up in learning and communication.

Even though ICT is the technology to improve the disability student's performance, assistive technology is the modern educational process to build their education more inclusive. It is used to overcome the impairments of the challenging students to improve their performance.

2 Literature Review

Susana Capitaoa et al. (2012) [1] promoted the access of ICT for the students with deafness to decrease the communication barriers that are Internet-based solution. Lora Aroyo et al. (2004) [2] suggested the semantic educational web and proposed a framework to share the knowledge of education systems that reflect the educational systems and complex task in the systems. Ryan S.J.D. Baker et al. (2010) [3] proposed work make a rules with the use of measures. Svetlana Obradovi et al. (2015) [4] examined the probable for present support and innovative to dyslexia students uses a sensory move toward the conquest over reading difficulty. Educators are prepared to lever the online learning resources. Shane Dawson et al. (2009) [5] contributed to the exploration of student online learning behavior in the academic field.

Philippe Fournier–Viger et al. (2015) [6] proposed partially ordered sequential rule and showed that it increases the prediction accuracy more than the normal sequential rule. Rule Growth algorithm relies on pattern growth approach [7] and makes the rules among any items which grows recursively by scanning the list enlarge left and right parts. Rule Growth takes the parameters sequence database, thresholds and generates all rules r1 of size 1*1 such that support (r1) > = minimum support and call the left- and right-recursive procedures for growing each rules. It repeatedly performs costly database projection operations which decreases the performance of datasets [7].

Enrique Garcial et al. (2007) [8] focused the difficulties in learning systems. Usef Faghihi et al. (2011) [9] addressed the problem of tutoring agent and investigated human learning capabilities. Thi-Thiet Pham et al. (2013) [10] applied several measures to generate all sequential rules from a sequence database. Cristobal Romero et al. (2013) [11] explained the EDM techniques to analyze data in educational practice. Musavir Ahamed et al. (2014) [12] focused the existing ICT infrastructure in colleges and emphasized the utilization of ELT.

Michele D' Andreagiovanni et al. [13] presented the pattern mining study attack sequences. V. Simic et al. (2011) [14] intended a descriptive evaluation of e-learning. Jiri Klema et al. (2008) [15] compared special approach of sequential

mining based on inductive logic programming. Johann M. Marquez Barja et al. (2014) [16] proposed FORGE to access latest ICT. Dr. Sotirios D. Nikolopoulos (2016) [17] implemented cloud-based accounting information systems using association rule mining. Agathe Merceron et al. (2008) [18] focused cosine and lift is well suitable measure for educational data. Luan Bekteshi (2015) [19] gives the suggestions that information system is efficient to improve disability person academic. Sylvia soderstrom et al. (2010) [20] examined the values of ICT–assistive technologies. Javier Bravo et al. (2015) [21] studied to improve learning process using ICT. M A Jabbar (2017) [22] presents the increase in amenities and improves the performance in medical sector. Patsakorn singto et al. (2014) [23] showed the improvement in career analysis using ICT.

3 Methodology

3.1 Problem Definition

The existing algorithm acquired the rules relies on pattern growth approach restricted only for any selected sequence database and is not flexible to changes. Our proposed work addressed SNEMiner Ranking algorithm to avoid rescanning the complete dataset every time. Figure 1 shows framework for the prediction of disability student's performance using the elastic rule.

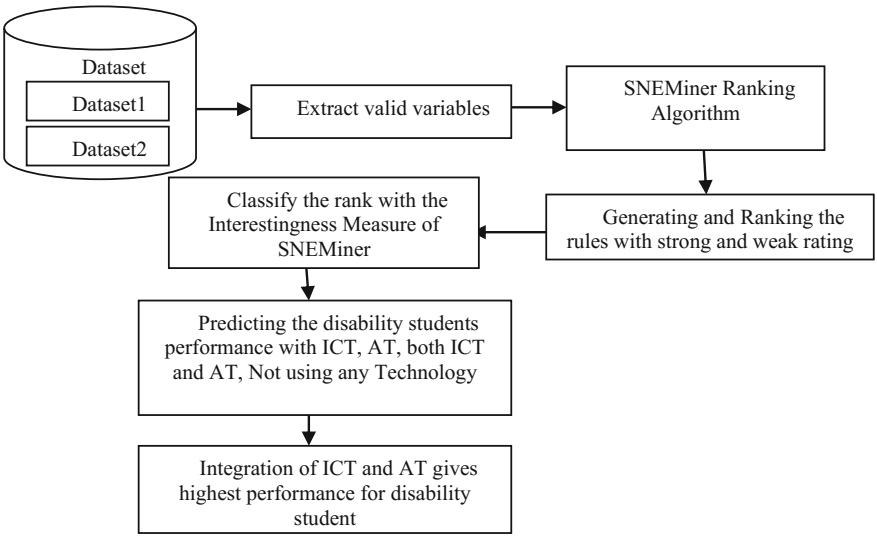


Fig. 1 Framework of predicting the performance of disability students using elastic rule

3.2 SNEMiner Ranking Algorithm

SNEMiner Ranking algorithm helps in obtaining the elastic rules where both the datasets, dataset1 and dataset2, are compared and rule rankings are restructured. The disability students detail without using any technology for their performance are collected and represented as dataset1, and collection of disability student details with ICT and AT are represented as dataset2. From Table 1, both the datasets are used to generate the rules which improve disability student's performance.

The ranking increased the value by 1 if it matches with dataset1 and the rule is said to be best. The left side of dataset1 is equal to any rule in dataset2 then the ranking of the rule in dataset1 decreased its value by 1 and the rule in dataset2 is increased its value by 2. Those rules in dataset1 are not occurred in dataset2 then it decreased its value by 2 and the newly occurring rules in the dataset2 increased its value by 1. By means of this technique, the rules are classified under best to worst category, i.e., from 1 to 5. This method is recursively followed to create the elastic rules. If rating is decreased to zero, it is eliminated and discarded. This algorithm is used to produce whether the rule is best or worst. Those rules with rating zero or less than zero are discarded.

Table 1 shows the strong rules for both the dataset1 without using any technology and dataset2 using the technology ICT and AT.

Algorithm:

Input: Dataset1 and Dataset2
Output: Elastic rules with ranking
1 Set r1 ranking as 1 in dataset1
2 Set r2 ranking as 5 in dataset2
3 Compare the dataset1 and dataset2
4 If (r1 ==r2) then increment r1and r2 by 1
5 If left (r1) is equal to left (r2) decrement r1 by 1 and increment r2 by 2
6 If (left (r1) is not equal to r2) decrement r1 by 2
7 If r1 is not equal to r2 increment r2 by 2
8 If ranking is <=0 Eliminate the rules
9 Store the rules with ranking

The SNEMiner Ranking algorithm implemented using WEKA mining tool to generate the elastic rules with ranking. In the preprocess operation, extracted variables are loaded into WEKA explorer. The threshold used for minimum support

Table 1 Rules generated for the dataset

Dataset1	Dataset2
{10} =>{20}	{10} =>{20},{10} =>{50}
{10,20} =>{40}	{10} =>{40},{10} =>{30}
{10,30} =>{50}	{10,40} =>{60},{10,40} =>{70}
{10,40} =>{50}	{90,40} =>{70}

Table 2 Ranking table for the rules using SNEMiner Ranking algorithm

1 Rated	2 Rated	3 Rated	4 Rated	5 Rated
{10} =>{20}	{10,40} => {50}	{10,30} => {50}	{10} =>{40}	{90,40} => {70}
{10,20} => {40}			{10} =>{50}	
			{10,40} => {60}	
{10} =>{30}			{10,40} => {70}	

is 0.5, and the minimum confidence is 0.6. It involves 12 iterations to generate the rules. The rule that does not meet minimum threshold values is discarded. Based on the minimum threshold values, the valid rules are generated with ranking increases the disability student's performance with the integration of ICT and AT.

Table 2 shows the ranking for the rules using the SNEMiner Ranking algorithm. The best rule is given the rating 1 and the rules which are less strong are rated as 2, 3 and 4. The worst rule is given the rating 5.

3.3 Interestingness Measure for SNEMiner

SNEMiner Ranking algorithm is used to classify rules based on the ranking table. It demonstrates the strength of a particular rule. The best rule of SNEMiner Ranking algorithm is obtained by measuring the interestingness measure using a key parameter. It is computed and displayed the rule with thresholds. Given a sequences database, the notation sequence id (A =>B) represents the sequence where the rule occurs that has support and confidence greater than or equal to the minimum thresholds, respectively.

Support measures the proportion of connections in the database that include both A and B. Equation 1 reduces the number of rules using support metrics. Equation 1 shows the downward closure property in the database that includes both the datasets.

$$\text{Support}(\text{A} => \text{B}) = \text{P}(\text{AUB}) \tag{1}$$

Equation 2 indicates the confidence metrics for the transaction in the datasets A and B. In that, minimum confidence threshold value is set and lifted the rules. It is not downward closure.

$$\text{Confidence}(\text{A} => \text{B}) = \text{P}(\text{B}/\text{A}) \tag{2}$$

Rules are strong, when it was satisfied by the threshold.

$$\text{SNEMiner} = \text{con1}(\text{confidence}(r) + (\text{con2}(\text{support}(r)*\text{length}(r)))) \quad (3)$$

From Eq. 3, SNEMiner value is calculated with constraints con1 and con2, r1, r2 € r represents the rules, support(r) and confidence(r) represent the calculated value for the threshold and length (r) represents the length of particular rule. The SNE-Miner used to predict the rules for the performance of disability students.

The measures lift and cosine used to give the recommendations for the weakest rules. In the lift, value > 1 indicates A and B present in both left and right rule otherwise less often together than expected. Its value is to obtain between 0 and infinity. The disqualified rule using lift value can assess their dependence range. Cosine measure is the geometric factor used in the vector space models. It is used to measure the similarity between dataset1 and dataset2. The closer cosine A => B is to 1, then the more transaction Item A contains Item B. Closer cosine A =>B is to 0, and then the more transaction Item A without containing Item B.

4 Experimental Analysis

The study contains 92 disability student details of which 52 students were men and 40 students were female. Table 3 shows that interestingness measures under the support, confidence, cosine and lift values entered from Table 2. In Table 3, the values of ICT, AT, ICT with AT and no technology are listed in the columnwise and these technologies are pointed with the corresponding values used in the interestingness measures.

Based on the interestingness measure, the experimental result in Fig. 2 shows that disability students improved their performance with the integration of ICT and AT technology. It shows the highest performance level for the disability students proved by the interestingness measure of SNEMiner Ranking algorithm.

Table 3 Interestingness measure with role of technology

Role of Technology	Interestingness Measures			
	Support	Confidence	Cosine	Lift
ICT	0.65	0.8	0.85	1.32
AT	0.6	0.75	0.73	1.25
Both ICT and AT	0.8	1	1	1.85
Not using any Technology	0.55	0.65	0.55	0.93

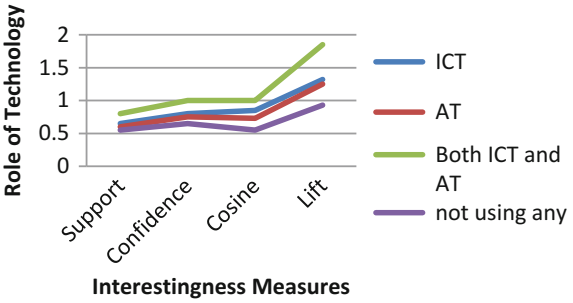


Fig. 2 Experimental analysis of interestingness measure for the performance level of disability students

5 Conclusion

In this research work, SNEMiner Ranking algorithm allocates ranking to the rules and it modified when some new rules were appended and unwanted rules were deleted. Interestingness measures for SNEMiner under the SNEMiner Ranking algorithm used to classify the rules. Based on this, metrics conclude the elastic rules and prove the performance increase in intellectual students. In this paper, experimental analysis shows that disability student used both the technology ICT and AT with integration which shows the highest performance level for inclusive education.

References

1. Capitao, S., Alneida, A.M.P., Vieira, R.M.: Connecting families and schools of students with deafness: describing the ICT and internet use in education. Elsevier Procedia Computer Science, vol. 14, pp. 163–172. Elsevier, Amsterdam (2012)
2. Aroyo, L., Dicheva, D.: The new challenges for e-learning: the educational semantic web. Educ. Technol. Soc. **7**, 59–69 (2004)
3. Baker, R.S.J.D.: Data mining for education. International Encyclopedia of Education, vol. 7, pp. 112–118 (2010)
4. Obradovi, S., Bjeki, D., Zlati, L.: Creative teaching with ICT support for students with specific learning disabilities. Elsevier Procedia – Social and Behavioral Sciences, vol. 203, pp. 291–296. Elsevier, Amsterdam (2015)
5. Dawson, S., MAcfadyen, L., Lockyer, L.: Learning or performance: predicting drivers of student motivation, University of Wollongong, pp. 184–193 (2009)
6. Fournier-Viger, C.-W.W., Tseng, V.S., Cao, L., Nkambou, R.: Mining partially – ordered sequential rules common to multiple sequences. IEEE Trans. Knowl. Data Eng. **27**(8), 2203–2216 (2015)
7. Fournier-Viger, P., Nkambou, R., Tseng, V.S.-M.: RuleGrowth: mining sequential rules common to several sequences by pattern growth. In: ACM Symposium on Applied Computing SAC 11, pp. 956–961 (2011)
8. Garcia, E., Romera, C., Ventura, S., Calders, T.: Drawbacks and solutions of applying association rule mining in learning management systems. In: International Workshop on Applying Data Mining in e- Learning (2007)

9. Faghihi, U., Fournier-Viger, P., Nkambou, R.: A cognitive tutoring agent with episodic and casual learning capabilities. In: International Conference on Artificial Intelligent in Education, pp. 72–80 (2011)
10. Pham, T.-T., Luo, J., Hong, T.-P., Vo, B.: An efficient algorithm for mining sequential rules with interesting measures. Int. J. Innov. Comput. Inf. Control **9**(12) (2013)
11. Romero, C., Ventura, S.: Data mining in education. WIREs Data Mining Knowledge Discovery, pp. 12–27 (2013)
12. Ahmed, M., Tariq Banday, M., Jan, T.R.: ICT4ELT: answer study with reference to Kashmir. Elsevier Procedia – Social and Behavioral Sciences, vol. 123, pp. 414–421 (2014)
13. Andreagiovanni, M.D., Baiardi, F., Lipilini, J., Ruggieri, S., Tonelli, F.: Sequential pattern mining for ICT risk assessment and prevention, Dipartimento di Informatica, Universita di pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy
14. Simic, V., Vojinovic, O., Milentijevic, I.: E-learning let's look around. Ser. A Appl. Math. Inform. Mech. **3**(2), 121–138 (2011). (Scientific Publications of the State University of Novi Pazar)
15. Klema, J., Novakova, L., Karel, F., Stepankova, O., Zelezny, F.: Sequential data mining: a comparative case study in development of atherosclerosis risk factors. IEEE Trans. Syst. Man Cybern. Appl. Rev. **38**(1) (2008)
16. Marquez-Bajra, J.M., Jourjon, G., Mikroyannidis, A., Tranoris, C., Domigue, J., Dasilva, L. A.: FORGE: enhancing learning and research in ICT through remote experimentation. In: IEEE Global Engineering Education Conference, E-ISBN 978-1-4799-3191-0 (2014)
17. Nikolopoulos, S.D. Dr.: Data mining association rules of ICT's adoption factors by greek accountants. In: 11th MIBES Conference- Heraklion, Crete, Greece, pp. 22–24 (2016)
18. Merceron, A., Yacef, K.: Interestingness measures for association rules in educational data. In: International Conference on Education Data Mining, Montreal, Canada, pp. 57–66 (2008)
19. Bekteshi, L.: Information and communication technology and students with disabilities. Eur. Sci. J. **11**(22) (2015)
20. Soderstrom, S., Ytterhus, B.: The use and non-use of assistive technologies from the world of information and communication technology by visually impaired young people: a walk on the tightrope of peer inclusion. J. Disabil. Soc. **25 (**2010)
21. Bravo, J., Romero, S.J., Luna, M., Pamplona, S.: Exploring the influence of ICT in online students through data mining tools. Int. J. Eng. Educ. 1622–1628 (2015)
22. Jabbar, M.A.: Information and communication technology (ICT) for health care in India: challenges and solutions. In: Computer Society of India, Knowledge Digest for IT Community, vol. 41, Issue 5 (2017)
23. Singto, P., Mingkhwan, A.: ICT career analysis using association rule. In: IEEE Xplore, International Conference on Digital Information Management, pp. 141–144 (2014)

A New Approach for Line Loadability Enhancement in Restructured Power System

D. Ragaleela and S. Sivanagaraju

Abstract The transmission network plays an important role in deregulated electricity markets because it is the key mechanism for generators to compete in supplying large users and distribution companies. Congestion in transmission network could be one of the primary problems of restructured electric industry. This paper proposes a new method to enhance the line loadability by congestion management using power vector graph method. In some methods like sensitivity analysis, only active power is taken to do congestion management so those methods are not accurate. But here, both active power and reactive power are considered for doing congestion management. The active and reactive power flow of individual sources in the network and contribution of each source to each line is evaluated by the use of superposition principle. Power vector graph is designed with the help of the power flows in the context of a competitive electricity market. Congestion can be reduced to enhance line loadability by adjusting relevant power components in four quadrants. Proposed method is illustrated on IEEE 30 bus system using MATLAB.

1 Introduction

In electric power industry, generation, transmission and distribution of power have single authority for its operation and control. Such systems are called as vertically integrated systems. Separation of large vertically integrated system is called restructured system or deregulated system. In deregulated system, competition exists in generation and distribution systems but not transmission system.

D. Ragaleela (✉)
Department of Electrical and Electronics Engineering, P.V.P. Siddhartha Institute of Technology, Kanuru, Vijayawada 520007, Andhra Pradesh, India
e-mail: raga_233@yahoo.co.in

S. Sivanagaraju
Department of Electrical and Electronics Engineering, JNTUK, Kakinada, Andhra Pradesh, India
e-mail: sirigiri70@yahoo.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_49

The transmission system is monopoly. Because regulation is necessary in transmission system in order to avoid over charging for its services. Due to the competition in deregulated system, electric power charge decreases and increases the quality of supply compared with regulated system [1].

Transmission lines capability to transmit electric power is constrained by several transfer limits such as voltage limit, thermal limit and stability limit. If the parameters of transmission line exceed any one of these limits, then the system is said to be congested. An enhance in power demand, sudden outages of generation, restriction on the erection of new lines, unanticipated power flow in the lines, transmission lines tripping or malfunctions of other equipments are some of the potential reasons for congestion [2, 3]. Transmission congestion system cannot be permitted over a scant duration as it can initiate cascaded outages which push the system to collapse. Hence, an effective congestion management is necessary to diminish the line overloads to the security limit in minimal time [4]. Congestion management can be done by different methods. Some of them are given below [5]:

(1) Cluster/zone-based approach; (2) transmission congestion distribution factors (TCDFs); (3) relative electrical distance (RED) method

(4) Domain of generator; (5) power flow tracing (6) load curtailment

Power flow tracing is used to find out, how much of power is flowing from particular source to particular load. It also finds the contributions of individual generators (or load) to individual line flows [6]. Power flow tracing can be done by Bialek's tracing method [7], Kirschen's tracing method [8] and power flow tracing using graph theory, etc.

In this paper, the concept of virtual flows is used to evaluate the contribution of each generator to line flows based on the principle of superposition, i.e., each transaction is calculated with one source at a time with all loads connected. Similarly calculations are done for all sources; then, the algebraic sum of transactions of all the sources gives the actual transaction. Thereafter, congestion management is computed with the use of the virtual flows using power vector graph method.

2 Congestion Management Model

Here, congestion in the transmission system is created by removing a line. By removing a line, the power supplied to the load by that line is transfered to other lines. So more power will be carried by the other lines than the previous [9]. Therefore, congestion may occur if power flowing in lines exceeds its transmission capacity limit.

2.1 Concept of Virtual Flows Based on the Principle of Superposition

In this section, Virtual network flows in a line due to various sources is evaluated using the principle of superposition. Superposition is a fundamental concept from

circuit theory as applicable to linear and bilateral network containing sources and loads [10]. This approach is easy and direct but needs solution of network equation with the use of sparse Y bus, and the number of times to be repeated is equal to number of sources. The initial data required are taken from power flow analysis for evaluation. Here, Newton–Raphson method is used to perform the power flow analysis. The bus voltage phasors (p.u), real and reactive power injections at all generator buses, connected loads and the network parameters are obtained.

Calculate modified Y bus matrix. It is obtained by adding loads as equivalent admittances to the Y bus matrix.

$$Y_{\text{mod}} = y_{bus} + y_k \tag{1}$$

Equivalent admittance of load at *k*th node

$$Y_k = S_k^* / |V_k|^2 \tag{2}$$

Current injected from ith source to corresponding bus

$$I_i = [S_{ci}/V_i]^* \tag{3}$$

The following equation is used to find out the bus voltage contribution by *i*th source when injecting equivalent current.

$$[Y_{\text{mod}}][V^n]_i = [I^n]_i \tag{4}$$

The vector sum of complex bus voltage contributions of individual sources for equivalent current injections is same as the original bus voltage profile of the network

$$\left[V^0\right] = \sum_{i=1}^{g} [V^n]_i \tag{5}$$

Branch currents supplied by particular source

$$I_{ij}^n = -(V_i^n - V_j^n) \times Y_{ij} \tag{6}$$

Branch power flows supplied by particular source

$$P_{ij}^n = V_i^0 \times (I_{ij}^n)^* \times BaseMVA \tag{7}$$

The line power flows in branches due to individual source action sum up to line power flow profile as obtained from initial condition.

$$\left[P_{ij}^0\right] = \sum_{i=1}^{g} \left[P_{ij}^n\right]_i \tag{8}$$

where

g = Total no of generator buses; V^0 = Initial bus voltage
y_k = Equivalent admittance of load at kth node; V_k = Bus voltage at k th node
S_{Ci} = Complex power injection to the network from ith source
S_k = Initial complex power of the load bus
I_{ij} = Current flow in line i–j supplied by the ith source
P_{ij} = Power flow in line i–j supplied by the ith source
y_{bus} = Bus admittance matrix of the network; Y_{mod} = Modified Y bus matrix
$[I^n]_i$ = The bus current injected from source when it is only doing transaction.
$[V^n]_i$ = The bus voltage developed for current injection from ith source individual action

2.2 Power Vector Graph Method

In power systems, the transmission capacity limit of each line is given in apparent power S = √(P^2 + Q^2). Q_{min} and Q_{max} can be calculated from voltage constrains, then P_{max} and P_{min} can also be calculated. The radius of the circle is obtained from the transmission capacity limit as shown in Fig. 1. If the power vector is in the shadow area in the circle means it does not exceed the transmission capacity limit and satisfies the voltage constraints. If the actual power is not in the shadow area, some power components have to be adjusted repeatedly till it enters. Adjustment of power components in four quadrants is required in this method [11]. The output of generator can be divided into two categories, first one is the cause of congestion (first and fourth quadrant power components) and second one is the relieve of congestion (second and third quadrant power components). If the power component length is greater, its influence on the line is greater. The power components having the longest length in the first and fourth quadrants should be cut. Similarly, the power components having the longest length in the second and third quadrants should be increased. In case the second and third quadrants have no power components, and then have to reduce the load and generator output. If there exists congestion, the following steps are to be considered for doing congestion management.

1. Make the power component vector graph with the power flow tracing result.
2. Judge which generator output is to cause congestion and which generator output is to relieve congestion
3. Calculate the gains of the generator or load which can participate in the adjustment.

$$G_{ij-ns} = \frac{S_{ij-ns}}{S_{ns}}, n \in G \tag{9}$$

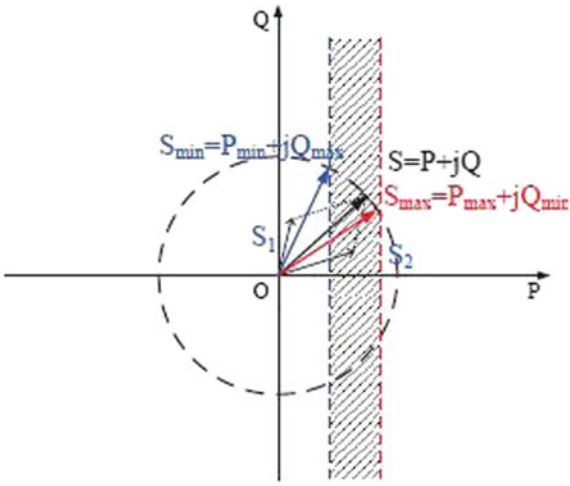


Fig. 1 Representation diagram of power vector graph

where

$G_{ij\text{-}ns}$ = Gain of the generator; $S_{ij\text{-}ns}$ = Power of i–j line due to nth source; S_{ns} = Output of the nth source

4. Calculate adjustment amount

$$\Delta S_{ij} = \frac{S_{ij} - \overline{S}_{ij}}{G_{ij-ns}^{+\max} - G_{ij-ns}^{-\max}} \tag{10}$$

Where

S_{ij} = Power of i–j line; $\overline{S}_{ij}$ = Transmission line capacity limit;
$G_{ij\text{-}ns}^{+max}$ = Gain of the generator which cause congestion;
$G_{ij\text{-}ns}^{-max}$ = Gain of the generator which relieve congestion

5. By the use of adjustment amount adjust the output of the generator or load. The process is repeated till the congestion is eliminated completely.

3 Simulation Result

An IEEE 30 bus system is considered in order to explain proposed concept. The line flows during normal case are obtained from Newton–Raphson load flow using MATLAB as shown in Table 1.

Table 1 Line flows of IEEE 30-bus from N-R load flow method

S.No.	From bus	To bus	Line power (MVA)
1	1	2	56.308−j8.752
2	1	8	42.603+8.513
3	2	11	30.511+j10.081
4	8	11	39.429+j7.045
5	2	5	45.041+j9.447
6	2	13	38.491+j8.752
7	11	13	35.172−j4.973
8	5	7	−0.075+j6.661
9	13	7	23.036+j2.843
10	13	3	12.016−j1.905
11	9	4	−20.000−j11.373
12	9	10	31.753+j14.785
13	12	6	−20.000-j8.085
14	12	14	8.267+j3.059
15	12	15	18.991+j9.572
16	12	16	7.966+j6.531
17	14	15	1.978+j1.274
18	16	17	4.373+j4.536
19	15	18	6.566+j3.245
20	18	19	3.310+j2.231
21	19	20	−6.200−j1.190
22	10	20	8.484+j2.073
23	10	17	4.654+j1.357
24	10	21	15.887+j9.401
25	10	22	7.680+j4.198
26	21	22	−1.729−j2.049
27	15	23	5.915+j4.545
28	22	24	5.896+j2.034
29	23	24	2.661+j2.836
30	24	25	−0.208−j1.940
31	25	26	3.547+j2.370
32	25	27	−3.762−j4.323
33	27	29	6.194+j1.676
34	27	30	7.097+j1.672
35	29	30	3.705+j0.608
36	3	28	1.998+j0.349
37	13	28	15.131+j0.123
38	13	9	11.753−j7.586
39	13	10	10.752+j0.255
40	11	12	26.425−j10.304
41	28	27	17.089+j0.346

4 Line Loadability Enhancement by Congestion Management

Remark 1 Single line congestion [12] is occurred at 13–7 line by removing 2–5 line. Power contribution from all generators to congested line (13–7 line) is obtained from Eqs. (1)–(8) as shown in Table 2.
Contribution of source power to line 13–7 in power vector graph is shown in Fig. 2.

Figure $S_{13,7\text{-}1s}$ has the longest length presented in 4th quadrant, and $S_{13,7\text{-}5s}$ has the longest length presented in 3rd quadrant. So to enhance line loadability and relieve congestion generator-1, output has to decrease and generator-5 output has to increase by the use of adjustment amount. Gain, from Eq. (9)

$G_{13,7\text{-}1s} = 0.3591\text{-}j0.0451$; $G_{13,7\text{-}5s} = -0.5220\text{-}j0.1638$

Adjustment amount, from Eq. (10) is $\Delta S_{13\text{-}7} = 42.4275\text{-}j17.7940$

Line power after adjustment of output of generators is 30.0000-j0.6290 MVA as shown in Fig. 3.

Remark 2 Similarly, multiple line congestion is occurred at 13–7 line by removing 2–5 line & 9–10 line and congestion is occurred in 13–7 line and 2–13 line by removing 2–5 line and 2–11 line. Table 3 shows the power flows in the lies during congestion and after congestion management using power vector graph method for line loadability enhancement

Table 2 Power contribution from all sources to the line 13–7 during congestion

Source-1 (MVA)	Source-2 (MVA)	Source-3 (MVA)	Source-4 (MVA)
36.7172−j6.2261	29.7631+j6.0091	7.6272+j12.3207	7.2016+j4.8228
Source-5 (MVA)	Source-6 (MVA)	Total power in line during Congestion	Total power in line during normal case
−18.7888−j31.4918	6.9787+j3.9285	69.497-j10.640	23.036+j2.843

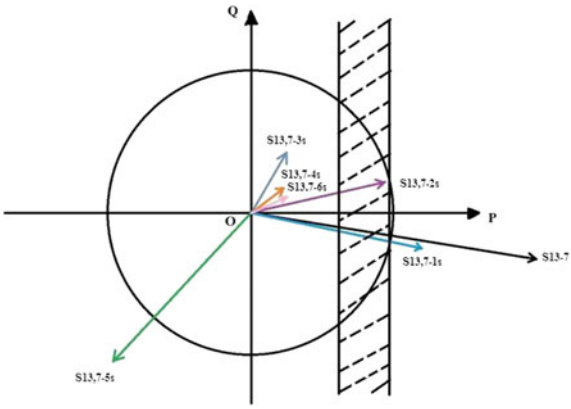


Fig. 2 Power vector graph during congestion for 13–7 line

Table 3 Line power flows during congestion and after congestion management

Remarks	Congested line	Power flow (MVA)		Transmission line capacity limit (MW)
		During congestion	After doing congestion management	
Single line contingency (2–5)	13–7	69.497-j10.640	**30.0000-j0.6290**	30
Multiple line contingency (2–5 & 9–10)		69.515-j11.086	**30.0001-j0.8654**	
Multiple line contingency (2–5 & 2–11)	13–7	69.581-j12.684	**29.9521+j0.0001**	30
	2–13	86.505+j1.303	**64.0222 +j10.3395**	65

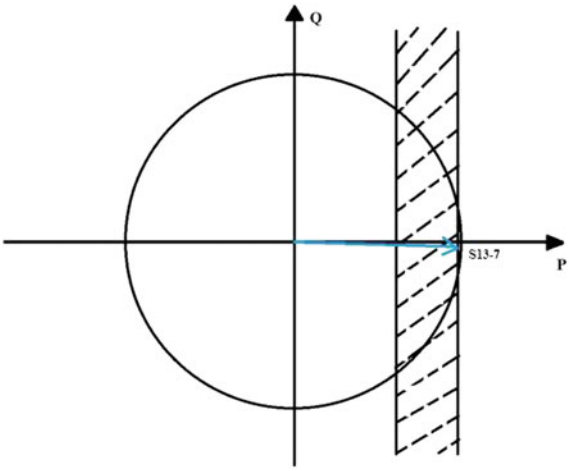


Fig. 3 Power vector graph after doing congestion management for 13–7 line

5 Conclusions

Congestion may occur due to the line outage or unit outage. So here line contingency is considered for creating congestion. Each generator output contributions to transmission lines are obtained by power flow tracing using super position principle. Here, line loadability enhancement is done by using power vector graph method. In this method, congestion is eliminated effectively by adjusting the power components in four quadrants. The information from the flows computed by the proposed approach is valuable for re-dispatch of generation and overload alleviation based on economics, environmental issues or any other creation. This technique can be used for loss allocation, reactive power optimization, etc.

References

1. Battacharya, K., Bollen, M.H.J., Daalder, J.E.: Operations of Restructured Power Systems. Kluwer Academic Publishers, Dordrecht (2001)
2. Verma, S., Mukherjee, V.: Optimal real power rescheduling of generators for congestion management using a novel ant lion optimizer. IET Gener. Transm. Distrib. **10**(10), 2548–2561 (2016)
3. Balaraman, S., Kamaraj, N.: Real time contingency management in deregulated electricity market using artificial neural network. Iran. J. Electr. Comput. Eng. **10**(1) (2011)
4. Ahamed Jeelani Basha, A., Anitha, M.: Power flow tracing based congestion management using firefly algorithm in deregulated electricity market. Int. J. Eng. Res. Dev. **12**(5), 38–48 (2016)
5. Mandala, M., Gupta, C.P.: Comparative studies of congestion management in deregulated electricity market. In: 16th National Power Systems Conference, Hyderabad (2010)
6. Shahidehpour, M., Yamin, H., Li, Z.: Market Operations in Electric Power Systems. Wiley, New York (2002)
7. Bialek, J.: Tracing the flow of electricity. IEE Proc. Gener. Transm. Distrib. **143**(4), 313–320 (1996)
8. Kirschen, D., Allan, R., Strbac, G.: Contribution of individual generators to loads and flow. IEEE Trans. Power Syst. **12**(1), 52–60 (1997)
9. Khan, M.T., Siddiqui, A.S.: Congestion management in deregulated power system using FACTS device. Int. J. Syst. Assur. Eng. Manag. **8**(1), 1–7 (2017) (Springer India)
10. Adhip, D.T.: Congestion management based on virtual real power flows. In: 2nd Biennial International Conference on Power and Energy Systems towards Sustainable Energy (PESTSE). IEEE Press, Bangalore (2016)
11. Wang, H., Bao, H.: A new method of congestion management based on power flow tracing. In: 2016 China International Conference on Electricity Distribution (CICED), (CICED 2016) Xi'an. IEEE Press, China (2016)
12. Rajalakshmi, L., Suganyadevi, M.V., Parameswari, S.: Congestion management in deregulated power system by locating series FACTS devices. Int. J. Comput. Appl. **13**(8), 19–22 (2011)

A Robust Gene Data Classification Model Using Modified Manhattan Distance-Based Weighted Gene Expression Graph Classifier

N. Sevugapandi and C. P. Chandran

Abstract Many computational algorithms were introduced to interpret the gene expressions data, and most of them were not robust enough to scale and classify large-scale gene population. Hence, a novel modified Manhattan distance-based weighted gene expression graph (GEG) classifier is proposed. Here, the gene data points were considered as one of the prominent features and extracted using the proposed modified Manhattan distance. Here a new classification scheme is attempted by combining modified Manhattan distance based on weighted GEG gene graph classifier. Further, the proposed model is experimentally validated and its performance is compared with conventional classifiers like naive Bayes, SVM and random forest, etc. The results show that the proposed model is robust and achieves better accuracy in classification.

1 Introduction

Data mining and bioinformatics provide the important and flourishing research area as well as development area for computational science. Bioinformatics means taking out needed information from enormous amount of biological datasets using computational and statistical techniques [1].

A DNA microarray is a solid slide, in which the DNA series are located in an organized way. A microarray is capable of holding up to thousands of DNA probes. By using DNA microarray, scientists can compute expression levels for thousands of gene expression activity at a time, which will facilitate to discover the complex

N. Sevugapandi (✉)
Research and Development Center, Bharathiar University, Coimbatore, Tamil Nadu, India
e-mail: sevugapandi1985@gmail.com

C. P. Chandran
Department of Computer Science, Ayya Nadar Janaki Ammal College,
Sivakasi, Tamil Nadu, India
e-mail: drcpchandran@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_50

relationship among the genes. It will be helpful to create databases for cancer tissues and normal tissues [2].

With the combination of gene expression graph (GEG) and Manhattan distance, a novel classification algorithm for above-stated problem was proposed. The existing method deals with the gene expression data. The possible view about the relationship between normal and unhealthy genes could be provided by gene expression graph. Depends on an association between GEG for familiar classes and graphs are decoded as samples to evaluate the construction of classifier. Additionally, Manhattan distance is used to measure the strength of relationship which exists between the genes [3].

The main shortcomings in all these above methods are limited while using for clinical analytics application. Moreover, these methods failed to distinguish class samples [4]. The proposed method is going to improve graph-based data structure by adding Manhattan distance to determine the relationships among genes. In this method, relevant genes are extracted by means of weight assigned to each gene, where larger weights point out a healthier rapport flanked by two genes. As a result, the increasing method is able to decrease the price acquired through classifying appropriate genes. One of the major works of the proposed method is having ability to appropriately categorize into the relevant classes and also properly distinguish class samples. In addition, it will significantly reduce computation time taken for classifying the data.

2 Methodology

2.1 Proposed System

The proposed weight-based gene expression graph performs classification of diseases by means of gene expression data along with carrying weight at the edge. In graph G(V, E), V represents the gene and the E represents the weighted edge. The weight present in the edge determines the strength of relationship among the gene. The graph is built from untreated microarray scanned image data to go through the following stages such as preprocessing, data modeling, classification and validation. The first level, cy3, represents the healthy condition of gene, and the second expression level, cy5, represents the diseased condition of gene [5]. This gene expression usually attained from the n samples is arranged as matrix called gene expression matrix. In preprocessing stage, this is done by means of normalization techniques. Normalization is particularly useful to train the input values for classification algorithm; this normalization technique helps to speed up the learning phase in classification algorithm [6]. There are many methods for data normalization. For that, the authors noted the standardized normal distribution techniques along with LN ratios of each gene. It can be computed by using Eqs. (1)–(4)

$$\text{LNratio(x)} = \frac{\ln(\text{cy5(x)}/\text{cy3(x)})}{\log\ 5} \tag{1}$$

$$\text{Mg} = \frac{1}{n}(x_1 + x_2 + x_3 + \cdots + x_n) \tag{2}$$

$$\underset{1}{\text{Sg}} = = \frac{1}{n}(|x_1 - \text{Mg}| + |x_2 - \text{Mg}| + |x_3 - \text{Mg}| + \cdots + |x_n - \text{Mg}|) \tag{3}$$

$$\text{Z}(\text{LNratio}(x), \text{M}_\text{g}, \text{S}_\text{g}) = \frac{\text{LNratio(x)} - \text{M}_\text{g}}{\text{S}_\text{g}} \tag{4}$$

where Mg represents the mean, S_g represents the standard deviation and LNratio(x) represents the input value x. For example, Table 1 considers the expression levels training set to compute standardized normal distribution.

Then relevancy between genes is computed by finding the relevancy of each gene by calculating Relevance Count (RC) [7]. Based on TRC value, the gene relevancy is categorized into 3 types. From Table 2, they are more relevant gene (1), quieted gene (–1) and irrelevant gene (0). RC is calculated by using threshold, cy5 and cy3 values. A threshold ε can be specified by using well-known methods for example, t test. Total Relevancy Count (TRC) is by doing summation of ll RC value in each gene. It is defined in Eq. 5. Table 3 shows the TC value for each gene

Table 1 Expression levels training set

Gene tissue	a		b		c	
	Diseased	Healthy	Diseased	Healthy	Diseased	Healthy
S1	30	10000	100	12000	2000	10
S2	50	30000	50	2000	1000	1099
S3	30	5000	13000	80	4000	150
S4	100	6000	80	15000	6000	25

Table 2 LNratio and normalization of the training dataset

Gene	LNratio(x)			Normalization (LNratio(x), M_g, S_g)			Mean	SD
	[1]a	[1]b	[1]c	[1]a	[1]b	[1]c	M_g	S_g
S1	−8.31	−6.85	7.58	−0.86	−0.64	1.5	−2.53	6.74
S2	−9.15	−5.28	−0.09	−1.43	−0.25	1.32	−4.44	3.30
S3	−7.32	7.28	4.7	−1.5	0.97	0.43	1.55	5.92
S4	−5.86	−7.49	7.84	−0.62	−0.88	1.5	−1.84	6.45

[1]a, b, c refers the gene tissue expression level

Table 3 Total Relevancy Count (TRC) value for threshold $\varepsilon = 0.5$

Gene	a	b	c
S1	−1	−1	1
S2	−1	0	1
S3	−1	1	0
S4	−1	−1	1
TRC	−4	−1	3

Table 4 The weight for gene expression tree as matrix

Gene	a	b	c
a	0	3	3
b	3	0	2
c	3	2	0

$$\mathrm{RC}(\varepsilon, \mathrm{cy5}, \mathrm{cy3}) = \begin{cases} 1 & \mathrm{Z}\left(\mathrm{LNratio(x)}, \mathrm{M_g}, \mathrm{S_g}\right) > \varepsilon \\ 0 & -\varepsilon \le \mathrm{Z}\left(\mathrm{LNratio(x)}, \mathrm{M_g}, \mathrm{S_g}\right) \le \varepsilon \\ -1 & \mathrm{Z}\left(\mathrm{LNratio(x)}, \mathrm{M_g}, \mathrm{S_g}\right) < -\varepsilon \end{cases} \tag{5}$$

The weight (v_a, v_b) between the genes are calculated by applying modified Manhattan distance formula. It is defined in Eq. 6. Table 4 represents the weight gene expression matrix [8].

$$\mathrm{WG}(v_a, v_b) = |v_{a1} {}^{*} v_{b1}| + |v_{a2} {}^{*} v_{b2}| + |v_{a3} {}^{*} v_{b3}| + \cdots + |v_{an} {}^{*} v_{bn}| \tag{6}$$

3 Experimental Results

The proposed methodology was experimentally tested using microarray dataset downloaded from UCI-CBCL research laboratory [9]. The dataset consists of 1000 gene profiles with both healthy and cancer-infected gene profiles which were used for training the proposed classifier and compare the other models such as random forest, SVM, naive Bayes, etc [10]. Figure 1 shows the results of varying threshold used for optimal feature selection (Fig. 2) by calculating modified Manhattan distance. The experimental results in Figs. 3, 4, 5 and 6 clearly show that the proposed model outperforms the conventional classifiers where the results obtained were 16.02% with absolute improvement from the conventional classifiers. A comparative analysis was carried out between the proposed GEG classifier and conventional classifiers, and from the observation the proposed GEG classifier achieves very low error rate in feature variance with 9.9% of the target genes as stated in Fig. 3. The proposed model is also tested in UCI-CBCL dataset which consists of 1650 gene

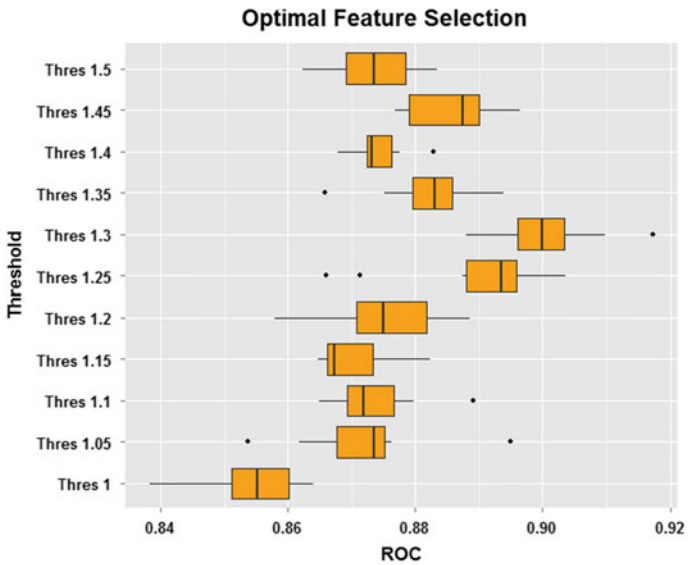


Fig. 1 Varying threshold used for optimal feature selection

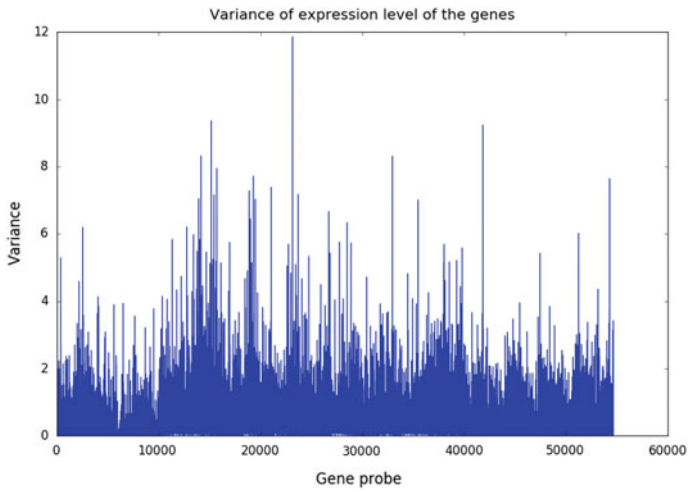


Fig. 2 Feature variance for UCI-CBCL dataset

profiles; still the improvement of the proposed GEG classifier is absolute with the rate of improvement of 7.8% and the error rate of 12.3%.

From Fig. 3 it is clearly shown that the low-level feature considered for gene classification is not enough to classify the cancer-infected genes with low-level indicators. From Fig. 4, it is clearly shown that the high-level feature considered for

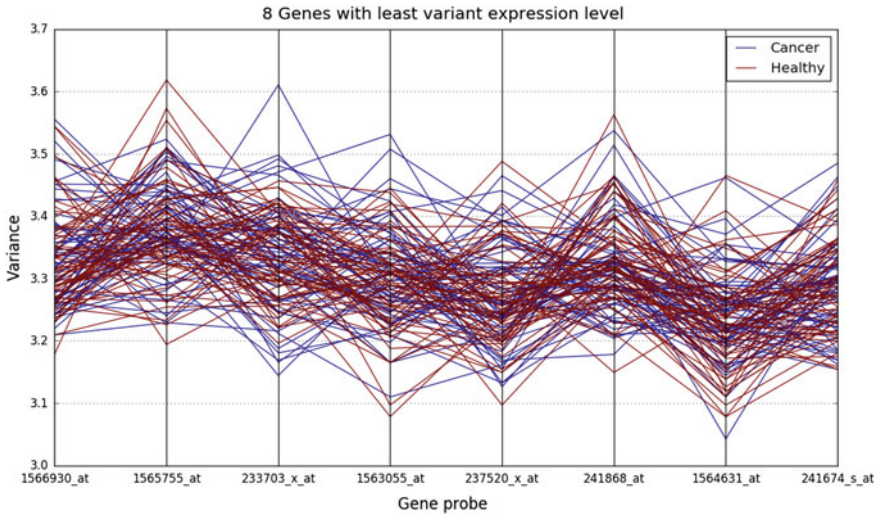


Fig. 3 Feature variance with low-level indicator for UCI-CBCL dataset

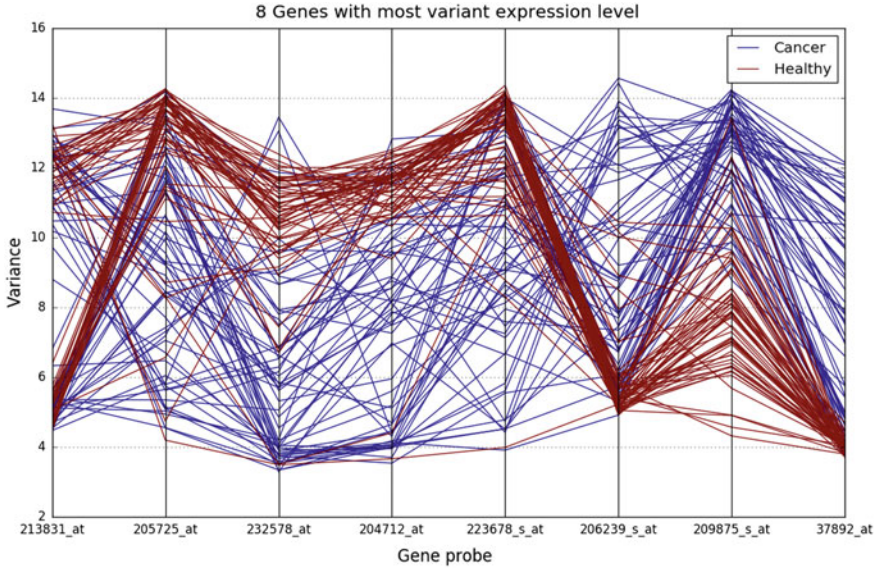


Fig. 4 Feature variance with high-level indicator for UCI-CBCL dataset

gene classification is robust enough to classify the normal gene ensembles from cancer-infected genes. The microarray-stored profiles saved in UCI-CBCL dataset are accurately classified, and results obtained were promising. The proposed model is also tested in UCI-CBCL dataset which consists of 1650 gene profiles; still the improvement of the proposed GEG classifier is absolute with the rate of improvement of 7.8% and the error rate of 12.3% as stated in Fig. 5. Finally, Fig. 6

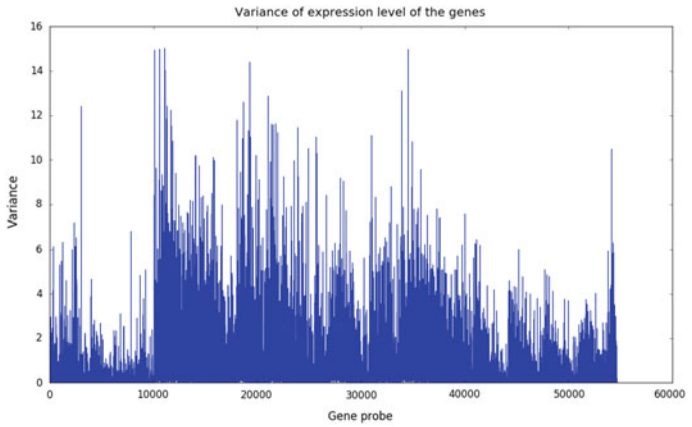


Fig. 5 Feature variance for UCI-CBCL dataset

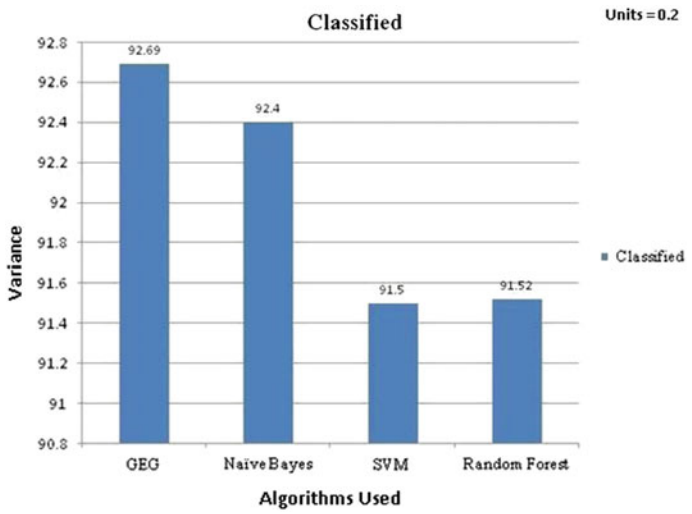


Fig. 6 Perform analysis plot for proposed GEG versus conventional classifiers

shows the performance comparison of the proposed GEG classifier versus conventional classifiers.

4 Discussion

Michael Young and David Craft [11] introduced the pathway gene classification systems for the analysis of cancer using gene expression data. Ayad Gahnee Ismail and Anar Auda Ablahad [12] say that, A novel method can predict the disease by mutations despite, the presence in gene sequence is not necessary it are malignant, so will be compare the patient's protein with the gene's protein of disease if there is difference between these two proteins then can say there is malignant mutations. Jianzhen Xu and Yongjin Li [13] says that the new opportunity for discovering hereditary disease genes with help of topological features in protein–protein interactions (PPIs) network could be provided by the availability of human genome-wide PPIs. Steffen Klasberg et al. [14] proposed that, Computational approaches are further prime methods that can be based on existing models or leveraging biological evidences from experiments. Wenying Yan et al. [15], the biological networks underlying cancer is undoubtedly important for understanding the molecular mechanisms of the disease and identifying effective biomarkers. Kavita Ganesan et al. [16] say that it finds an unsupervised realistic approach to mine related concepts by influencing the quantity within large sum of experimental notes.

5 Conclusion

This paper proposed a modified Manhattan distance-based weighted GEG classifier to classify a gene expression data. The experimental outcome confirms that the efficiency of the proposed method gives better result than the existing approaches. The suggested method is able to distinguish appropriately among the class samples. More than that, the suggested method could properly classify samples in the relevant classes. From the experimental results, it is clearly shown that the proposed GEG classifier outperforms well in terms of classification accuracy than the conventional classifiers.

References

1. Benso, A., Di Carlo, S., Politano, G., Sterpone, L.: A Graph-based representation of gene expression profiles in DNA microarrays. In: Proceedings of IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), pp. 75–82 (2008)

2. Breiman, J., ad Friedman, L., Stone, C.J., Olshen, R.: Classification and Regression Trees. Taylor and Francis, New York (1984)
3. Zhang, H., Yu, C.-Y., Singer, B.: Cell and tumor classification using gene expression data: construction of forests. Proc. Nat. Acad. Sci. USA **100**(7), 4168–4172 (2003)
4. Vapnik, V.N.: An overview of statistical learning theory. IEEE Trans. Neural Netw. **10**(5), 988–999 (1999)
5. Khan, J., Wei, J.S., Ringner, M., Saal, L.H., Ladanyi, M., Wester-mann, F., Berthold, F., Schwab, M., Antonescu, C.R., Peterson, C., Meltzer, P.S.: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat. Med. **7**(6), 673–679 (2001)
6. Benso, A., Di Carlo, S., Politano, G.: A cDNA microarray gene expression data classifier for clinical diagnostics based on graph theory. IEEE/ACM Trans. Comput. Biol. Bioinform. **8**(3), 577–591 (2011)
7. Zheng, H., Ng, T.Y., Zhang, L., Shiu, C.K., Wang, H.Q.: Tumor classification based on non-negative matrix factorization using gene expression data. IEEE Trans. Nanobiosci. **10**(2), 86–93 (2011)
8. Chiang, J.H., Ho, S.H.: A combination of rough-based feature selection and RBF neural network for classification using gene expression data. IEEE Trans. Nanobiosci. **7**(1), 91–99 (2008)
9. Das, R., Saha, S.: Gene expression classification using a fuzzy point symmetry based PSO clustering technique. In: 2015 Second International Conference on Soft Computing and Machine Intelligence (ISCMI), Hong Kong, pp. 69–73 (2015)
10. Bontempi, G.: A blocking strategy to improve gene selection for classification of gene expression data. IEEE/ACM Trans. Comput. Biol. Bioinform. **4**(2), 293–300 (2007)
11. Young, M., Craft, D.: Pathway-informed classification system (PICS) for cancer analysis using gene expression data. Cancer Inform. **15**, 151–161 (2016)
12. Ismail, A.G., Ablahad, A.A.: Novel method for mutational disease prediction using bioinformatics techniques and backpropagation algorithm. IRACST – Eng. Sci. Technol. Int. J. (ESTIJ) (2013). ISSN: 2250-3498
13. Xu, J., Li, Y.: Discovering disease-genes by topological features in human protein–protein interaction network. Bioinformatics (2006)
14. Klasberg, S., Bitard-Feildel, T., Mallet, L.: Computational identification of novel genes: current and future perspectives, current and future perspectives. Bioinform. Biol. Insights **10**, 121–131 (2016)
15. Yan, W., Xue, W., Chen, J., Hu, G.: Biological networks for cancer candidate biomarkers discovery. Cancer Inform. **15**(S3), 1–7 (2016)
16. Ganesan, K., Lloyd, S., Sarkar, V.: Discovering related clinical concepts using large amounts of clinical notes. Biomed. Eng. Comput. Biol. **7**(S2), 27–33 (2016)

Identity-Based Security Scheme in Internet of Vehicles

Rakshanda Agarwal, Sai Satya Pranay, K. Rachana and H. Parveen Sultana

Abstract There has been a rapid growth in the number of vehicles on road, and this makes it vital to establish communication between the vehicles for safe transportation. This has given rise to importance of new industries as a part of automotive engineering such as cloud computing, IoT, network security and communications. The collaboration of these fields is what is today known as IoV (Internet of vehicles). Although a security mechanism exists across domains of Internet of things, the existing methods fail for Internet of vehicles due to their collaboration of ad hoc and intelligence systems. Through this paper we would like to propose a security model to establish secured communication between the individual vehicles and the infrastructure. In this model, we propose a mechanism to establish authentic communication between the RSUs (roadside units) and vehicles in turn storing the secret key in cloud and securing it by continuously updating.

1 Introduction

There has been a rapid growth in the number of vehicles on road, and according to Raymond James' industry report "in 2011 the number of Internet-connected devices surpassed the number of people on the planet, and it is expected to reach between 26 billion and 50 billion by 2020" [1]. This has given rise to importance of new

R. Agarwal (✉) · S. S. Pranay · K. Rachana · H. Parveen Sultana
Vellore Institute of Technology, Vellore 632014, India
e-mail: rakshuagarwal@gmail.com

S. S. Pranay
e-mail: saisatyapranay@icloud.com

K. Rachana
e-mail: rachanak96@gmail.com

H. Parveen Sultana
e-mail: hparveensultana@vit.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_51

industries as a part of automotive engineering such as cloud computing and IoT, thus increasing the M2M communication. Although the term of vehicular ad hoc networks is a decade old with all the vehicles connected under the Internet, the need for smart vehicles has given way to the concept of Internet of vehicles (IoV). IoV is an integration of vehicle's networking and vehicle's intelligence including the allied aspects such as humans, roads, environment and traffic.

Bringing all the vehicles under a common Internet involves distributed systems and efficient communication specifically designed for vehicles—"vehicular cloud." The key characteristics to be analyzed in deciding smart vehicles include space and time validity, content privacy and management. The main functions of cloud IoV include vehicular computing and information-centric networking. IoT devices are connected together to share information, because there is a high chance of hacking system or blocking services during transmission of information. The technologies used in IoT are product code of electronics (EPC) in which IoT devices are simple RFID tags, and EPC aims to support the use of RFID and to spread the RFID globally, which is used to automatically detect an object to store and retrieve information using the assistance of radio signals. The data collected by the RFID are transmitted in the form of radio waves between the sender (tag) and receiver (reader). Internet of vehicles (IoV) forms a part of Internet of things (IoT) which is used for communication between vehicles and public networks, and the characteristics of IoV include huge network scale, dynamic topological structure, limitations in the mobile phones, non-uniform distribution of nodes, complex granularities.

The major problems in this system include security, privacy, access control, interoperability and standardization, data deluge. In this paper I have majorly concentrated on the security breach aspect. The security risks of IoV are mainly connected vehicles, intelligent devices, V2X communication, service platform security, data security.

2 Literature Survey

IoT is fronting security challenges since IoT devices provide security and are linked to each other by sharing data throughout any haphazard condition. The main challenge areas in IoT include security, privacy, access control, data deluge, interoperability and standardization. Susceptible areas of attack to IoT device comprise places where there is minimum security. IoT devices do not have the capability to update; this advances the chances of attacks if any. These devices can also be attacked by weather conditions like moisture and high temperature. There are a few currently available solutions including software update of devices and review security codes in a timely manner so as reduce unauthorized access and connecting devices to the central power stations. The motive behind introduction of the IoT is to unite world with the Internet by using sensors and RFID technology so

that human life can be more relaxed and easy in spite of many issues which are affecting this technology [1].

The Internet of vehicles has storage, astuteness, communications and learning abilities to forestall the customers' purposes. A perception that can help in the advancement to the Internet of vehicles is the vehicular cloud, the correspondent of Internet cloud for vehicles which provide all the facilities obligatory by the self-directed vehicles. These resourceful distributed handling environment and communications can be provided by a newly developed network and total standard precisely aimed for vehicles, known as the vehicular cloud. This mobile cloud offers numerous important services which include dissemination, routing to content search, attack protection, spectrum sharing to self-directed vehicle applications via standard. The urban taskforce of these vehicles is embryonic from an assembly of sensor platforms to Internet of self-directed vehicles. Vehicular cloud is the central system environment which makes the development probable, and that in the future, the self-directed driving will be the foremost receiver in the cloud design. A vehicular cloud model sample has been detailed and explained; also potential design viewpoint has been discussed with emphasis on autonomous vehicle [2]. To solve the complications produced due to the growing transportation concerns, a vehicular data cloud policy has been proposed by the use of cloud computing and IoT together. Two pioneering vehicular data cloud services, an intelligent parking cloud service and a vehicular data mining cloud service in the IoT setting, are also offered reviews. This reading made assistance by offering an exclusive code proposal for the transportation information clouds in collaboration with the IoT environments, which has the capability to assimilate wide-ranging devices existing inside devices and vehicles within the road framework. IoT-based transportation information cloud units are estimated to be the pillar of forthcoming ITSs having the aim to build safe driving. Even after too many encounters, cloud computing and IoT provide marvelous occasions for invention of technology into the industry and provides sactionative arrangements to develop transportation information clouds [3].

Wicked users in IoV may deceive the whole transportations and create confusion on the road. Data falsification attack is one of the chief security problems in IoV where vehicles depend on data acknowledged from other peers/vehicles. Data falsification attack detection using hashes has been presented for improving network security and performance by becoming accustomed to contention window size to forward precise information to the neighboring vehicles on time. A clustering approach has also been presented which reduces travel period in case of traffic congestion. Performance of the proposed approach has been evaluated using numerical results attained from simulations. The proposed adaptive approach averts IoV from data falsification attacks and provides advanced throughput with minor delay. Furthermore, when exact statistics is transmitted to vehicles, they can make knowledgeable choice to decrease their waiting time when different routes are obtainable [4].

Due to IoV features of active topological configurations, non-uniform spreading of nodes, huge network scale and mobile restriction, IoV structures have to face various types of attacks like routing attacks, authentication and identification

attacks, confidentiality attacks, data authenticity attacks, availability attacks. These attacks end in several challenging necessities in security and privacy. According to this paper, the STRIDE threat model categorizes information security, threats and attacks into six foremost types: tampering with data, spoofing identity, information disclosure, repudiation, elevation of privilege and denial of service. There are various security requirements and countermeasures in IoV. The security requirements include availability, key distribution management, high mobility IoV entities, low error tolerance and private information in routing, while the countermeasures for them are threat model (STRIDE), intrusion detection system, honey spot and secure routing protocols, routing privacy guard device and important key administration [5].

A protected and dependable communication is a necessary element of any security application in IoV. The usage of a healthy security algorithm marks communications for security applications safety; it might also lessen application QoS in line for security processing delays and enlarged packet overheads. Predominantly, in high-density situations where the vehicles obtain huge number of security packets from the zone, suitable authentication of these packets cannot be assured. Consequently, critical safety packets go on unsubstantiated which results in a cryptographic loss. Two security mechanisms have been proposed which aims to diminish cryptographic loss rate. The principal tool is random transmitter security level section, whereas the subsequent one is adaptive scheme that repeatedly selects the top probable security level at the transmitter depending on the current cryptographic loss rate. The initial method prefers arbitrary level of security at every CAM transmission. The secondary suggested method will repeatedly select the finest promising security level conferring to a degree of security queue congestion recognized as the cryptographic loss ratio. Simulation outcomes demonstrate that the efficiency of the suggested mechanisms is in contrast to the static security technique suggested by the ETSI standard. It also shows that these techniques considerably develop the submission QoS in regard to packet delivery ratio and delay [6].

The novel notion of vehicular ad hoc networks (VANETs) has been altered to an innovative perception named the Internet of vehicles (IoV). In IoV, many diverse technologies, standards and services need to be assimilated. Still, heterogeneity and also the enormous quantity of vehicles intensify the necessity for data security. Internet of vehicles, with other skills, has some security susceptibilities. Vehicles work in susceptible and insecure surroundings with serious complications of security in cloud communications and vehicle to infrastructure. Internet of vehicles can become also be extremely susceptible to cyber-attacks. Certain security requirements have been proposed which need to be addressed to make IoV secure [7]. Another method known as the identity-based batch verification (IBV) scheme is one such scheme that creates VANET more secure and efficient. Preserving confidentiality through anonymity and lessening of authentication time of messages by confirming them in batch are the main objectives of this system. This paper points out the security problems of the existing IBV scheme and brings together the concept of the random change of anonymous identity with time and location, to

prevent the security attack and to preserve the privacy. Here, performances are calculated in terms of delay and transmission overhead [8].

The PKI-based validation protocols are intensely suggested for solving security issues in VANETs. However, they have following inadequacies: lengthy certificates lead to communication and calculation overheads, and the absence of privacy preservation due to revealing of vehicle individuality, transferred in broadcasting safety message. Symmetric cryptography-based protocols are quicker because of a solitary secret key and straightforwardness; nevertheless, it does not guarantee non-repudiation. An efficient, scalable and privacy-preserving authentication (ESPA) protocol has been presented for protected VANETs. This protocol employs cross-cryptography. In ESPA, the symmetric hash message authentication code (HMAC)-based authentication and the asymmetric PKI-based pre-authentication are implemented throughout vehicle-to-vehicle (V2V) communications and vehicle to infrastructure (V2I), respectively. Widespread simulations are accompanied to authenticate proposed ESPA protocol, and it is also compared with the present work based on PKI and HMAC. The performance analysis displayed that ESPA is more efficient, scalable and privacy-preserving secured protocol than the existing protocols. The complete network is distributed into some domains, which are locally monitored by ATAs. The protocol comprises of two phases, phase I (V2I), where vehicles pre-authenticate with RSUs, and phase II (V2V), where the vehicle authenticates with other vehicles in VANET. PKI system has been used for pre-authenticating vehicles by RSUs during V2I communication. HMAC is used for vehicle-to-vehicle verification during V2V communication [9].

Wireless communication networks are also exposed to cyber-attacks such as denial of service (DoS). Thus, a vehicular control scheme which is robust to such cyber-attacks is needed to improve the safety and security. In this paper, a resilient control scheme has been developed for a section of connected vehicles well appointed with cooperative adaptive cruise control (CACC) to lessen the DoS attack. The conventional CACC algorithm has been modified here by adding an approximation algorithm consisting of a set of Luenberger's observer and a delay estimator. The suggested estimation algorithm consists of three main components: a Luenberger observer for no DoS case, a model-based observer and delay estimator for the under-attack perception. This estimation algorithm is proficient of perceiving DoS attack in DSRC communication network as well as approximating states of earlier vehicle. These probable states are used in the modified CACC to improve the performance under the attack [10].

Inter-vehicle communication schemes are an innovative hypothesis of networking. It is greatly associated with mobile ad hoc networks along with their diverse, self-organizing assembly, also being responsible for the introduction of new threats. To evaluate these threats, a new model has been introduced in [11]. It is based on the attacks on the inter-vehicle communication system. The model has been recycled to enhance the system model of the NoW communication system and also to the discovery of possible flaws throughout the description phase of the NoW communication system. Safety-associated applications need a protected, trustworthy system. Hence, here a summary is presented on the numerous possible outbreaks and

countermeasures which need to be considered intensively. According to this paper, attack trees offer a beneficial instrument to measure the security of a system progressively. The top-down method influences the organization design at an initial development phase concerning security contemplations, while also having the capability to create a more meticulous study immediately after the system's stipulations be more exact. Two procedures can augment complete security, undertaking indigenous credibility checks in regular system and car checks on the nodes, particularly RSUs. Plausibility checks include assessment of the established statistics to internal sensor data, assessing messages from diverse information foundations regarding an individual event and scenario building; here in individual traffic events are linked by means of statistics. This can greatly increase the strength of the attacker, but will require an appropriate model for each application. Secondly, consistent system checks would authenticate the appropriate function of a unit and consequently diminish the numeral of malfunctioning units. This also includes the possibility of updating the software [11].

A roadside unit (RSU) cloud, which is a vehicular cloud, acts as the operative pillar of the vehicle grid in IoV with the incorporation of sensors and microcontrollers in vehicles and a permanent roadside arrangement to practice an intellectual vehicle grid and also to accommodatingly proliferate road safety and traffic flow. Here, a unique vehicular cloud design known as RSU cloud has been proposed. It comprises of traditional RSUs along with specific microscale data centers. The uniqueness of this RSU cloud design lays in promoting from the elasticity and deep programmability presented in software-defined networking (SDN). A SDN comprises of two communication planes, an abstracted control plane and the physical data plane. This combination of control and data planes enables the elasticity and programmability of the SDN. The Android application collects all information from server through web server about the occupancy and vacancy of vehicle space and then determines overall numeral of slots, occupied and vacant slots. An IoV has been planned and developed and also developed an incorporation of sensors and microcontrollers in vehicles along with a static roadside infrastructure to produce an intelligent vehicle grid to compliantly proliferate road safety and traffic flow [12].

3 Proposed Methodology

a. RSU: It is an embedded system that consists of sensors to identify vehicles and mapped to the secret key.
b. OBU: It is an embedded system chip that is fixed in the vehicles to enable discoverability to the RSUs and help communicate the vehicle details to the RSU.
c. Secret key: It is the unique key generated to map the vehicle ID to the cloud to provide privacy.
d. Cloud technology: It is used to store the details and activities of the vehicles and gets updated.

The junction of the roads is embedded with RSUs. As the vehicles pass by, a communication is established between vehicles' OBU and RSUs. A unique key also known as anonymous ID is generated based on vehicle identification and stored in the cloud. In the existing methods, an anonymous ID is created for each vehicle which represents the vehicle in the network. This anonymous ID is used to identify and track the vehicle using RSUs. This model is vulnerable to several attacks like traceability attack. In this model, once if an anonymous ID is created and mapped to a particular vehicle, ID won't change till the end of the session. Due to this constant nature of ID throughout the session, the attacker can easily get the data about any vehicle in that network. This might cause privacy breach of a particular vehicle. To close this loop hole, we have proposed a model in which the anonymous ID will get changed periodically when the vehicle travels from one RSU to another. When the vehicle moves out of the range of one RSU as shown in Figs. 1 and 2, the existing anonymous ID will be replaced by the newly generated anonymous id. This periodic change of anonymous IDs helps to decrease the vulnerabilities and makes it harder for the attacker to trace a vehicle based on their anonymous IDs.

Once the vehicle moves out of RSUs range, it is mapped to the next RSU and the new secret key is generated which is then mapped to the cloud. Since the key is unique for a vehicle, the previously stored key is replaced by the new secret key. This facilitates security since the vehicles' identification is constantly mapped to a new secret key soon after generating it. Initially, the car is in range B, as indicated in Fig. 1. Hence, a communication is established between the OBC of the car and RSU 2. A secret key is generated and stored in the cloud. Once the car moves out of zone B and reaches range A as indicated in Fig. 2, a new communication range is detected. Hence, a communication is established between OBC and RSU 1. The new secret key generated replaces the old one in cloud. In this manner newly generated keys are unique and replaced the old keys providing security.

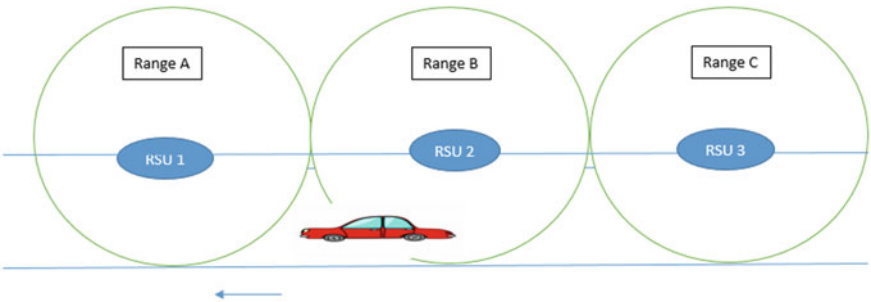


Fig. 1 Vehicle connected to RSU 2 when in range B

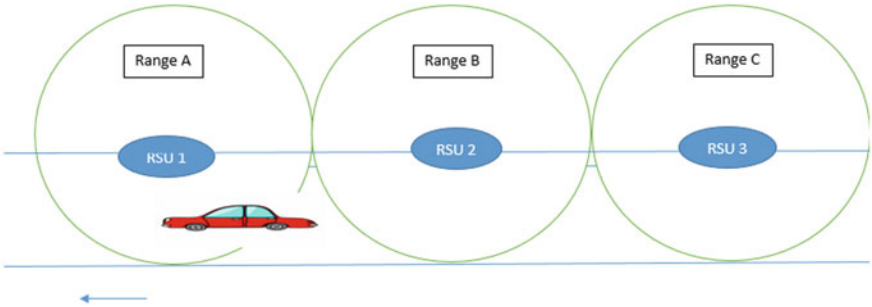


Fig. 2 Vehicle connected to RSU 1 when in range A

4 Conclusion

This security model facilitates secured communication in V2M connection. VANETs are a special class of ad hoc networks. While establishing V2I connection, the concept of MANET also needs to be taken care of. In this context the use of RSU which is in continuous communication with the Internet is used. The proposed model provides security by continuously updating the anonymous ID and mapping to the vehicle when it moves from one RSU to another, thereby providing a secured communication in an ad hoc environment for the vehicles. Our scheme makes it harder for the attacker to trace a vehicle because of the dynamic allocation of the anonymous IDs with respect to time. Hence, the proposed model improves the security and protects the private data of the connected vehicles.

5 Future Work

As a future work, encryption algorithms can be applied on the secret keys to provide higher level of security during key exchange. Vehicle authentication, i.e., verification of the anonymous IDs, can be done through a secured transmission channel using SSL. An improvement can be done by storing the secret key as a hashed value in the cloud for faster fetch and for faster change of anonymous IDs.

References

1. Parashar, A., Rishishwar, S.: Security challenges in IoT. In: 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), pp. 446–449. IEEE (2017)

2. Gerla, M., Lee, E.-K., Pau, G., Lee, U.: Internet of vehicles: from intelligent grid to autonomous cars and vehicular clouds. In: 2014 IEEE World Forum on Internet of Things (WF-IoT), pp. 241–246. IEEE (2014)
3. Ashokkumar, K., Sam, B., Arshadprabhu, R.: Cloud based intelligent transport system. Procedia Comput. Sci. **50**, 58–63 (2015)
4. Rawat, D.B., Garuba, M., Chen, L., Yang, Q.: On the security of information dissemination in the Internet-of-Vehicles. Tsinghua Sci. Technol. **22**(4), 437–445 (2017)
5. Sun, Y., Wu, L., Wu, S., Li, S., Zhang, T., Zhang, L., Xu, J., Xiong, Y.: Security and privacy in the internet of vehicles. In: 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI), pp. 116–121. IEEE (2015)
6. Javed, M.A., Hamida, E.B.: Adaptive security mechanisms for safety applications in internet of vehicles. In: 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 1–6. IEEE (2016)
7. Contreras, J., Zeadally, S., Guerrero-Ibanez, J.A.: Internet of vehicles: architecture, protocols, and security. IEEE Internet Things J. (2017)
8. Mahapatra, P., Naveena, A.: Enhancing identity based batch verification scheme for security and privacy in VANET. In: 2017 IEEE 7th International Advance Computing Conference (IACC), pp. 391–396. IEEE (2017)
9. Tangade, S., Manvi, S.S.: Scalable and privacy-preserving authentication protocol for secure vehicular communications. In: 2016 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), pp. 1–6. IEEE (2016)
10. Biron, Z.A., Dey, S., Pisu, P.: Resilient control strategy under denial of service in connected vehicles. In: American Control Conference (ACC 2017). IEEE (2017)
11. Aijaz, A., Bochow, B., Dötzer, F., Festag, A., Gerlach, M., Kroh, R., Leinmüller, T.: Attacks on inter vehicle communication systems-an analysis. In: Proceedings of the WIT (2006), pp. 189–194 (2006)
12. Salahuddin, M.A., Al-Fuqaha, A., Guizani, M.: Software-defined networking for RSU clouds in support of the internet of vehicles. IEEE Internet Things J. **2**(2), 133–144 (2015)

IoT-Based Smartbots for Smart City Using MCC and Big Data

Praveen Kumar Singh, Rajesh Kumar Verma and P. E. S. N. Krishna Prasad

Abstract To build a smart city that connects various instruments such as cars, refrigerators, sensors, actuators, smart lights, smart mobile devices is useful to build a smart and intelligent city. In order to have the real-time systems to capture huge amount data of the various IoT services running in parallel, the complex data gathering and subsequent processing can be a very challenging task. In this paper, we proposed a Smartbot which gathers huge amount of data and process it to provide intelligent solution to the user, using MCC and Big Data technology. The architecture uses Hadoop and Spark ecosystem for the real-time simulation and presents the realization of Smartbots which are integrated with the IoT devices, smart data processing with analytics, networking with cloud, and other data visualization systems. The novelty of this paper lies in the integration of Smartbots in the IoT environment by making use of MCC and Big Data. Our proposed architecture has been developed considering the smart city as the base and all enhancements are built on the top of this base framework.

1 Introduction

The current trend of living is changing towards making cities smarter than yesterday, by adding infrastructure and sensors which can alert us about any changes in the current setup and take appropriate steps. Users needed good services and infrastructures on demand through the smart devices. Reference [1] reports Cisco estimating connected devices will be 50 Billion and connected devices per person will be 6.58

P. K. Singh
Bank of America, Mumbai, India
e-mail: praveenhelp78@gmail.com

R. K. Verma
Biju Patnaik University of Technology, Rourkela, Odisha, India
e-mail: rajeshverma_chicago2004@yahoo.com

P. E. S. N. Krishna Prasad (✉)
Prasad V. Potluri Siddharta Institute of Technology, Vijayawada, India
e-mail: Surya125@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_52

by 2020. Urbanization is the factor where most of the private and government organization are keen for digitally connected world with advancements. A report by Evans Data Corporation (EDC) more than 50% of IoT developers are targeting implementation of connected home, transportation, etc [2] Many Organization are very keen to use IoT platform for live survey, live monitoring, live quality measures (such as water quality measurement, vehicle monitoring system, pollution, humidity), mapping with population, gathering all crucial parameters from various perspective through various IoT devices for urban development. The IoT devices such as sensors (light, temperature, humidity, motions, gestures, etc) are deployed at various locations for various purposes and connected across the network, resulting a huge amount of heterogenous data is being generated. According to EDC [2] most of the IoT developers are primarily focussed on Big Data, mobility, and cloud for providing services using IoT devices. Available platforms from Microsoft Azure IoT, Intel IoT, IBM IoT, Cisco IoT, etc, for business purposes to provide services and infrastructures in quick fashion.

This paper provides architectures to analyze Big Data on the basis of user needs using Big Data framework that enables the various user can use different applications using the smart mobile devices to interact with the real-time smart city data. Cloud computing is used to get highly available resources to collect smart city data present across the network from various IoT devices as a service.

The rest of the paper is organized as follows. In next section, we will discuss the related work, the IoT-based system architecture using MCC and Big Data, smart data processing, and analytics, proposed unified Smartbot architecture using MCC and Big Data and impact of data analytics on smart city.

2 Related Work

Author [3] proposed a IoT-based framework for the realization of smart city using data management, networking, and cloud-based integration. References [4] and [5] proposed approach for providing cognitive assistance for smart home and provides assistance for persons suffering from cognitive deficits in smart home, respectively. Reference [6] has done lot more survey on smart city and proposed heat and electricity management system application for smart city.

Author [7] given the prototype system for smart parking management system (SPARk) based on wireless sensor network for parking monitoring as well as for parking reservation. MCC is used for computing resources by SMDs integrated with the cloud [8]. Reference [9] proposes the IoT-based urban planning system using Big Data analytics which consists of various IoT-based smart system including smart home, smart parking, weather, and water but limited to the Hadoop ecosystem. Our proposed approach is closed to this work but additional support for real-time processing using Spark with rich resources availability for computation and processing using cloud computing.

3 IoT-Based System Architecture using MCC and Big Data

3.1 Architectural Components Used

The architecture is mainly used for developing smart system which is connected to all the available components and resources. Figure 1 depicts the main components which are used in different layers.

- Data Connected Layer: Used for collecting the data generated from various IoT devices to the one place. It is very important to get correct information from right devices to take decision which is useful for smart city planning.
- Network Connected Layer: are various services that are running on smart system to provide gateway and all API management system. Various available IoT devices use different API and to provide generic type of API management system to complete certain tasks.
- Smart Data Processing and Analytic Layer: helps to streamline the various data of various data forms through the smart data preprocessing, smart data processing, and smart data analytics. This is required for intelligent decision making in real-time for smart city data.
- Smart Data Visualization System: helps to visualize the real-time smart city data, getting pattern from the smart city data, understand the trend getting from visualization system and to get some good insights from the smart visualization system. Further, helps to make better decision-making system for smart city.
- Applications (Apps): Traffic monitoring applications, water applications, pollution application, smart health monitoring application, smart city application, smart home application, surveillance application, simulation-based applications, safety application, disaster detection and recovery application, object captures, smart parking applications, smart transportation applications, urban planning applications, and image classification applications come into this category. User uses these applications as a package with proper notification and through the real-time environments.

The sequence of processing a different IoT devices data is as follows

Sensors are deployed on different places in order to get the real-time smart city data. Different places are having different application in order to classify the data comes from which applications such as traffic monitoring, water monitoring, pollution monitoring, health monitoring, and many more applications. Challenge is to collect all generated data into one place to make better decision making for smart city.

The generated data comes to one place which we call it as IoT-based smart system. System is of distributed and parallel in nature so that Big Data frameworks can work on it along with that disaster recovery factor can be taken into consideration.

The API management system/gateway is responsible for accessing and controlling the RESTful Web APIs for IoT devices. Many existing open-source solutions are present to create easily APIs for IoT devices. These IoT devices considered to be

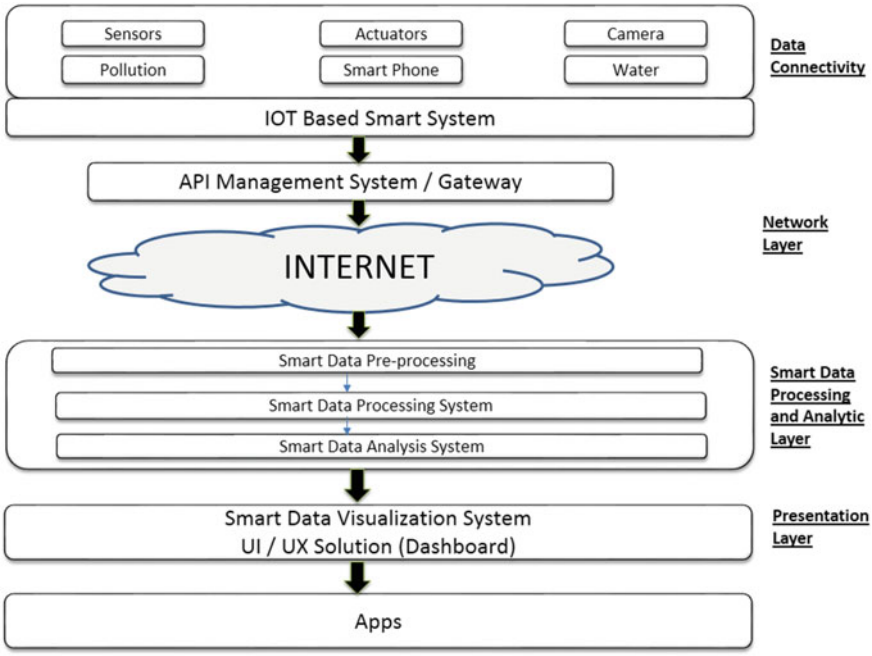


Fig. 1 IoT-based system architecture

connected with Chatbots through the available interfaces. Data generated through the IoT devices are of heterogenous type including lot of metadata. Captured data can be of structured, semi-structured, and unstructured which may have redundancy as well as not required information, results slow down the smart system and creating extra overhead computation. Therefore, we are preprocessing the data in such a way that filtering algorithm applied just after data collection and attribute reduction algorithm [10, 11] applied to remove the redundant attribute. Resultant classified data is in the organized form which can be stored in any supported file format. Further classification required based on our built-in XML identifier for identifying which Big Data processing framework is useful for analysis.

Huge amount of data is being collected and need to be processed, which requires mobile cloud computing as well as Big Data ecosystem. Big Data ecosystem uses the most popular framework named as Hadoop [12] and Spark [13]. These frameworks required distributed filesystem such as HDFS, S3. HDFS divides the data into number of blocks and each block is of same size, default block size is already defined in the framework and can be customized based on the requirements. These frameworks follow typical master–slave architecture, where master is known as NameNode and slave is known as DataNode. DataNodes are the nodes where data resides in an distributed manner and Namenode contains the metadata of DataNodes. Parallel

processing is performed on each DataNode using MapReduce programming model. Spark preferred over Hadoop in case of real-time analytics, in-memory computation, and machine learning based analytics.

Finally, the decision making and result generation are done at the ecosystem itself and appropriate data need to be displayed across all the connected smart city applications through the smart visualization system.

4 Smart Data Processing and Analytics

The different components that are used for any applications of smart city using MCC and Big Data are depicted in Fig. 2. The different components and their functionalities are detailed as follows

- Data Sources: This component hosts the various IoT-based applications used in smart city that require data processing, data analytics, and data visualization. Several smart city applications such as traffic applications, water applications,

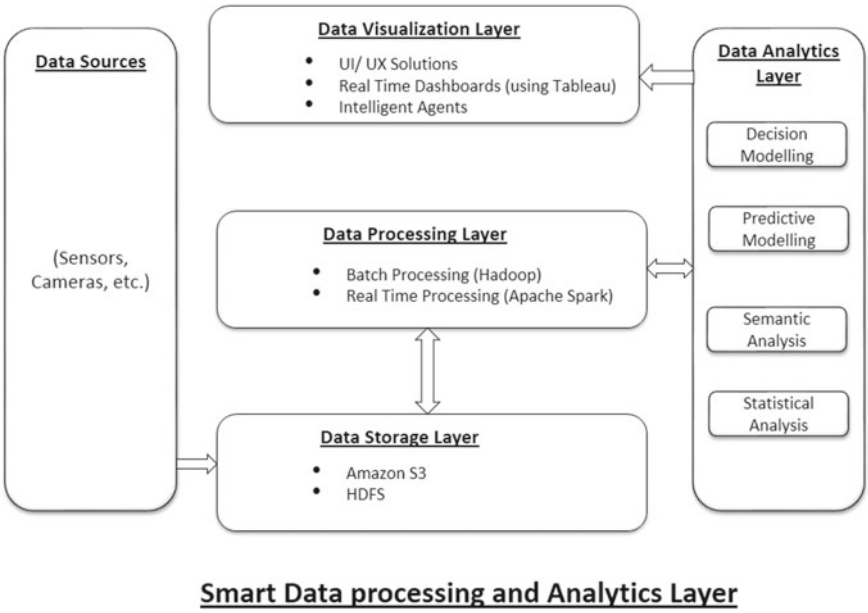


Fig. 2 Smart data processing and analytics

pollution application, smart home application, surveillance application, simulation-based applications, safety application, object captures, smart parking applications, Urban planning applications, image classification, and detection-based applications getting data into the various IoT devices, for example, sensors, camera, lights.

- Data Storage Layer: This layer is used for various data storage which supports distributed file system such as HDFS, s3, and local file system as well.

- Data Processing Layer: This layer is useful for batch processing, real-time processing, and hybrid processing using various available Big Data frameworks such as Hadoop, Spark, Twister, Flink.

- Data Analytic Layer: This layer contains various analytics and modeling used for analysis and derives hidden meaningful information from smart city data. Decision modeling, predictive modeling, semantic analysis, and statistical analysis are used in this layer to perform the computation.

- Data Visualization Layer: This layer is responsible for decision making and performing an action. User Interface (UI/UX) provide to the user to get real-time solution using various dashboards and Intelligent Agents. Tableau, QlikView, and many more tools can be used for visualization.

Smart city data is collected from various IoT devices. Data processing, analysis, and visualization can be done in the Hadoop cluster as well as in the cloud using visualization server connectivity. Big Data frameworks are responsible for Offline, Batch, hybrid and real-time processing based on the data classification. The subsequently transmitted results from processing can be transmitted to the user devices.

5 Proposed Unified Smartbot Architecture Using MCC and Big Data

The Fig. 3 depicts the Generic Bot architecture for MCC and Big Data, which contains the four modules as described below.

- User/bot Layer: The physical user interacts with the Smartbot to clarify all the queries in his mind regarding the different facilities, location, etc, in the smart city. Subsequently, the bot layer interacts with the cloud and network layer through the gateway using available API's to pass the queries (input parameters) and also gathers the results which are pass back the users.

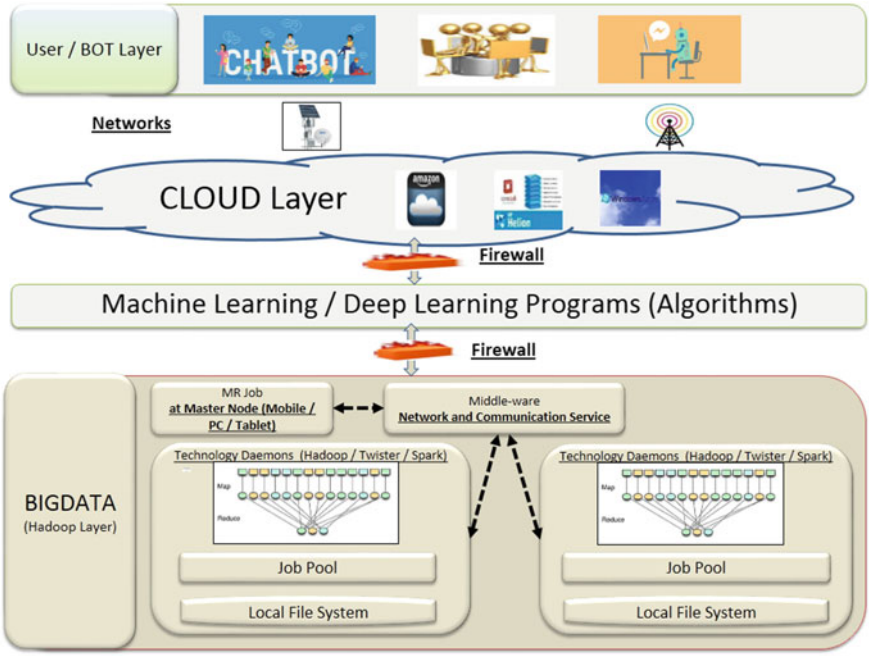


Fig. 3 Generic bot architecture using Big Data

- Cloud Layer: The cloud layer has immense resources which are used for fast data processing and storing the large amount of data.
- Machine Learning: The data gathered is processed via the intelligent algorithm of machine learning and deep learning.
- Big Data Layer: The Big Data layer is used to store huge amount of data coming from various IoT devices. The MapReduce and Spark programs help to process this huge amount of data and provide significant insight and knowledge to the user.

6 Impact of Data Analytics on Smart City

The proposed architecture having solid integration among all the smart systems which are connected to the IoT-based smart system in Fig. 1 and real-time data gathers at centralized place in a distributed manner. Analytic engine such as Hadoop and Spark is responsible for data insights.

6.1 Scenario 1 (A Case Study)

For smart city water quality measurement system: We have considered different parameters responsible for water quality across all the locations with their acceptable range along with water temperature. Parameters are chosen with the help of domain expert system and availability of sensors. Parameters are of independent in nature and depend on the environment changes. Deployed sensors at various locations are sending data to the server at data connectivity module in Fig. 1 through the network. The sample data gathered is converted to appropriate format, and subsequently passed to the big data ecosystem after removing any outliers that may be present in the data. This refined data is then classified and different categories of alarms are generated, based on the input categories that are fed into the reference XML file. Generated alarm sends to the end users, and it is being saved as a triggered event in the smart visualization system. Useful trends available to the end users as well as to the administrator such as top ten locations affected through the water quality, user's feedback demonstration, water quality measurement of entire city, generated triggered events, parameters values across various locations meet the respective standards, pH, and total dissolved solids are positive deviated or negative deviated, comparison with weighted arithmetic index and NSF water quality index.

6.2 Scenario 2 (A Case Study)

For the smart city vehicle management and monitoring system: Various sensors are deployed at various locations such as parking, stops, building main gate along with cameras. These sensors and cameras are used to generate real-time data, the considered parameters are vehicleId, time, speed, locations, and many more, parameters are of dependent in nature. Real time data is got from the sensors and is sent to the data connectivity module via the network. Subsequently, preprocessing is required and then the refined data is sent to the Big Data frameworks for analytical processing. The real-time data displayed to the users to aware about the real-time traffic, the visualization system shows the number of vehicles and the average speed of the vehicles on the road which is integrated with the GPS. The smart system also shows the parking slots are available or not the particular location, depending on that particular user can take decision for the parking zone. These smart system can be utilized by the government as well as by the private organization to take run-time decision for traffic control as well as for commercial vehicle availability to the users at particular location. Useful trends can be displayed to the users as well as to the admin, the trends can be topmost traffic path, most filled parking zone, traffic analysis of entire city.

7 Conclusions and Future Work

The integration of IoT devices, MCC, and Big Data helps to create applications for the smart city. All applications are provided as a package to the user. The intervention of cutting-edge technologies such as Hadoop, Apache Spark is used in a big manner. User can smartly visualize the required and appropriate information as per their need in order to make intelligent decisions and act upon them in real-time using mobile devices. The architecture proposed by the author helps to integrate various smart systems using Big Data framework to address issues related to processing and scalability in order to get real-time data visualization for taking quick decisions. This architecture can be easily validated with the requirements for any particular smart city facility, implemented using Hadoop and Spark and tested before the live implementation takes place. Typical examples are the smart city water quality measurement system and the smart city vehicle management and monitoring system.

In future, we are planning to implement full-fledged package with actual integrated deployment of all the smart city applications with the consideration of quality of service (QoS) and service-level agreement (SLA).

References

1. Evans, D.: The internet of things: how the next evolution of the internet is changing everything. In: White Paper, The Internet of Things, pp. 346–360 (2011)
2. E.D. Corporation. Internet of things - vertical research service. http://www.evansdata.com/reports/viewRelease.php?reportID=38 (2016)
3. Jin, J., Gubbi, J., Marusic, S., Palaniswami, M.: An information framework for creating a smart city through internet of things. IEEE Internet Things J. **1**(2), 112–121 (2014)
4. Giroux, S., Bauchet, J., Pigot, H., Lussier-Desrochers, D., Lachappelle, Y. Pervasive behavior tracking for cognitive assistance. In: Proceedings of the of PETRA08, pp. 1–7. ACM, New York, USA (2008)
5. Bouchard, B., Giroux, S., Bouzouane, A.: A smart home agent for plan recognition of cognitively-impaired patients. J. Comput. **1**(5), 53–62 (2006)
6. Kyriazis, D., Varvarigou, T.A., White, D., Rossi, A., Cooper, J.: Sustainable smart city IoT applications: heat and electricity management and eco-conscious cruise control for public transportation. In: WOWMOM, pp. 1–5. IEEE Computer Society (2013)
7. Srikanth, S.V., Pramod, P.J., Dileep, K.P., Tapas, S., Patil, M.U., Babu, N.S.C.: Design and implementation of a prototype smart parking (spark) system using wireless sensor networks. In: AINA Workshops, pp. 401–406. IEEE Computer Society (2009)
8. Hoang, D.T., Lee, C., Niyato, D., Wang, P.: A survey of mobile cloud computing: architecture, applications, and approaches. Wirel. Commun. Mob. Comput. **13**(18), 1587–1611 (2013)
9. Rathore, M.M., Ahmad, A., Paul, A., Rho, S.: Urban planning and building smart cities based on the internet of things using big data analytics. Comput. Netw. **101**, 63–80 (2016)

10. Sai Prasad, P.S.V.S., Subrahmanyam, H.B., Singh, P.K.: Scalable iqra_ig algorithm: an iterative mapreduce approach for reduct computation. In: Krishnan, P., Krishna, P.R., Parida, L. (eds.), ICDCIT. Lecture Notes in Computer Science, vol. 10109, pp. 58–69. Springer, Berlin (2017)
11. Singh, P.K., Sai Prasad, P.S.V. S.: Scalable quick reduct algorithm: Iterative mapreduce approach. In: Marathe, M., Mohania, M.K., Mausam, Jain, P., (eds.), CODS, pp. 25:1–25:2. ACM (2016)
12. Apache hadoop. http://hadoop.apache.org/
13. Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., Stoica, I.: Spark: cluster computing with working sets. In: Nahum, E.M., Xu, D. (eds.), HotCloud. USENIX Association (2010)

Generation of Random Fields for Image Segmentation Based on MRF Energy Level Set Method

Rambabu Pemula and C. Nagaraju

Abstract This research paper focuses on image segmentation by random fields generated by using MRF energy level set function. The focus of the researcher is on this technique of how MRF energy is embedded in the standard level set energy method. The connection among all the pixels and its neighborhood is established by MRF energy function, and it attempts to have all the pixels in the same region. Comparing the MRF level set technique with surviving manifold learning technique on different standard images. In comparison, the proposed technique is fast and healthy for image segmentation tough the statistical tables present the researcher the superiority of the proposed technique.

1 Introduction

Image segmentation plays a vital role in object recognition, and computer vision applications and its main aims are to separate the given image into object and background. Image segmentation is a significant problem due to a variety of distressing factors like noise, shadow, clustered data, missing parts in the literature, and several techniques are proposed to deal with this problem in the perspective of manifold learning techniques, Markov random fields (MRF), entropy techniques, different techniques of fuzzy, etc. The vital complexity of the image segmentation aim is to present a speedy and robust process for effectively even for the under low-quality images.

R. Pemula (✉)
Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Kakinada, Kakinada, Andhra Pradesh, India
e-mail: rpemula@gmail.com

C. Nagaraju
Department of Computer Science and Engineering, YSR Engineering College of YV University, Proddatur, Andhra Pradesh, India
e-mail: nagaraju.c@yogivemanauniversity.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_53

A new technique shape prior manifold was proposed by Etyngier et al. [1], which will handle a lot of universal shape priors and prototypical a class of shapes as a finite dimensional manifold. But, it does not give an exact prediction operator on the manifold. Later, [2] Etyngier et al. proposed a model for deformable model framework based on nonlinear shaper prior by using the diffusion maps, and it is called as shape prior manifold. An outline of manifold linear learning and applications of manifold to medical image examination was given by Aljabar Paul et al. [3]. Duc et al. [4] introduced a novel method for medical image segmentation using multi-atlas segmentation using manifold technique. It provides improved medical image segmentation results. A random geometric prior forest system was introduced by Xiao et al. [5] to get object-adaptive geometric priors with efficiency and robustly. During this process, manifold preserving grouping approach is used to rebuild the test geometries by storing the feature space manifold. Rambabu et al. [6] introduced a new method for image segmentation by generation of random fields with the help of manifold learning technique, and it provides improved image segmentation in comparison with occlusion technique and entropy division. Applications of manifold learning techniques were discussed in [7–10]. Using the hierarchical Markov random field models, Barker et al. [11] proposed two unsupervised segmentation techniques for separating the textured and noisy images. Each technique finds the foremost possible range of classes and their measures attached and fosters an equivalent split of the images hooked on these classes. MRF and filters, random and maximum entropy model (FRAME) [12] were used for unsupervised segmentation of images that contain multiple textures. This approach can segment the images of complex textures effectively and is strong to noise. Image segmentation using generation of random fields using different techniques is deliberated in [13–17]. A new 3D tumor segmentation technique was proposed by Taheri et al. [18], based on the global threshold-based scheme that uses level set speed function. A vital region-based segmentation technique based on level set method was introduced by Li et al. [19], and it deals with the depth in homogeneities in the segmentation. This technique is better than the piecewise smooth model in terms of robust to initialization, faster, and accurate. This technique is more useful for segmentation of magnetic resonance images. A novel level set technique for segmentation for the image was introduced by Yang et al. [20], which is fast and strong than the other segmentation techniques. Here, MRF energy function is entrenched with the conventional level set energy function. Pemula et al. [21] projected a new technique for image segmentation with fuzzy manifold learning technique that yields improvised segmentation results.

2 Existing Method

A weighted adjacent neighboring algorithm with symbolic features method for learning that makes usage of the adjacent neighbor's distance and the weights of nearest neighbors decided by the searching point. The approach based on distance

ignored the structure data of the input space looked by different adjacent neighbors. This makes performance poor. To overcome this problem, locally linear embedding (LLE) [6], [25] algorithm is used for the reconstruction of the test geometries protecting the feature space of manifold structure. Image set $\{I_1, \ldots., I_N\}$ defined by N images $I_i \in \mathrm{R}^D$, each being defined as an intensity vector, where D is the image pixels. Assuming that $\{I_1, \ldots., I_N\}$ lie on or near the dimensional d manifold $\mathcal{M}$ embedded in R^D and $d \ll D$, low-level dimensional representation $\{y_1, \ldots., y_N\}$ with $y_i \in \mathrm{R}^d$ of input images. The local neighborhood of the high-dimensional data in the low-dimensional space was safeguard by using manifold for low-dimensional built of the locally linear embedding (LLE). This presumes a locally linear bond among the adjoining data points. The notion is to each data point I_i as weighted combination of k, which is the nearby neighbors in high-dimensional space. It demonstrates the group of weights w_{ij} for the neighbors k of I_i. Its goal is toward identifying a low-dimensional illustration y_i that admirations this. The objective method of LLE is defined in Eq. (1) as

$$\varnothing(Y) = \sum_i \|y_i - \sum_{j=1}^{k} w_{ij} y_{ij}\|^2 \; subject\; to \; \left\|y^{(k)}\right\|^2 = 1 \tag{1}$$

W is the weight matrix, the embedding is attained form the eigenvectors d, equivalent the smallest nonzero eigenvalues of $(I-W)^T(I-W)$. In LLE, least constructed cost used to identify the local coordinates then maps these coordinates' maps to the global one. Experiments proved that LLE is better than the other manifold learning techniques. But LLE algorithm has some limitations like out-of-sample problem and neglects the class information problem.

3 Proposed Method

In this experiment, the researcher introduced a new technique for image segmentation by using generating random fields with the help of MRF energy level set function. In this context, MRF energy method is explained in Eq. (2) with the help of posteriori object segmentation energy $Eng(Ob|In)$ in addition posteriori background segmentation energy $Eng(Bg|In)$ as

$$E_{MRF} = Eng(Ob|In)U(\varnothing) + Eng(Bg|In)(1 - S(\varnothing)) \tag{2}$$

where unit step function $S(\varnothing)$, and it is represented in Eq. (3) that frequently used rather in a smoothened form in application, i.e.,

$$S_\varepsilon(\varnothing) = \begin{cases} 1, & \varnothing > \varepsilon \\ 0, & \varnothing < -\varepsilon \\ \frac{1}{2}\left(1 + \frac{\varnothing}{\varepsilon} + \frac{1}{\pi}\sin\left(\frac{\pi\varnothing}{\varepsilon}\right)\right) & |\varnothing| \ll \varepsilon, \end{cases} \tag{3}$$

ε is a constant and its value is 1.5. Segmentation posteriori probability energy method can be represented in Eq. (4) as

$$\begin{aligned} &Eng(T|In) = -\log Pr(T|In) \approx -\log\left(Pr(T)Pr(In|T)\right) \\ &\approx \sum_{f\in F}\left(\frac{\left(In_s-\mu T_f\right)^2}{2\sigma_{T_f}^2} + \log\left(\sqrt{2\pi}K\sigma T_f\right) + \sum_{c\varepsilon\mathbb{N}^{\mathcal{W}m}} V_c\left(T_f\right)\right) \end{aligned} \tag{4}$$

$Pr(T)$ is the segmentation priori probability is associated with its region $\mathbb{N}^{\mathcal{W}m}(f)$ based on the property of Markovian. Then, $Pr(T)$ can be represented in Eq. (5) as

$$Pr(T) = \prod_{f\in F}\frac{1}{K}exp\left(-\sum_{c\in\mathbb{N}^{\mathcal{W}}(f)} V_c\left(T_f\right)\right) \tag{5}$$

where c is the set of all the cliques, K is a constant, and $V_c(T_s)$ is the clique energy method, it is represented in Eq. (6) as

$$V_c\left(T_f\right) = \begin{cases} -\gamma\left(e\left(-\left|In_f - In_{n(f)}\right|\right)\right), & T_f = T_{n(f)} \\ +\gamma\left(\frac{2}{1+e\left(-\left|In_f - In_{n(f)}\right|\right)} - 1\right), & T_f \neq T_{n(f)} \end{cases} \tag{6}$$

where $n(f)\in\mathbb{N}^{\mathcal{W}m}(f)$, $\left|In_f - In_{n(f)}\right|$ is the intensity absolute difference between center pixel and one of its neighbors, and γ is constant. MRF energy function used to maximize the posterior segmentation, and it identifies the segmentation label for each and every pixel with its neighborhood information as in Eq. (7)

$$Pr(T|In) = \frac{Pr(T|In)Pr(T)}{Pr(In)} \tag{7}$$

Because, here, $Pr(In)$ is a constant, the segmentation probability $Pr(T|In)$ is proportionate to the segmentation prior probability $Pr(T)$ and the segmentation conditional probability $Pr(In|T)$ is given in Eq. (8).

$$Pr(T|In) \propto Pr(T)Pr(In|T) \tag{8}$$

Generally, pixels in the background and object regions in image are independent. It is the elementary for the all level set functions. But it results in the segmentation scattered by noise. So, we constructed the $Pr(In|T)$ and $Pr(T)$ using the MRF theory, which considers the neighboring relationship and defines $Pr(In|T)$. $Pr(T)$ of each pixel is relied on its neighborhood $\mathbb{N}^{\mathcal{W}m}(f)$. The MRF energy method builds a $(2\mathcal{W}m+1)\times(2\mathcal{W}m+1)$ neighborhood structure $\mathbb{N}^{\mathcal{W}m}(f)$, where sites $f\varepsilon F$ refer to individual variables of type random elements. The eight neighborhoods of f_x^y are well-defined in Eq. (9) as

$$\mathbb{N}^1(f) = \left\{ f_{x-1}^{y-1}, f_x^{y-1}, f_{x+1}^{y-1}, f_{x-1}^{y}, f_{x+1}^{y}, f_{x-1}^{y+1}, f_x^{y+1}, f_{x+1}^{y+1} \right\} \tag{9}$$

A combined random field is represented by $M = \{In, T\}$ where $In = \{In_f, f \in F\}$ is the image intensities field, and $T = \{T_f, f \in F\}$ is the object (Ob) from background (Bg) separated field label. In the segmentation conditional probability $Pr(In|T)$ and Gaussian distribution are followed by the pixels in the object as well as background. Thus, with various values of means μ_{T_f} and standard deviations σ_{T_f}, the method of $Pr(In|T)$ is given in Eq. (10) as

$$Pr(In|T) = \prod_{s \in F} \frac{1}{\sqrt{2\pi}\sigma T_f} e \left(- \frac{\left(In_f - \mu_{T_f}\right)^2}{2\sigma_{T_f}^2} \right) \tag{10}$$

where μ_{T_f} mean and σ_{T_f} standard deviation of background or object pixel intensities separately, i.e.,

$$\begin{cases} \mu_{Ob_f} = average\left(In_f\right), \ in\{\varnothing \geq 0\} \\ \mu_{Bg_f} = average\left(In_f\right), \ in\{\varnothing < 0\} \end{cases} \tag{11}$$

$$\begin{cases} \sigma_{Ob_f} = std\left(In_f\right), \ in\{\varnothing \geq 0\} \\ \sigma_{Bg_f} = std\left(In_f\right), \ in\{\varnothing < 0\} \end{cases} \tag{12}$$

where μ_{Ob_f} and μ_{Bg_f} are the object (Ob) and background (Bg) mean values of pixel intensities, respectively, and is represented in Eq. (11). σ_{Ob_s} and σ_{Bg_s} are the object (Ob) and the background (Bg) standard deviation values of pixel intensities, respectively, and is represented in Eq. (12).

The proposed method designed an energy function structure in Eq. (13) as

$$E = E_{int} + E_{ext} + E_{MRF} \tag{13}$$

where E_{int} internal energy method, E_{ext} external energy method and E_{MRF} MRF energy method. Curvature, length, and area are the some of the characteristics of the contour which are used to define the internal energy method. The evolution force driven by the image information was the main concern of the external energy function. The MRF energy method is added to the energy method structure that reflects the connection among pixels and its corresponding neighbors.

Level set methods (LSM) indirectly representation of the planar closed curve C by the zero-level set of the level set method $\varnothing(i, j, t)$ are given in Eq. (14).

$$C(t) = \{(i, j) | \varnothing(i, j, t) = 0\} \tag{14}$$

The level set equation in lieu of the evolution of $\varnothing(i,j,t)$ can be denoted in Eq. (15) as

$$\frac{\partial \varnothing}{\partial t} + H|\nabla \varnothing| = 0 \tag{15}$$

where H is the speed of the evolution and ∇ is the gradient operator, respectively. The problem of reducing a certain energy method defined depending on the level set method $\varnothing$ is explained in Eq. (16).

$$\frac{\partial \varnothing}{\partial t} = -\frac{\partial E}{\partial \varnothing} = T_i + T_e + T_{MRF} \tag{16}$$

where T_i, T_e, and T_{MRF} are the internal and external terms and MRF term resultants from MRF energy method (2), respectively. T_{MRF} is given in Eq. (17).

$$T_{MRF} = -\delta(\varnothing)(Eng(Ob|In) - Eng(Bg|In)) \tag{17}$$

where $\delta(\varnothing)$ is the unit impulse function (18) by the following form

$$\delta_\varepsilon(\varnothing) = \begin{cases} 0, & |\varnothing| > \varepsilon \\ \frac{1}{2\varepsilon}\left(1 + cos\left(\frac{\pi\varnothing}{\varepsilon}\right)\right), & |\varnothing| \le \varepsilon \end{cases} \tag{18}$$

To avoid the re-initialization, the internal term T_i (19) is computed on the basis curvature of the evolving contour and external term T_e (20) is tallied grounded on Li's penalty.

$$T_i = \lambda|\nabla \varnothing|\nabla . g\left(\frac{\nabla \varnothing}{|\nabla \varnothing|}\right) \tag{19}$$

$$T_e = \beta[\nabla \varnothing - div\left((\nabla \varnothing)/|\nabla \varnothing|\right)] \tag{20}$$

Here λ and β are +ive constants.

The information of the initial contour C^0 used to initialize the level set method $\varnothing$ (21), i.e.,

$$\varnothing(i,j)^0 = \begin{cases} a, & (i,j) \in in(C^0) \\ 0, & (i,j) \in C^0 \\ -a, & (i,j) \in out \in (C^0) \end{cases} \tag{21}$$

Here, a is +ive constant value.

Several constraints affect the segmentation functioning in the proposed method, including the Gibbsian parameter γ, the parameter $\mathcal{W}m$ controls the dimensions of the neighborhood, and the parameters λ and β determining the weight of the internal and external terms in the energy method framework.

4 Result Analysis

Comparative analysis is conducted between the proposed technique and the current technique on one hundred images data. The functioning is calculated using mean, standard deviation, and entropy. There are degradations in shadows, low contrast, noise, background intensity of data images. Comparison of proposed method with the current manifold technique on data is notified in terms of quantity. The proposed method results analysis is subdivided. They are subjective and experimental evaluation, respectively.

4.1 Subjective Evaluation

In this section presents a subjective evaluation of the proposed MRF energy method against the current manifold learning method. The proposed method is applied to the 100 images. On the result analysis, ten image outcomes are given of the current manifold method and expressed the approach. In every case, comparison of the current and the proposed methods is given. Figure 1 shows the original input image, output images of the current manifold technique and proposed MRF technique. For the given image ImgA, proposed MRF approach perfectly segmented the object image and their characteristics are identified exactly which includes mouth, nose. In contrast, surviving method did not recognize these features exactly. For the given image ImgB, the proposed MRF technique identified the four quadrangles in contrast to the surviving approach identified only three quadrangles. For given images ImgC and ImgF, the proposed MRF technique recognized all the characteristics of the horsetail, legs, body, and head, whereas the existing method could not identify them. For the given image ImgD, the proposed method identifies the man and horse perfectly which the current approach couldn't. For given images ImgI and ImgE, the proposed method recognized the person exactly, whereas the current approach identified only few. The proposed MRF model recognized the stones and human exactly of the input image ImgG, but current model could not do. The rice recipes are identified perfectly by the proposed MRF technique, but existing technique does not recognize for the ImgH input image. Animal exactly segmented by the proposed technique for the ImgJ input image, but existing technique does not segment the animal perfectly. From the observation of the results of the existing and proposed techniques, proposed MRF technique recognized the objects perfectly compared with the existing manifold learning technique.

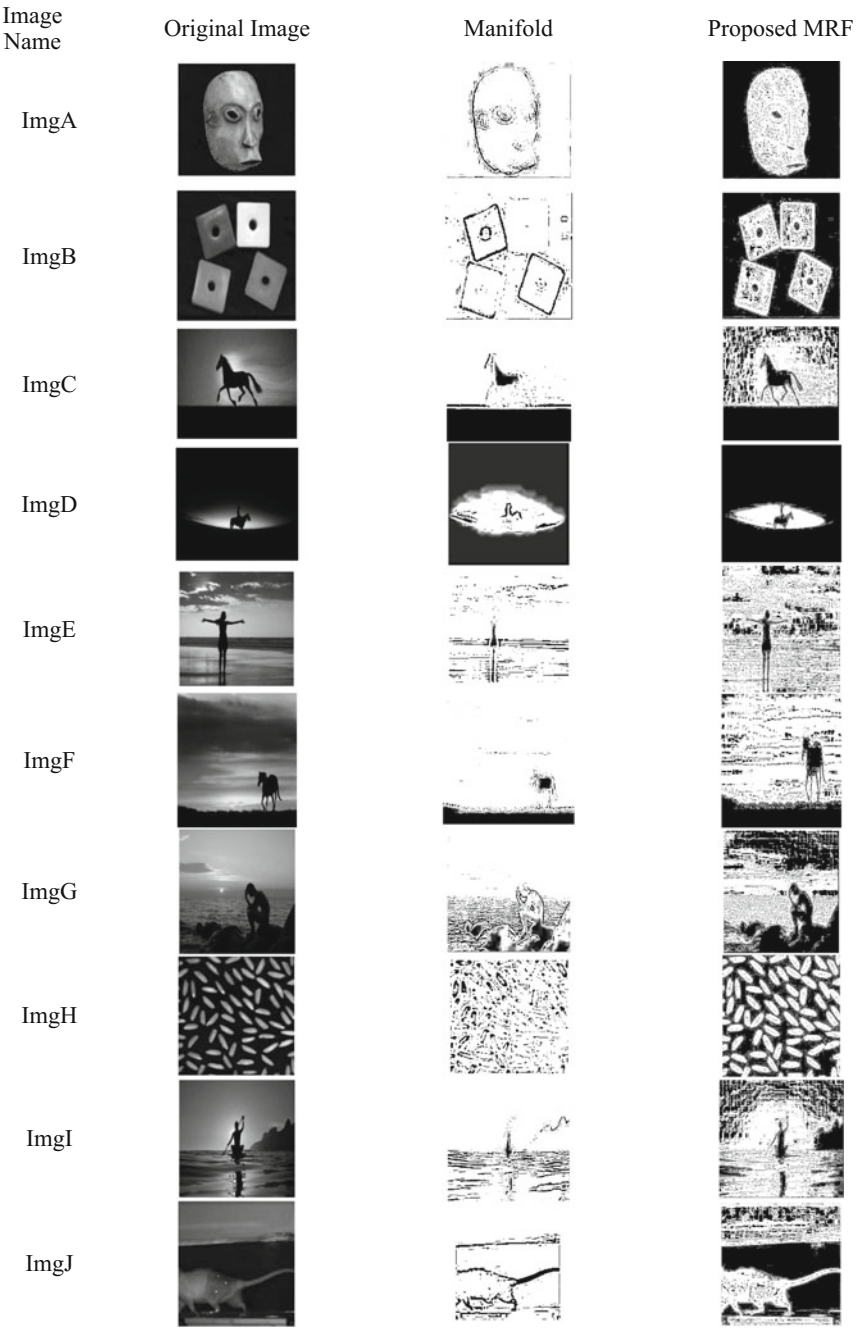


Fig. 1 From left to right, image name, original image, existing manifold, and proposed method output images

4.2 Experimental Evaluation

This section provides four performance metrics methods. They are mean, standard deviation, and entropy. By applying the qualitative evaluation methods on the existing manifold learning technique and proposed MRF method, proposed MRF technique gives the quality image segmentation results in contrast to existing manifold learning method. With test application of each test image, *Mean* (μ) values are shown in Table 1 for existing manifold learning method and proposed MRF method. The optimum threshold value of the proposed method in comparison with the current method affirms the qualitative superiority of MRF approach against the current. Standard deviation (σ) results are presented in Table 2 for current manifold learning technique and MRF technique. MRF method yields the optimum threshold value in comparison with manifold learning approach that reinstates the quality measure against the current method. *E*ntropy *H* results are presented in Table 3 for the current manifold learning technique and MRF technique. MRF technique yields the maximum entropy value in comparison to manifold technique. The proposed MRF method confirms the qualitative enhancement over the current manifold method.

Table 1 Mean

Image	Mean	
	Manifold	Proposed MRF
ImgA	0.9008	0.3512
ImgB	0.8859	0.3556
ImgC	0.6557	0.4560
ImgD	0.2994	0.1049
ImgE	0.8910	0.7045
ImgF	0.8544	0.6145
ImgG	0.7932	0.4663
ImgH	0.7708	0.4882
ImgI	0.8700	0.6228
ImgJ	0.8961	0.4416

Table 2 Standard deviation

Image	Standard deviation	
	Manifold	Proposed MRF
ImgA	0.0965	0.1681
ImgB	0.0690	0.1533
ImgC	0.0595	0.0656
ImgD	0.1052	0.1330
ImgE	0.0580	0.0631
ImgF	0.0506	0.0613
ImgG	0.0615	0.0648
ImgH	0.0576	0.0707
ImgI	0.0596	0.0702
ImgJ	0.0525	0.0739

Table 3 Entropy

Image	Entropy	
	Manifold	Proposed MRF
ImgA	0.6645	1.7748
ImgB	0.8719	2.1031
ImgC	1.0990	1.6399
ImgD	0.8831	1.6607
ImgE	0.6351	1.9329
ImgF	0.7808	1.7100
ImgG	1.3269	2.1855
ImgH	1.3379	3.2259
ImgI	0.7240	2.0395
ImgJ	0.6378	2.0581

5 Conclusion

This research project an innovative technique for image segmentation with the generating random fields using the MRF energy level set mode. The success of the proposed method is confirmed on the basis of the images in comparison with the manifold technique using the qualitative measures standard deviation, mean, and entropy. The proposed MRF energy level set model gives better segmentation results compared with the current manifold technique. But the results of the proposed method contain some noise in the images, and we will overcome this problem in future.

References

1. Etyngier, P., Keriven, R. Segonne, F.: Projection onto a shape manifold for image segmentation with prior. In: ICIP 2007. IEEE International Conference on Image Processing, vol. 4, pp. IV-361–IV-364, 16 Sept 2007–19 Oct 2007
2. Etyngier, P., Segonne, F., Keriven, R.: Shape Priors using Manifold Learning Techniques.In: IEEE 11th International Conference on in Computer Vision, 2007. ICCV 2007, pp. 1–8, 14–21 Oct 2007
3. Aljabar, P., Wolz, R., Rueckert, D.: Manifold learning for medical image registration, segmentation, and classification. Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis: Medical Imaging Intelligence and Analysis, p. 351 (2012)
4. Duc, A.K.H., et al.: Using manifold learning for atlas selection in multi-atlas segmentation. PloS one **8.8**, e70059 (2013)
5. Liu, X., Song, M., Tao, D., Bu, J., Chen, C.: Random geometric prior forest for multiclass object segmentation. IEEE Trans. Image Process. **24**(10), 3060–3070 (2015)
6. Pemula, R., Nagaraju, C.: Generation of random fields for image segmentation using manifold learning technique. Int. J. Pattern Recognit. Artif. Intell. **30**(10), 1654007 (14 pp) (2016)
7. Li, B., Zheng, C.-H., Huang, D.-S.: Locally linear discriminant embedding: an efficient method for face recognition. Pattern Recognit. **41**, 3813–3821 (2008)

8. Golchin, E., Maghooli, K.: Overview of manifold learning and its application in medical data set. Int. J. Biomed. Eng. Sci. (IJBES) **1**)(2) (2014)
9. Li, M., Luo, X., Yang, J., Sun, Y.: Applying a locally linear embedding algorithm for feature extraction and visualization of MI-EEG. Hindawi Publishing Corporation, J. Sensors, Article ID 7481946 (2016)
10. Yin, H., Huang, W.: Adaptive nonlinear manifolds and their applications to pattern recognition. Inf. Sci. **180**, 2649–2662 (2010)
11. Barker, S.A., Rayner, P.J.W.: Unsupervised image segmentation using Markov random field models. Pattern Recognit. **33**, 587–602 (2000)
12. Bing, C., Wang, Y., Zheng, N., Jia, X., Bian, Z.: MRF model and FRAME model-based unsupervised image segmentation. Sci. China Ser F Inf. Sci. **47**(6), 697–705 (2004)
13. Nagaraju, C., Siva Shankar Reddy, L.: Abnormality Detection by generating random fields based on Markov random field theory. IJCSNS Int. J. Comput. Sci. Netw. Secur. **8**(5) (2008)
14. Nagaraju, C., Siva Shankar Reddy, L.: Generation of entropy based binary random fields for image bundary detection based on fuzzy semantic rules. IJCSNS Int. J. Comput. Sci. Netw. Secur. **8**(5) (2008)
15. Nagaraju, C., Srinivasa Rao, P.S.V., Girija Rani, G.: A novel method for generation of random fields for boundary detection and classification of images. IJCSNS Int. J. Comput. Sci. Netw. Secur. **8**(7) (2008)
16. Nagaraju, C., Reddy, L.S.S., Ramesh Babu, D.: Automatic extraction of linear features using morphological operations and Markov randomfields. Int. J. Inf. Technol. Knowl. Manag. **2**(1), 5–8 (2009)
17. Rambabu, P., Naga Raju, C.: Generation of random fields for object recognization using binarization technique. Int. J. Emerg. Trends Technol. Comput. Sci. (IJETTCS) **4**(5)(2), 8–15 (2015)
18. Taheri, S., Ong, S.H., Chong, V.F.H.: Level-set segmentation of brain tumors using a threshold-based speed function. Image Vis. Comput. **28**, 26–37 (2010)
19. Li, C., Huang, R., Ding, Z., Gatenby, J.C., Metaxas, D.N., Gore, J.C.: A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE Trans. Image Process. **20**(7), 2007–2016 (2011)
20. Yang, X., Gao, X., Tao, D., Li, X., Li, J.: An efficient MRF embedded level set method for image segmentation. IEEE Trans. Image Process. **24**(1) (2015)
21. Pemula, R., Raju, C.N.: Generation of optimal random fields for image segmentation using fuzzy multi-region technique. In: 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, pp. 1–6 (2016)

Recognition of Islanding Data for Multiple Distributed Generation Systems with ROCOF Shore Up Analysis

Ch. Rami Reddy, K. Harinadha Reddy and K. Venkata Siva Reddy

Abstract In this paper, a new hybrid islanding detection method is presented for protecting the distributed system with unintentional islanding, with low-frequency current harmonic injection and over/under frequency relay for inverter-based multiple distributed generation (DG) systems. Low-frequency current harmonic component is injected into the system through the q-controller of the grid-side controller. The injected low-frequency current component causes the system frequency to deviate during islanding. This deviation separates islanding and non-islanding events. This paper evaluates the performance of this hybrid ROCOF relay when load and generation are matched and also at different active power imbalances. This paper will also clearly differentiate between islanding and non-islanding conditions for mixed types of DG's connected to the grid. The test system results are carried out on the MATLAB/Simulink setting that shows the strength of this technique.

1 Introduction

Nowadays, the global energy demand is supplied by the combustion of fossil fuel resources. Due to the continuous combustion of fossil fuels, the environment is affecting and fossil fuels are decreasing. So the alternate solution for global energy consumption demand and pollution-free country is the use of renewable energy

Ch.R. Reddy (✉) · K. V. S. Reddy
Department of Electrical and Electronics Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur 522502, Andhra Pradesh, India
e-mail: crreddy229@gmail.com

K. V. S. Reddy
e-mail: sivareddy.v.k@gmail.com

K. H. Reddy
Department of EEE, Lakireddy Bali Reddy College of Engineering (Autonomous), Mylavaram, Krishna 521230, Andhra Pradesh, India
e-mail: kadapa.hari@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_54

resources like solar, wind. Distributed generation is the smallest scale power generation source which is connected at the consumer level of power system [1]. Most of the DG systems are using renewable energy resources and interfaced to the grid with suitable power converters. Depending on the nature of the source, the energy may be generated in AC or DC forms. Due to failures in the grid, if the DG is disconnected from the main grid, the DG along with local load forms an islanding [2]. The islanding is unsafe to field persons and equipments connected because the servicing persons are not mindful that the frame up is connected and supplying with DG near. The main causes of such unintentional islanding is due to the failures detected by the grid, accidental opening of circuit breakers at the grid, intentional opening of CB for maintenance, human errors, and an act of nature [3]. The basic grid interfacing rules such as IEEE 1547 and UL 1741 needs that it is necessary to disconnect the DG source within 2 s, because if the island load is more or less, and then it leads to variations in the voltage, frequency, current, THD, active, reactive powers outside the standards, which may hazardous to customer loads connected to it [4, 5]. Different control systems and inverters are implemented to connect DG sources to grid with proper synchronization and to inject high-quality power into the grid [6, 7]. PLCC and SCADA are the remote islanding detection techniques; they detect the islanding by gathering information from DG side and utility side. Utility side signals are monitored by the PLCC [8]. If these signals are not appearing, then islanding is detected. On the other hand, islanding was detected by SCADA with information from circuit breaker auxiliary contacts. The implementation of these methods is very difficult because the cost and implementation of other monitoring devices, transmitters, and receivers are more [9]. By injecting some signal at PCC for some cycles and observing the deviations in the output signal active methods will detect the islanding. In the grid-connected system, the system absorbs the local disturbance and considerable deviations are not observed. However, more deviations are observed in the output signal if the system is islanded [10–13]. By observing these deviations, the islanding is detected by active methods. The active methods are more efficient than passive methods, but they reduce the power quality [14–16]. Large samples of data collected from non-islanding and islanding events such as voltage, current, power, p.f and processing it with the learning algorithms, and the islanding is detected by machine learning detection techniques [17–19]. In passive techniques, regional parameters such as voltage, frequency, current, phase angle, total harmonic distortion (THD) are monitored at the PCC, if there are changes beyond a certain threshold level, then islanding is detected [20–24]. Passive islanding techniques monitor the parameters like frequency, phase jump, rate of change of reactive power, THD and others to detect islanding [25–28]. The low-frequency current injected through the DQ controller of inverter-based distributed generation presented [29, 30] will work inside the NDZ of a conventional frequency relay. In this paper, a new hybrid islanding detection method is presented with a passive method of over/under frequency relay and active method of low-frequency current injection through the q-controller of the synchronous rotating frame controller of the grid-side converter. The over/under frequency relay works when the frequency is more or less than 50.5 and 49.5 as per

Indian standards, and the current injection-based relay works inside the NDZ of the conventional frequency relay. The proposed hybrid ROCOF relay is also differentiating between islanding and non-islanding events. Furthermore, this technique has no non-detection zone (NDZ). The paper is organized as follows. System under the study along with control system is discussed in Sect. 2, the proposed hybrid ROCOF relay is presented in Sect. 3, the simulated results and comparison with the existing methods are presented in Sect. 4, and conclusions are drawn in Sect. 5.

2 Paper Preparation

The test system considered as per distribution energy sources interconnection standards, such as IEEE 1547 [4, 5] and UL 1741 is shown in Fig. 1. During the grid-connected mode, the CB is closed, DG is synchronized with grid and quality power is injected into the grid, but when the CB is opened due to failures in the grid, the DG, and local load forms islanding, so that customer equipment may damage. During this condition, islanding should be detected within 2 s after creation of the island as per interconnection standards. A 100 KW PV array is connected to a 120 kV grid via a DC-DC boost converter three-phase three-level voltage source converter (VSC) and a step-up transformer The PV array uses 330 sun power modules (SPR-305E-WHT-D) in which 66 strings of 5 series connected modules are connected in parallel to produce 1 KW power (66 * 5 * 305.2 W = 100.7 KW) with an open circuit voltage of 64.2 V and a short circuit

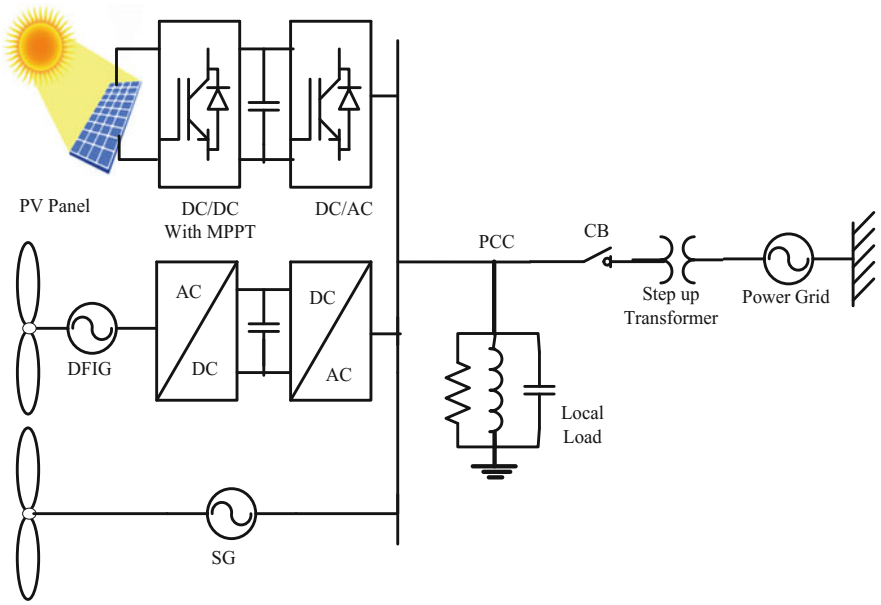


Fig. 1 Multiple distributed generation system under study

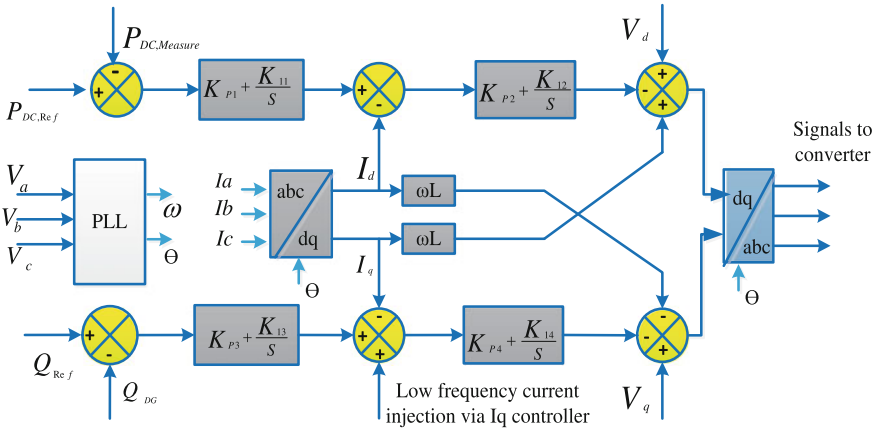


Fig. 2 Control system for injecting current harmonic

current of 5.96 A at 1000 W/m^2 sun irradiance and 25 °C ambient temperature. The boost converter uses maximum Power Point tracking (MPPT) technique to extract maximum power from the panel under irregular weather conditions, which is implemented using incremental conductance and integral regulator technique.

The boost converter increases the voltage level to 500 V DC and is given to the input of the inverter. Wind farm 1.5 MW with doubly fed induction generators consisting wind turbines. Wind turbines using a doubly fed induction generator consist of a wound rotor induction generator and an AC/DC/AC IGBT-based PWM converter with synchronous reference controller [7, 14] as shown in Fig. 2. Third DG of 10 MW synchronous generator-based DG is considered. The DFIG technology allows extracting maximum energy from the wind for low wind speeds by optimizing the turbine speed, while minimizing mechanical stresses on the turbine during gusts of wind. The output of the wind and PV inverters is connected to the low voltage bus and is connected to the grid through circuit breaker and step-up transformer.

3 Proposed Hybrid ROCOF Relay Theory and Structure

The conventional ROCOF relay shown in Fig. 3 [10] will work on Eqs. (1–3) [10, 13]. The frequency is found with three phase-Phase Locked Loop module (PLL) from voltage signals at the point of common coupling (PCC). The effective value of ROCOF is calculated by Eq. (1), having a measurement window of 500 ms, and the trip signals are sent to the CB if the calculated effective value of ROCOF is more than a threshold value or if the frequency is more or lesser than the threshold settings. The threshold value for large mismatch between load and generation is

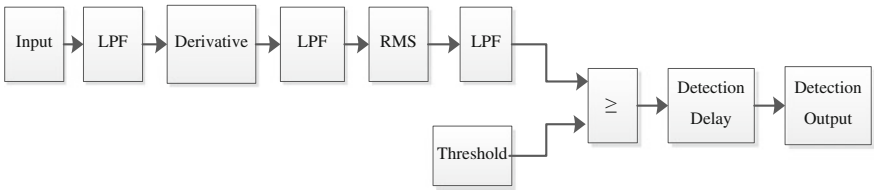


Fig. 3 Conventional ROCOF relay structure

considered as 10 Hz/s for a large measurement window of 500 ms, but has an advantage; it cannot trip for non-islanding cases

$$\text{ROCOF} = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{df}{dt}(k)\right| \tag{1}$$

where $n = \frac{\Delta t}{T_s}$

$n =$ number of samples for a window size of Δt.
$k = 1, 2, 3 \ldots\ldots\ldots n$th sample and
$T_s =$ Sampling time

$$\frac{df}{dt}(k) = \frac{f(t_k) - f(t_k - \Delta t)}{\Delta t} \tag{2}$$

$f(t_k)$—calculated value of frequency at the time of kth sample, i.e., t_k.

$f(t_k - \Delta t)$—measured value of frequency, Δt before the kth sample time, i.e., $t_k - \Delta_t$.

The ROCOF relay operation depends on the imbalance between load and generation and frequency changes at PCC. The changes in the frequency wave will depend on the mismatch between load and generation [28] expressed in (13),

$$\frac{df}{dt} = \frac{P_{DG} - P_{LOAD}}{2 \times S_{DG} \times H_{DG}} \tag{3}$$

where P_{DG}—output power of DG
P_{load}—load power
f_0—rated fundamental frequency
S_{DG}—rating of DG
H_{SG}—inertia of wind machines

So, if the difference between load and generation is less or zero, then the dynamic changes in frequency are zero and islanding cannot be detected. To improve the performance of the conventional ROCOF relay and to eliminate the

wrong operation during non-islanding, hybrid islanding detection relay is presented. It has a small measurement window of 50 ms with low-frequency current injection and new threshold settings. In this method, the deviations in frequency are used to detect worst case islanding [25–30].

The deviations in frequency are achieved by injecting a low-frequency current wave of 1% of DG capacity. By considering the reduced window size of 50 ms, the deviations in the frequency are clearly observed. The developed ROCOF relay and its algorithm with new threshold settings are shown in Fig. 4. A low-frequency signal of 30 Hz is injected through the q-controller of grid-side synchronous reference controller shown in Fig. 2. Under grid-connected operation the injected wave has no effect on frequency at PCC, but when the system is islanded the low-frequency signal is forced to load through low impedance path and leads to frequency deviations. The three-phase PLL connected at PCC will measure the frequency, and the effective value is calculated for every 50 ms of time. For injected current frequency of 30 Hz, with a large difference between load and generation, the threshold value is set as 10 Hz/s. A trip signal is sent if the ROCOF value is more than the active ROCOF and will trip the DG if ROCOF is greater than T1 for Hz/s. The effective value is calculated for two times for every 50 ms and compared with threshold. If it is more than the threshold value for two times, then it is treated as permanent change and trip signal will send to turn off the inverter. So the total time taken for decision making during large mismatch is 100 ms. During zero power mismatch condition and non-islanding condition, the ROCOF compares with a threshold value T2 for two times, i.e., 100 ms, if it is more for two times, then

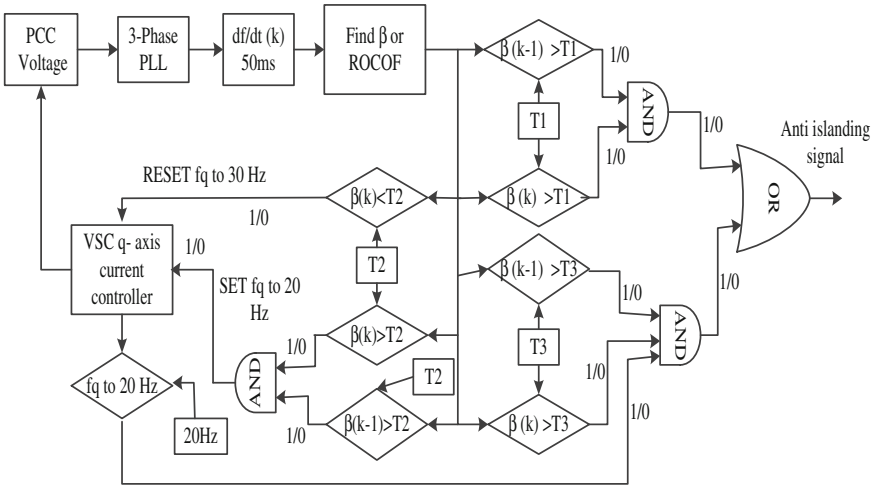


Fig. 4 Proposed ROCOF relay structure

injected frequency is changed to 20 Hz and ROCOF is calculated for another two consecutive instants. For these two instants, if it is more than the threshold value of T3 Hz/s, then it is treated as islanding with zero power imbalance condition, i.e., the total time for detecting islanding during low power imbalance is 200 ms. In non-islanding cases, it is observed that the effective ROCOF is more than T2 for times but less than T3 for another two times. So there is no possibility of nuisance tripping. So the proposed ROCOF has the ability to detect islanding during zero power imbalance condition with new threshold settings, and there is no possibility of nuisance tripping. For non-islanding cases, ROCOF is less than threshold value T3, then the injected signal frequency is reset to 30 Hz and the algorithm is repeated.

4 Results and Discussion

The voltages, currents, active power, and reactive power of the DG during grid-connected and islanding modes are depicted in Fig. 5a–d. The simulation is run for 1.1 s, and islanding is created at 1 s. The oscillations in waveforms are due to islanding. The hybrid ROCOF relay is found to be effective in detecting islanding during zero power imbalance condition, islanding is simulated at 1 s, and simulation is run for 2 s. The performance of the proposed hybrid ROCOF relay has tested to a load of half of DG capacity. The simulated results are shown in Fig. 6a–c. The effective value of ROCOF is calculated for every 50 ms after islanding at 1 s, and it

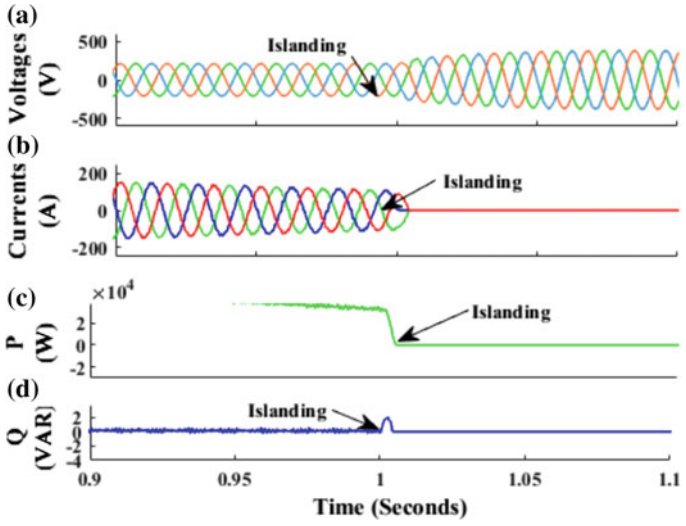


Fig. 5 Performance during pre- and post-islanding **a** Voltages of DG **b** DG currents **c** Active power of DG **d** Reactive power of DG

is found to be more than T1 of 10 Hz/s for two successive calculations. As per our algorithm, the trip signal is initiated at 1.1 s, i.e., 100 ms. Figure 7a–c shows the waveforms of a load of 90% of DG capacity.

Here the effective value is calculated for every 50 ms after 1 s, i.e., after islanding. The effective value is found more than T2 of 5 Hz/s for two successive calculations. At 1.1 s, the frequency of injected current is set to 20 Hz and the calculated effective value of ROCOF is found to be more than T3 of 8 Hz/s for other 100 ms and is considered as permanent change and islanding detection signal is generated at 1.2 s. Figure 8a–c shows the waveforms of zero power imbalance condition.

Here the effective value is calculated for every 50 ms after 1 s, i.e., after islanding. The effective value is found more than T2 of 5 Hz/s for two successive calculations. At 1.1 s, the frequency of injected current is set to 20 Hz and the

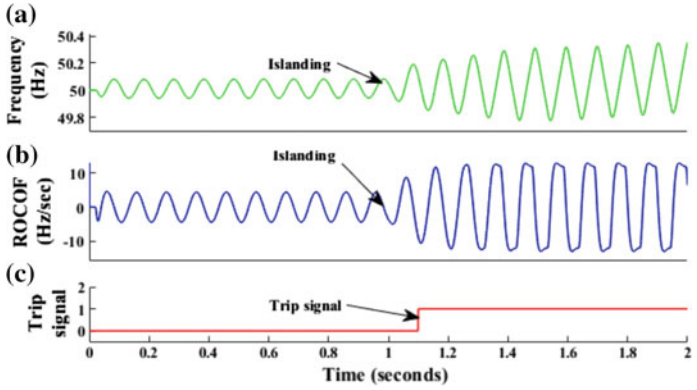


Fig. 6 Proposed relay performance when load is half of DG capacity **a** Frequency of DG **b** ROCOF **c** Trip signal

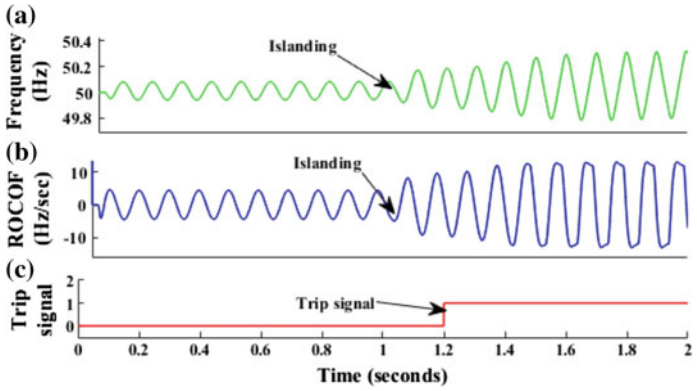


Fig. 7 Proposed relay performance when the load is 90% of DG capacity **a** Frequency of DG **b** ROCOF **c** Trip signal

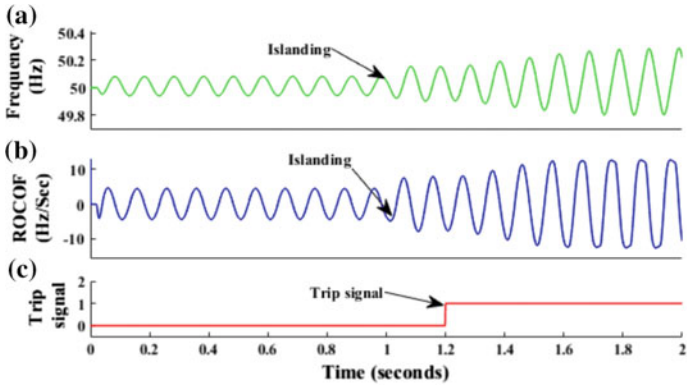


Fig. 8 Proposed relay performance when load capacity is same as DG capacity **a** Frequency of DG **b** ROCOF **c** Trip signal

calculated effective value of ROCOF is found to be more than T3 of 8 Hz/s for other 100 ms and is considered as permanent change and islanding detection signal is generated at 1.2 s. From Figs. 6, 7, and 8, for zero power imbalance condition, the islanding is detected within 200 ms and large power mismatch loads it is 100 ms.

The non-islanding cases of capacitor bank switching, same capacity DG switching are simulated as shown in Fig. 9a–c and from Fig. 10a–c. It is found that the effective value is less than T2, and these cases are considered as non-islanding cases and no trip signals are initiated as per our algorithm.

4.1 Comparison with the Existing Methods

The general ROCOF proposed in the paper works if the difference between load and generation are more, and it cannot work during generation and load are matches

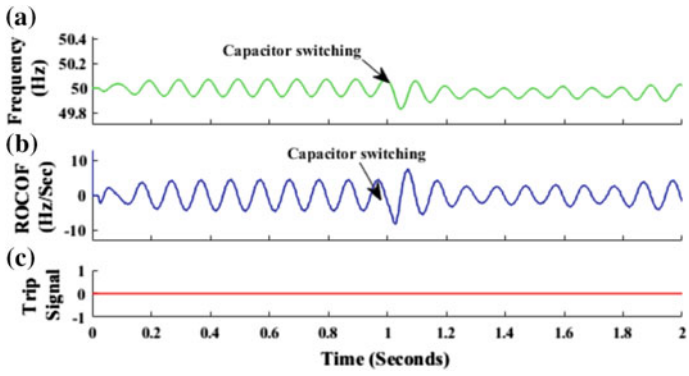


Fig. 9 Proposed relay performance for capacitor bank switching **a** Frequency of DG **b** ROCOF **c** Trip signal

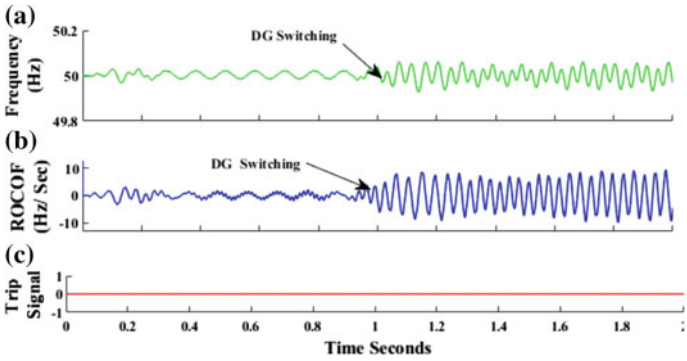


Fig. 10 Proposed relay performance during switching of another DG of the same capacity **a** Frequency of DG **b** ROCOF **c** Trip signal

[13]. These are not detecting islanding when load and generation are matching. It has a measurement window of 500 ms, and it is not observing the deviations in the frequency during islanding due to large measurement window, but has an advantage it cannot trip for non-islanding cases. The NDZ of a conventional frequency relay is more and is not responding to zero power imbalance conditions. The hybrid frequency relay proposed in this paper is using a measurement window of 50 ms and is having the capability of detecting deviations in frequency due to low frequency current injection. Due to low-frequency current injection, the frequency of the system is deviated with some angle. It responds only if there is a continuous deviation of frequency for 100 ms for high power difference. For less than 100 ms, it will treat as a non-islanding condition due to switching of loads, transients, etc.

5 Conclusion

The hybrid ROCOF relay with the proposed algorithm proposed in this paper capable of detecting islanding during the DG and load generation are matched and low power mismatches. The injected low-frequency current harmonic will deviate the system frequency and leads to increased oscillations in ROCOF during balanced islanding or zero or low power imbalance conditions. To analyze the ROCOF at PCC, a measurement window of 50 ms is sufficient with new threshold settings proposed in this paper than a conventional frequency relay of 500 ms. This method detects islanding if there is a continuous deviation of the effective value of ROCOF for 100 ms for more difference between the DG and load power and 200 ms for zero power imbalances. From the method presented in this paper, there is no continuous deviation for 200 ms during non-islanding conditions. So it is clearly differentiating between the islanding and non-islanding conditions and nuisance trips are prevented. The amplitude of injected current is kept 1% of DG capacity for

maintaining good power quality and unity power factor due to disturbance. The different islanding and non-islanding cases are simulated and obtained trip signals for islanding and non-islanding cases. The future work comprises of anti-islanding protection with this method and hardware implementation.

References

1. Bindra, A.: A Unified Interface for Integrating Renewable Energy Sources with the Smart Grid. IEEE Power Electron. Mag. **3**(4), 4–6 (2016)
2. Liserre, M., Sauter, T., Hung, J.Y.: Future energy systems: Integrating renewable energy sources into the smart power grid through industrial electronics. IEEE Ind. Electron. Mag. **4** (1), 18–37 (2010)
3. Han, H., Hou, X., Yang, J., Wu, J., Su, M., Guerrero, J.M.: Review of power sharing control strategies for islanding operation of AC micro grids. IEEE Trans. Smart Grid **7**(1), 200–215 (2016)
4. IEEE Standard conformance test procedures for equipment interconnecting distributed resources with electric power systems. IEEE 1547.1 Standard (2005)
5. UL1741 standard for safety for static converters and charge controllers for use in photovoltaic power systems. Underwriters Laboratories (2001)
6. Bhatt, Y.P., Shah, M.C.: Design, analysis and simulation of synchronous reference frame based phase lock loop for grid connected inverter. In: IEEE International Conference on Power Electronics Intelligent Control and Energy Systems (ICPEICES-2016), pp. 1–5 (2016)
7. Jana, J., Saha, H., Bhattacharya, K.D.: A review of inverter topologies for single phase grid connected photovoltaic systems. Renew. Sustain. Energy Rev. **72**, 1256–1270 (2017)
8. Sarkar, S.J., Kundu, P.K.: A proposed method of load scheduling and generation control using GSM and PLCC technology. Michael Faraday IET International Summit, pp. 273–277 (2015)
9. Zhang, J.Y., Bush, C.M.: PMU based islanding detection to improve system operation. In: IEEE Power and Energy Society General Meeting (PESGM), pp. 1–5 (2016)
10. Reddy, Ch.R., Reddy, K.H.: A passive islanding detection method for neutral point clamped multilevel inverter based distributed Generation using rate of change of frequency analysis. Int. J. Electr. Comput. Eng. **08**(04), 1967–1976 (2018)
11. Gupta, P., Bhatia, R., Jain, D.: Average absolute frequency deviation value based active islanding detection technique. IEEE Trans. Smart Grid **6**(1), 26–35 (2015)
12. Jia, K., Wei, H., Bi, T., Thomas, D., Sumner, M.: An islanding detection method for multi-DG systems based on high frequency impedance estimation. IEEE Trans. Sustain. Energy **8**(7), 74–83 (2017)
13. Reddy, Ch.R., Reddy, K.H.: Islanding detection for inverter based distributed generation with low frequency current harmonic injection through Q controller and ROCOF analysis. J. Electr. Syst.**14**(02), 179–191 (2018)
14. Guo, Z.: A harmonic current injection control scheme for active islanding detection of grid-connected inverters. In: IEEE International Telecommunications Energy Conference (INTELEC), pp. 1–5 (2015)
15. Dhar,S., Chandak, S., Naeem, M.H.: Harmonic profile injection based active islanding detection for PV-VSC based grid. In: 2015 IEEE Power Communication and Information Technology Conference (PCITC), pp. 489–496 (2015)
16. Pouryekta, A., Ramachandaramurthy, V.K., Selvaraj, J.: Active islanding detection for synchronous generators using speed disturbance injection technique. In: IEEE Conference on Clean Energy and Technology (CEAT), pp. 54–59 (2013)

17. Salman, S., King, D., Weller, G.: Investigation into the development of a new ANN -based relay for detecting loss of mains of embedded generation. In: IEEE International Conference on Developments in Power System Protection, vol. 2, pp. 579–582 (2014)
18. Faqhruldin, O., El-Saadany, E., Zeineldin, H.: A universal islanding detection technique for distributed generation using pattern recognition. IEEE Trans. Smart Grid **5**(4), 1985–1992 (2015)
19. Matic-Cuka, B., Kezunovic, M.: Islanding detection for inverter based distributed generation using support vector machine method. IEEE Trans. Smart Grid **5**(6), 2676–2686 (2014)
20. Vyas, S., Kumar, R., Kavessari, R.: Data analytics and computational methods for anti-islanding of renewable energy based distributed generators in power grids. Renew. Sustain. Energy Rev. **69**, 493–402 (2017)
21. Velasco, D., Trujillo, C., Garcer, G., Figueres, E.: Review of anti islanding techniques in distributed generators. Renew. Sustain. Energy Rev. **14**(6), 1608–1614 (2010)
22. Mahat, P., Chen, Z., Bak-Jensen, B.: Review of islanding detection methods for distributed generation. In: Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), pp. 2743–2748 (2016)
23. Satish Kumar, P.: FPGA Implementation of space vector pulse width modulated neutral point clamped three-level inverter fed induction motor Drive. In: IEEE Conference on Power, Control, Communication and Computational Technologies for Sustainable Growth (PCCCTSG), pp. 222–230 (2015)
24. Wada, K., Yoshida, H.: Improvement of Power Quality for Three-Phase Grids using Single-Phase DG with Active Filter Function Units. In: IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia) (2016)
25. Huayang, W., Zuyuan, Y., Long, Z., Yuan, C.: Based on hybrid algorithm of distributed power islanding detection. In: IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), pp. 1360–1364 (2016)
26. Azim, R., Li, F., Zhao, X.: A hybrid islanding detection technique for inverter based distributed generations. In: IEEE Electrical Power and Energy Conference (EPEC), pp. 239–243 (2015)
27. Reddy, Ch.R., Reddy, K.H.: Islanding detection method for inverter based distributed generation based on combined changes of ROCOAP and ROCORP. Int. J. Pure Appl. Math. **117**(19), 433–440 (2017)
28. Chen, X., Li, Y.: An islanding detection method for inverter-based distributed generators based on the reactive power disturbance. IEEE Trans. Power Electron. **31**(5), 3559–3574 (2016)
29. Reigosa, D., Briz, F., Blanco, C., García, P., Guerrero, J.M.:"Active islanding detection for multiple parallel connected inverter-based distributed generators using high frequency signal injection. In: IEEE Energy Conversion Congress and Exposition (ECCE), pp. 2719–2726, (2012)
30. Alsharidah, M., Ahmed, N.A., Alothman, A.K.: Negative sequence injection for islanding detection of GRID interconnected distributed generators. In: 2014 IEEE Electrical Power and Energy Conference, pp. 267–274 (2014)

PSO Algorithm Support Switching Pulse Sequence ISVM for Six-Phase Matrix Converter-Fed Drives

Ch. Amarendra and K. Harinadha Reddy

Abstract A matrix converter (MC) is an array of power electronic switches that are directly connected from input to output. The six-phase matrix converter (SPMC) is an advanced power electronic converter having three-phase input and six-phase output. The SPMC provides the six-phase output for the six-phase applications. In this paper, the SPMC is operated with the indirect space vector modulation (ISVM). This ISVM is unable to bring the harmonic content below standard value. The value of harmonics should be less than 5% as per the standard IEEE value. The optimization technique is able to reduce the harmonic content in the output to below 5%. The switching state can be optimized to get the best results. Every time the optimum switching state will be selected based on the cost function using PSO algorithm, and the same will be given to the MC. This optimization brings the THD below 5%. The simulation results of the proposed technique are shown in this paper for validating the method.

1 Introduction

The matrix converter is a direct AC–AC converter, so many topologies, modulation methods, and advantages were discussed in the literature [1–6]. The matrix converter-based multi-phase drive system (MPDS) gained more attention in the past decade, and wide research has been carried out for getting the best performance [7, 8].

Ch. Amarendra (✉)
Department of EEE, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur, AP, India
e-mail: amarendraamrita@gmail.com

K. Harinadha Reddy
Department of EEE, Lakireddy Bali Reddy College of Engineering (A),
Mylavaram, Krishan District, AP, India
e-mail: kadapa.hari@gmail.com

Ch. Amarendra
Department of EEE, AEC, Surampalem, AP, India

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_55

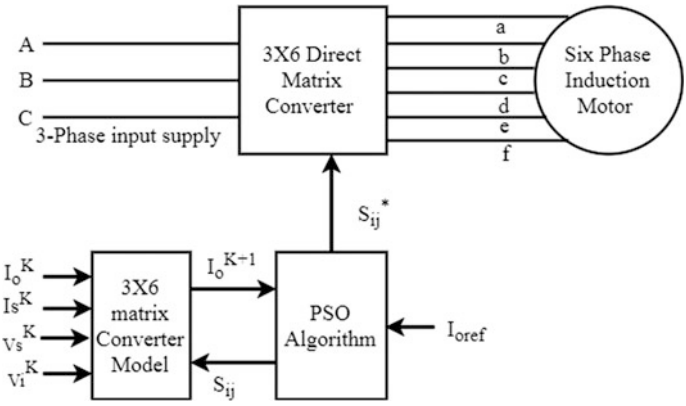


Fig. 1 Block diagram representation of six-phase MC for six-phase IM

Compared with the conventional three-phase drive system, MPDS has many advantages: (i) higher reliability, (ii) enhanced power per rms ampere ratio, (iii) lower torque pulsation, and (iv) reduced rated current of the power semiconductor switching. MPDS has a wide range of applications including transportation applications, industrial applications, and traction locomotives. Various MPDSs were discussed in the literature for various applications [9–12]; among all the MPDSs, six-phase induction motor (SPIM) is a popular drive system because of its identical and symmetrical windings separated by 30 electrical degrees. The proposed technique represented in the block diagram of Fig. 1, it contains the three-phase supply, three-phase to six-phase conversion SPMC, SPIM, PSO algorithm and matrix converter model. The paper is organized as follows: The SPIM mathematical modeling is discussed in Sect. 2, the modeling of SPMC is represented in Sect. 3, the conventional ISVM- and PSO-based ISVM modulation methods are discussed in Sects. 4 and 5, respectively, the simulation results are discussed in Sect. 6, and the conclusions are presented in Sect. 7.

2 Mathematical Modeling of Six-Phase Induction Motor

The state equation of six-phase induction motor can be written as (1–25)

$$[V_{abc}] = [R_1][I_{abc}] + \frac{d[\varnothing_{abc}]}{dt} \tag{1}$$

$$[V_{def}] = [R_2][I_{def}] + \frac{d[\varnothing_{def}]}{dt} \tag{2}$$

$$\left[V_{xyz}\right] = \left[R_r\right]\left[I_{xyz}\right] + \frac{d\left[\varnothing_{xyz}\right]}{dt} \tag{3}$$

$$J\frac{d\Omega}{dt} = T_{em} - T_r - K_f\Omega \tag{4}$$

J = moment of inertia of the moving part
K_f = viscous friction coefficient
T_{em} = load torque

$$\frac{\mathrm{d}\varnothing}{\mathrm{dt}} = [\mathrm{A}][\varnothing] + [\mathrm{B}][\mathrm{V}] \tag{5}$$

$$\varnothing = \begin{bmatrix} \varphi_{\mathrm{d1}} \\ \varphi_{\mathrm{d2}} \\ \varphi_{\mathrm{q1}} \\ \varphi_{\mathrm{q2}} \\ \varphi_{\mathrm{dr}} \\ \varphi_{\mathrm{qr}} \end{bmatrix} \tag{6}$$

$$\mathrm{V} = \begin{bmatrix} \mathrm{V_{d1}} \\ \mathrm{V_{d2}} \\ \mathrm{V_{q1}} \\ \mathrm{V_{q2}} \end{bmatrix} \tag{7}$$

The equation of electromagnetic torque is given by

$$\mathrm{T_{em}} = \frac{\mathrm{L_m}}{(\mathrm{L_m} + \mathrm{L_r})\left(\varphi_{\mathrm{dr}}\left(\mathrm{i_{q1}} + \mathrm{i_{q2}}\right) + \varphi_{\mathrm{qr}}(\mathrm{i_{d1}} + \mathrm{i_{d2}})\right)} \tag{8}$$

The equation of the flux is

$$\varnothing_{\mathrm{md}} = \mathrm{L_s}\left(\frac{\varnothing_{\mathrm{d1}}}{\mathrm{L_1}} + \frac{\varnothing_{\mathrm{d2}}}{\mathrm{L_2}} + \frac{\varnothing_{\mathrm{dr}}}{\mathrm{L_r}}\right) \tag{9}$$

$$\varnothing_{\mathrm{mq}} = \mathrm{L_s}\left(\frac{\varnothing_{\mathrm{q1}}}{\mathrm{L_1}} + \frac{\varnothing_{\mathrm{q2}}}{\mathrm{L_2}} + \frac{\varnothing_{\mathrm{qr}}}{\mathrm{L_r}}\right) \tag{10}$$

$$[\mathrm{A}] = \begin{bmatrix} \mathrm{a_{11}} & \mathrm{a_{12}} & \mathrm{a_{13}} & \mathrm{a_{14}} & \mathrm{a_{15}} & \mathrm{a_{16}} \\ \mathrm{a_{21}} & \mathrm{a_{22}} & \mathrm{a_{23}} & \mathrm{a_{24}} & \mathrm{a_{25}} & \mathrm{a_{26}} \\ \mathrm{a_{31}} & \mathrm{a_{32}} & \mathrm{a_{33}} & \mathrm{a_{34}} & \mathrm{a_{35}} & \mathrm{a_{36}} \\ \mathrm{a_{41}} & \mathrm{a_{42}} & \mathrm{a_{43}} & \mathrm{a_{44}} & \mathrm{a_{45}} & \mathrm{a_{46}} \\ \mathrm{a_{51}} & \mathrm{a_{52}} & \mathrm{a_{53}} & \mathrm{a_{54}} & \mathrm{a_{55}} & \mathrm{a_{56}} \\ \mathrm{a_{61}} & \mathrm{a_{62}} & \mathrm{a_{63}} & \mathrm{a_{64}} & \mathrm{a_{65}} & \mathrm{a_{66}} \end{bmatrix} \tag{11}$$

$$[\mathrm{B}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{12}$$

$$a_{11} = a_{32} = \frac{R_1 L_a}{L_1^2} - \frac{R_1}{L_1} \tag{13}$$

$$a_{12} = a_{34} = \frac{R_1 L_a}{L_1 L_2} \tag{14}$$

$$a_{56} = -a_{65} = \omega_{gl} \tag{15}$$

$$a_{11} = a_{32} = -a_{31} = -a_{42} = \omega_s \tag{16}$$

$$a_{14} = a_{16} = a_{23} = a_{26} = a_{32} = a_{35} = 0 \tag{17}$$

$$a_{41} = a_{45} = a_{52} = a_{54} = a_{61} = a_{62} = 0 \tag{18}$$

$$a_{15} = a_{36} = \frac{R_1 L_a}{L_1 L_r} \tag{19}$$

$$a_{21} = a_{42} = \frac{R_2 L_a}{L_1 L_2} \tag{20}$$

$$a_{22} = a_{44} = \frac{R_2 L_a}{L_2^2} - \frac{R_1}{L_1} \tag{21}$$

$$a_{25} = a_{46} = \frac{R_2 L_a}{L_r L_2} \tag{22}$$

$$a_{51} = a_{62} = \frac{R_r L_a}{L_1 L_r} \tag{23}$$

$$a_{52} = a_{64} = \frac{R_r L_a}{L_r L_2} \tag{24}$$

$$a_{55} = a_{66} = \frac{R_r L_a}{L_r^2} - \frac{R_r}{L_r} \tag{25}$$

3 Theory of Six-Phase Matrix Converter

The six-phase matrix converter (SPMC) is a 3 × 6 converter, i.e., three-phase input and six-phase output. Any phase of the output can be connected to any phase of the input. A total of 18 bidirectional switches are used in the topology of the SPMC. The input phases cannot be short-circuited, and the output phases cannot be open-circuited. The detailed SPMC is represented in Fig. 2. The modeling of SMC is expressed in (26–34)

The switching function can be defined as

$$S_{ij} = \begin{cases} 1 & \text{switch } S_{ij} \text{is close} \\ 0 & \text{switch } S_{ij} \text{is open} \end{cases} i = \{a, b, c\} j = \{A, B, C, D, E, F\} \tag{26}$$

The output voltage can be expressed as

$$V_o(t) = \begin{bmatrix} V_A(t) \\ V_B(t) \\ V_C(t) \\ V_D(t) \\ V_E(t) \\ V_F(t) \end{bmatrix} = \begin{bmatrix} V_o \sin(\omega_o t) \\ V_o \sin(\omega_o t - 2pi/3) \\ V_o \sin(\omega_o t - 4pi/3) \\ V_o \sin(\omega_o t - pi/6) \\ V_o \sin(\omega_o t - 5pi/6) \\ V_o \sin(\omega_o t - 3pi/2) \end{bmatrix} \tag{27}$$

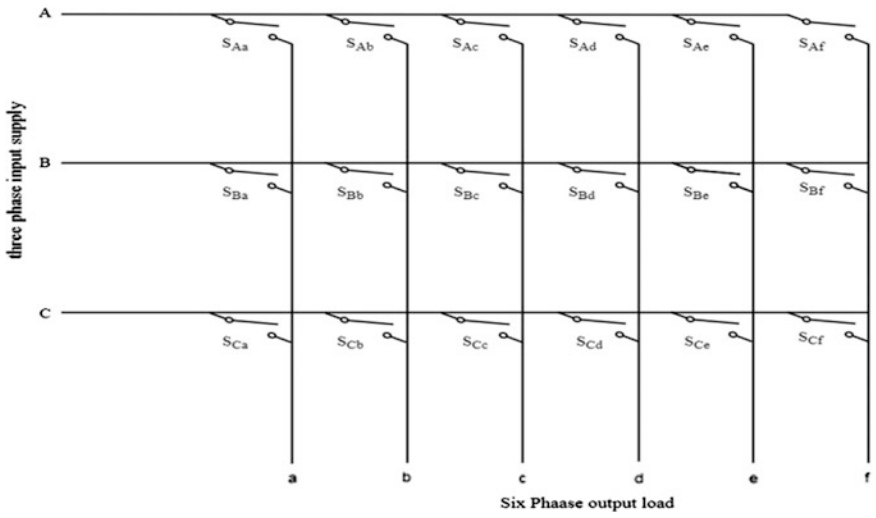


Fig. 2 Six-phase matrix converter topology

The output current can be expressed as

$$I_o(t) = \begin{bmatrix} I_A(t) \\ I_B(t) \\ I_C(t) \\ I_D(t) \\ I_E(t) \\ I_F(t) \end{bmatrix} = \begin{bmatrix} I_o \sin(\omega_o t - \varnothing_o) \\ I_o \sin(\omega_o t - 2pi/3 - \varnothing_o) \\ I_o \sin(\omega_o t - 4pi/3 - \varnothing_o) \\ I_o \sin(\omega_o t - pi/6 - \varnothing_o) \\ I_o \sin(\omega_o t - 5pi/6 - \varnothing_o) \\ I_o \sin(\omega_o t - 3pi/2 - \varnothing_o) \end{bmatrix} \tag{28}$$

Where $\varnothing_o$ is output power factor angle
The input voltage can be expressed as

$$V_i(t) = \begin{bmatrix} V_a(t) \\ V_b(t) \\ V_c(t) \end{bmatrix} = \begin{bmatrix} V_a \sin(\omega_i t) \\ V_b \sin(\omega_i t - 2pi/3) \\ V_c \sin(\omega_i t - 4pi/3) \end{bmatrix} \tag{29}$$

The input current can be expressed as

$$I_i(t) = \begin{bmatrix} I_a(t) \\ I_b(t) \\ I_c(t) \end{bmatrix} = \begin{bmatrix} I_a \sin(\omega_i t) \\ I_b \sin(\omega_i t - 2pi/3) \\ I_c \sin(\omega_i t - 4pi/3) \end{bmatrix} \tag{30}$$

$$V_o(t) = M(t) V_i(t) \tag{31}$$

$$I_i(t) = M(t)^T I_o(t) \tag{32}$$

Where

$$M(t) = \begin{bmatrix} m_{aA}(t) & m_{bA}(t) & m_{cA}(t) \\ m_{aB}(t) & m_{bB}(t) & m_{cB}(t) \\ m_{aC}(t) & m_{bC}(t) & m_{cC}(t) \\ m_{aD}(t) & m_{bD}(t) & m_{cD}(t) \\ m_{aE}(t) & m_{bE}(t) & m_{cE}(t) \\ m_{aF}(t) & m_{bF}(t) & m_{cF}(t) \end{bmatrix} \tag{33}$$

$$S_{aj} + S_{bj} + S_{cj} = 1 \;\; j = \{A, B, C, D, E, F\} \tag{34}$$

4 Indirect Space Vector Modulation for SPMC

The SPMC can be equivalently represented as the combination of current source rectifier and 6-leg voltage source inverter, which are virtually connected by a dc-link. The switching function is divided into the product of rectifier switching and the inverter switching. The switching function realization is expressed in (35)

$$\begin{bmatrix} S_{Aa} & S_{Ba} & S_{Ca} & S_{Da} & S_{Ea} & S_{Fa} \\ S_{Ab} & S_{Bb} & S_{Cb} & S_{Db} & S_{Eb} & S_{Fb} \\ S_{Ac} & S_{Bc} & S_{Cc} & S_{Dc} & S_{Ec} & S_{Fc} \end{bmatrix} = \begin{bmatrix} S_7 & S_8 \\ S_9 & S_{10} \\ S_{11} & S_{12} \\ S_{13} & S_{14} \\ S_{15} & S_{16} \\ S_{17} & S_{18} \end{bmatrix} \begin{bmatrix} S_1 & S_3 & S_5 \\ S_2 & S_4 & S_6 \end{bmatrix} \tag{35}$$

5 Particle Swarm Optimization Based ISVM

The particle swarm optimization (PSO) is applied to optimize the switching state in this paper. This optimization technique is preferred because of the short code and fast convergence speed. Switching selection should be completed within 50 μs because the switching frequency of the SPMC is taken as 20 kHz. The PSO code is computation time not more than the time taken for switching selection time. The PSO code should take the value of output voltage simultaneously. Because of the above reasons, the PSO code should be short and have fast convergence speed. The weight factor W is taken as 0.4–0.5. If the weight factor is high about 0.9–1, the global optimization capability increases. If the weight factor is about 0.4–0.5, the local optimization capability increases. C_1, C_2 and R_1, R_2 are the acceleration factor and random numbers, respectively. Depending on the cost function of the switching state, the optimum switching state will be applied to the MC; in the simulation every time MC model should run to determine the output voltage THD, this THD value is the cost function of that particular switching state. Figure 3 shows the detailed flowchart of the proposed methodology. The PSO equations for updating the velocity and position of the each particle are (36, 37)

$$V_i^{t+1} = W*V_i^t + C_1R_1\left(Pbest_i^t - x_i^t\right) + C_2R_2\left(Gbest_i^t - x_i^t\right) \tag{36}$$

$$x_i^{t+1} = x_i^t + V_i^{t+1} \tag{37}$$

6 Simulation Results

The ISVM and PSO-based ISVM techniques were tested on the same test system. The simulation was carried out on SPIM fed by SPMC. The performance of the SPIM is analyzed by keeping the supply voltage constant and varying the load torque. Figure 4 shows the speed response of the test system under no-load

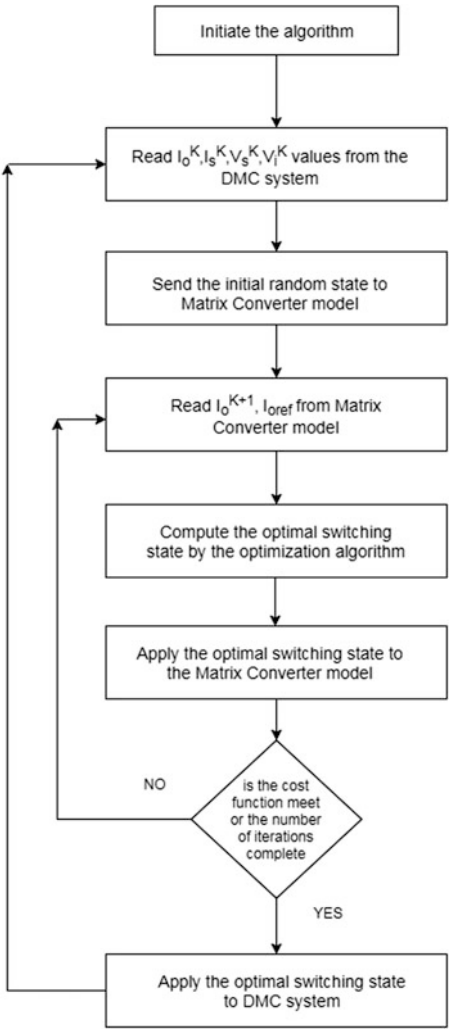


Fig. 3 Flowchart representation of the PSO algorithm

condition. Figure 5 shows the torque response under no-load condition. Figure 6 shows the output voltages of the SPMC operated with PSO-based ISVM. Figures 7 and 8 show the FFT analysis of SPMC operated with ISVM and PSO-based ISVM, respectively.

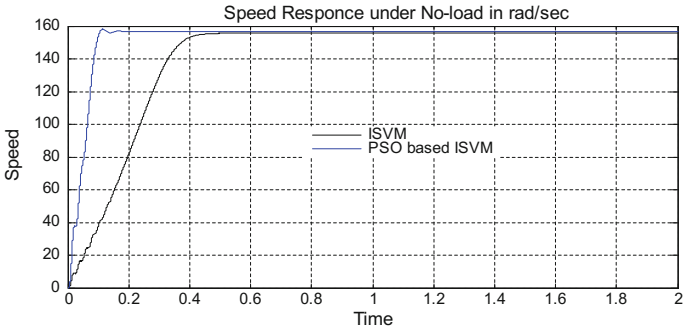


Fig. 4 Speed response under no-load condition

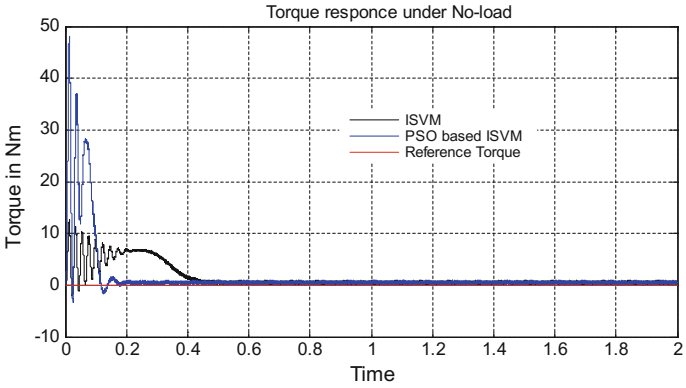


Fig. 5 Torque response under no-load condition

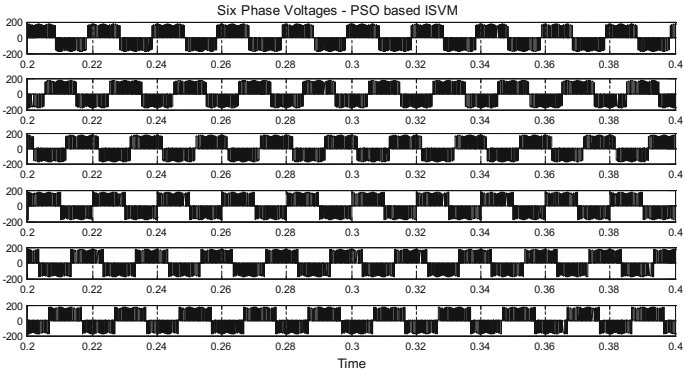


Fig. 6 Six-phase output voltages (phase to phase) operated with PSO-based ISVM

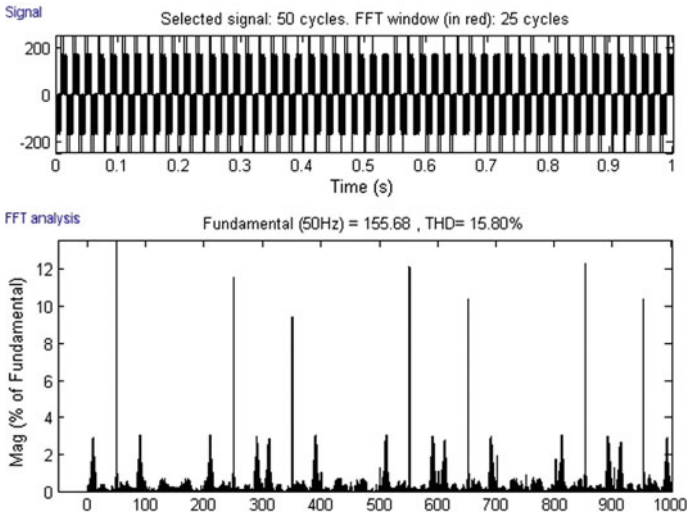


Fig. 7 FFT analysis of output voltage of SPMC operated with ISVM

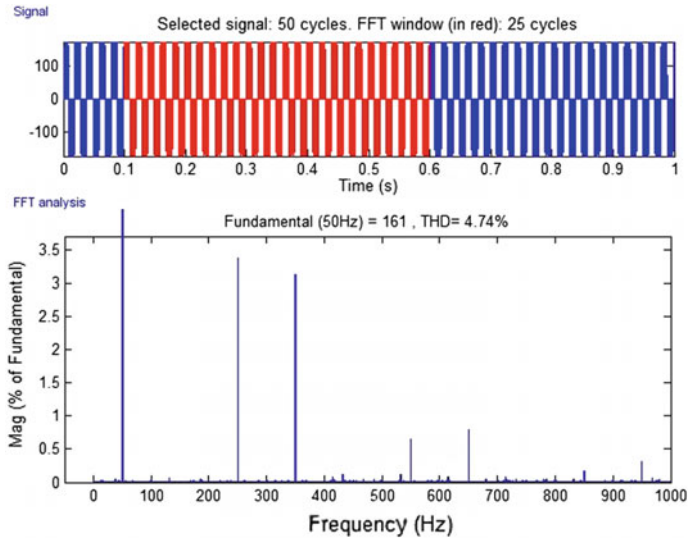


Fig. 8 FFT analysis of output voltage of SPMC operated with PSO-based ISVM

7 Conclusion

The MC is the best alternative solution for conversion of AC–AC. The MC is facing challenges like complex modulation techniques for implementation. Using conventional and easy techniques, the value of THD will be high. The two techniques ISVM and PSO-based ISVM were implemented successfully. The results of both the methods were compared. The method which was proposed to reduce the harmonic has the disadvantage that it is complex to implement, the overall complexity of the modulation technique may increase, but when compared with the amount of decrease in the percentage, THD is high. With conventional ISVM, the output voltage percentage THD is 15.80%. The proposed technique output voltage percentage THD is 4.74%.

References

1. Casadei, D., Serra, G., Tani, A., Zarri, L.: Matrix converter modulation strategies: a new general approach based on space vector representation of the switch state. IEEE Trans. Ind. Electron. **49**(2), 370–381 (2007)
2. Venturini, M., Alesina, A.: Generalised transformer: a new bidirectional, sinusoidal waveform frequency converter with continuously adjustable input power factor. In: The IEEE Power Electronics Specialists Conference, Atlanta, GA, USA, pp. 242–252 (1980)
3. Alesina, A., Venturini, M.G.B.: Analysis and design of optimum-amplitude nine-switch direct AC-AC converters. IEEE Trans. Power Electron. **4**, 101–112 (1981)
4. Huber, L., Borojevic, D.: Space vector modulated three phase to three-phase matrix converter with input power factor correction. IEEE Trans. Ind. Appl. **31**(6), 1234–1246 (1995)
5. Amarendra, Ch., Reddy, K.H.: Investigation and analysis of space vector modulation with matrix converter determined based on fuzzy C-means tuned modulation index. IJECE **6**(5), 1939–1947 (2016)
6. Wheeler, P.W., Rodriguez, J., Clare, J.C., Empringham, L.: Matrix converter—a technology review. IEEE Trans. Ind. Electron. **49**(2), 275–288 (2002)
7. Rodas, J., Gregor, R., Takase, Y., Gregor, D., Franco, D.: Multimodular matrix converter topology applied to the six-phase wind energy generator. In: Power Engineering Conference (UPEC), pp. 1–6 (2015)
8. Bachir, G., Bendiabdellah, A.: Scalar control for six phase matrix converter fed double star induction motor. Adv. Electr. Comput. Eng. **10**,121–126 (2010)
9. Iqbal, A., Rahman, K., Alammari, R.: Space vector PWM for a three-phase to six phase direct AC/AC converter. In: Industrial Technology (ICIT), pp. 1–6 (2015)
10. Omrani, K., Dami, M.A., Jemli, M.: Three to six-phase direct matrix converters. In: Renewable Energy Congress (IREC), pp. 1–5 (2016)
11. Martin, J.-P., Semail, E., Pierfederici, S., Bouscayrol, A., Meibody-Tabar. F., Davat, B.: Space vector control of 5-phase PMSM supplied by 5 H-bridge VSIs. In: Conference on Modeling and Simulation of Electric Machines, Converters and Systems (ElectrIMACS'02), Montreal, Canada (2002)
12. Aware, M.V.: Six-phase inverter operation with space vector pulse-width modulation for a symmetrical single neutral load. Electr. Power Compon. Syst. **41**(16), 1635–1653 (2013)

Spam Mail Detection Using Data Mining: A Comparative Analysis

Soumyabrata Saha, Suparna DasGupta and Suman Kumar Das

Abstract In the era of digitization communication, the commercial transaction takes place through the Web, email and be one of the most authoritative and fastest forms of communication, email fame has led to the scratchy email spam upload. The ensuing increase in superfluous and unsolicited spam received through email not only increases network communication and memory space, but it also becomes severe security intimidation for the end user. Automatic spam filtering is a promising and worthy research area where extensive works have been reported for the cataloging of email spam, but none of the methodologies guarantees complete solutions. Due to the rapid expansion of digital data, knowledge discovery and data mining have engrossed much attention with an imminent need to turn that data into useful information and knowledge. In this paper, authors have focused on how email communications are affected by spam and focus on various classification-based data mining techniques in a spam data set to identify spam and ham to analyze the performance of all classifiers and identify the best classifiers in terms of performance. To carry out the purpose of the work, an open source WEKA data mining tool has used to explore the performance analysis of the different classifiers and finally the superlative classifier has identified for the classification of email spam and has developed the knowledge flow model.

S. Saha (✉) · S. DasGupta · S. K. Das
Department of Information Technology, JIS College of Engineering, Kalyani, West Bengal, India
e-mail: som.brata@gmail.com

S. DasGupta
e-mail: suparnadasguptait@gmail.com

S. K. Das
e-mail: itsmesuman999@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_56

1 Introduction

Evolving from a small exasperation to a foremost crisis, given the excessive circulating volume and the already apposite content of some of the communications, the spam is to diminish the reliability of the electronic mail. Email spam is referred as spam mail or junk mail in which unsolicited emails are sent to waste network bandwidth and the delivery of legitimate emails are also affected by large amount of spam traffic. Spam can include malicious software such as scripts or other executable attachments and one of the major factors that influence the environment and that is uncomfortable for email users. Spam filters are highly looked for to automatically filter spam to clean mailboxes. Initial spam filters require a user to create regulations primarily by looking at patterns in a typical junk email, such as the presence of specific words, combinations of words, phrases. To ignore these rules, spammers used content from obfuscation techniques. A business model that depends on the requested marketing mail is not particularly noticeable because the sender is much less, tends to a lot of messages too, this aggressive behavior is probably the most characteristic main spammers.

Data mining is the new predominant technology that has engrossed lot of attention in academics, business and as well as industrial research. Due to colossal accessibility of large amount of data and the requirement for whirling such data into functional information and knowledge, data mining techniques have widely used. Data mining have been used for extraction of concealed information from large databases and known as knowledge discovery in databases. Researchers have investigated that data mining techniques can be used to envisage future trends and activities, allowing business to make upbeat, knowledge-driven decisions. Mining tools and techniques can respond business to questions that conventionally were too time intensive to resolve. The results obtained from data mining are essentially used for making analysis and prediction. There are different techniques that have used in data mining, such as association, clustering, classification, prediction, outlier, detection, and regression.

In this work, authors have used Naive Bayes—a Bayesian-based classification, SMO—a functional-based classification, J48 and random forest—a tree-based classification technique in order to classify the spam mail. This paper presents a brief review of literature carried out in the area of data mining for the betterment of research works in this area. The remnant of the article is organized as follows. Comprehensive surveys of related works are discussed in Sect. 2. In Sect. 3, proposed methodology has been presented. Performance analysis has been figure out in Sect. 4. The final conclusion is offered in Sect. 5.

2 Related Works

There are many researches and their experimental work carried out using diverse data mining techniques for the uncovering of superfluous emails. This section provides an analysis of available literature of some algorithms employed by different researchers to utilize various techniques. In [1], machine learning techniques have been applied to the email rundown where the RIPPER classifier used the sentence resolution that would be included in a summary. It involves positive examples in huge numbers, and machine learning technique is further varied and extensively used for the cataloging of spam. Bayesian [2] networks located as the most preferred practice for spam detection, but it is quite intricate to swell on many elements to come out with the trial. In [3], the author formalizes a problem of grouping the collection of spam messages through the criterion function. A genetic algorithm along with penalty function has introduced to unravel the clustering problem. The extraction of symbiotic data is a dispersed mining strategy that unifies filtering based on the content with the collaborative filtering [4]. The objective is to use local filters to hold up custom filtering in the context of privacy. In [5], the email classifiers canted again in the feeding method, the neural propagation communities, and the Bayesian classifiers are evaluated. In [6], for detection of intrusions and Web-based attacks of and all activities as ID3 is a classification algorithm used to construct a decision tree attacks and intrusions. In [7], the authors have proposed a hybrid SMS classification system to detect spam or ham, using the Naïve Bayes algorithm and Apriori algorithm. In [8], authors have offered the classification which is a significant technique of data mining with broad applications to classify the various types of data used in everyday life. Symbiotic Data Mining [9] that unifies content based filtering with collaborative filtering. In [10], authors evaluated the efficacy of email classifiers based on the neural network of forward feedback propagation and Bayesian classifiers. A totally Bayesian approach for grouping and classifying smoothness using mixed membership models based on four-level assumptions was implemented in [11].

In this paper, authors have chosen classification technique as other data mining techniques like clustering, whereas association is not capable of prediction related issues. Clustering is capable of partitioning the related data elements in same set, and on the other hand, association is generally used for establishing relationships between attributes that exists in a data set.

3 Proposed Methodology

The overall proposed system architecture is illustrated in Fig. 1.

- **Data Collection**: Data were donated by George Foreman, and the collected data set is used in UCI machine learning repository [13].

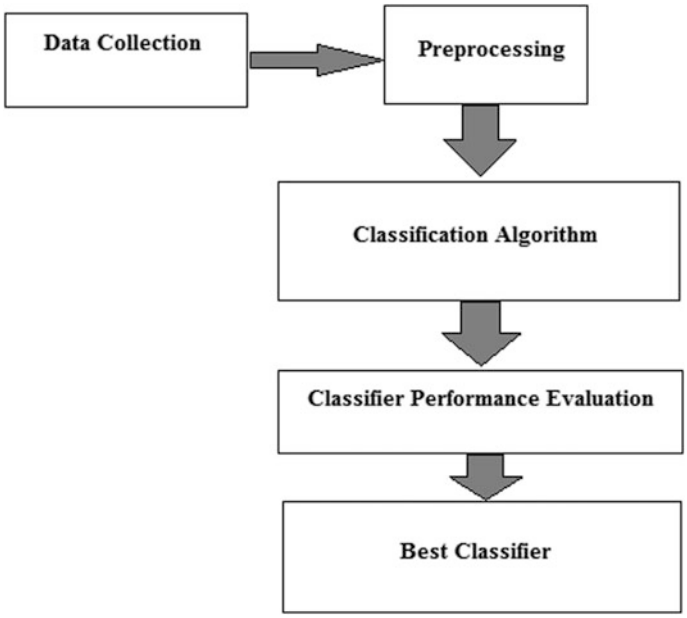


Fig. 1 Proposed system architecture

- **Data Pre-processing**: In the real-world data set, which consists of many mistakes, they are cleaned and removed in order to have accurate results of the data sets. In this step data set it is transformed and integrated into an appropriate format before classifiers are applied in the data set. The file email-spam.arff has properly processed before classifiers are applied on it.
- **Applying Algorithm**: After having the pre-processed file, all the classification algorithms, namely Naive Bayes, SMO, J48 decision tree, and random forest, have applied in order to find features on the basis of which spam being identified.
- **Performance Evaluation**: After applying all classifiers, each of them was evaluated on the basis of performance metrics in order to figure out the best classifier.

Description of data set: The data set named email-spam.arff has 4601 instances and 58 attributes. The data set contains 57 continuous and 1 nominal class label attribute. Description of all attributes is mentioned below in Table 1.

Tools and techniques: For implementing the required task, authors have used WEKA tool which is an open-source tool for data mining and the classifiers used in this work are: (i) Naïve Bayes, (ii) SMO, (iii) J48, (iv) random forest.

Table 1 Attribute description of used data set

Attribute	Attribute type	Description
A1–A48 (real)	char_freq_CHAR	Characters in the email that matches characters lists in percentage
A49–A54 (REAL)	capital_run_length_average	Average length of uninterrupted sequences of capital letters
A55 (real)	capital_run_length_longest	Length of longest uninterrupted sequence of capital letters
A56 (integer)	capital_run_length_longest	Length of longest uninterrupted sequence of capital letters
A57 (integer)	capital_run_length_total	Total number of capital letters in the email
A58 (0, 1)	Class label	1 for spam, 0 for ham

4 Result Analysis

The data set used in this work is tested and analyzed with four different classification techniques that use cross-validation which are the following: (i) Naïve Bayes, (ii) SMO, (iii) J48, (iv) random forest. After applying all the classifiers, the performance of each classifier has been analyzed on the basis of various performance metrics as: true-positive rate, false-positive rate, precision, recall, F-measure, ROC area, time taken to build classifier model. Figures 2, 3, 4, and 5 shows output of Naive Bayes, SMO, J48, and random forest, respectively.

All the statistics results are provided in Table 2, and the statistical results obtained are plotted graphically in Fig. 6.

Accuracy for each algorithm has been calculated based on the obtained confusion matrix generated individually on WEKA workbench as shown in Figs. 2, 3, 4, and 5 for Naive Bayes, SMO, J48 decision tree, random forest decision tree, respectively. Typical representation of confusion matrix and how accuracy is calculated from it is shown below in Fig. 7.

From Fig. 2 (for Naive Bayes), authors have true-positives (TPs) as 1923 and false-positives (FPs) as 1725 and total number of instances is 4601, so accuracy in terms of confusion matrix is

```
=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.690    0.049    0.956      0.690    0.801      0.632    0.936     0.941     0
                0.951    0.310    0.666      0.951    0.784      0.632    0.939     0.892     1
Weighted Avg.   0.793    0.152    0.842      0.793    0.794      0.632    0.937     0.922

=== Confusion Matrix ===

    a    b   <-- classified as
 1923  865 |    a = 0
   88 1725 |    b = 1
```

Fig. 2 Output obtained by Naive Bayes

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.952    0.169    0.896      0.952    0.923      0.798    0.891     0.882     0
                 0.831    0.048    0.918      0.831    0.872      0.798    0.891     0.830     1
Weighted Avg.    0.904    0.122    0.905      0.904    0.903      0.798    0.891     0.862

=== Confusion Matrix ===

    a    b   <-- classified as
 2654  134 |    a = 0
  307 1506 |    b = 1
```

Fig. 3 Output obtained by SMO

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.944    0.092    0.940      0.944    0.942      0.853    0.939     0.945     0
                 0.908    0.056    0.913      0.908    0.911      0.853    0.939     0.874     1
Weighted Avg.    0.930    0.078    0.930      0.930    0.930      0.853    0.939     0.917

=== Confusion Matrix ===

    a    b   <-- classified as
 2632  156 |    a = 0
  167 1646 |    b = 1
```

Fig. 4 Output obtained by J48

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.972    0.071    0.955      0.972    0.963      0.906    0.987     0.989     0
                 0.929    0.028    0.955      0.929    0.942      0.906    0.987     0.984     1
Weighted Avg.    0.955    0.054    0.955      0.955    0.955      0.906    0.987     0.987

=== Confusion Matrix ===

    a    b   <-- classified as
 2709   79 |    a = 0
  128 1685 |    b = 1
```

Fig. 5 Output obtained by random forest

- ACCURACY = (1923 + 1725)/4601 = 0.7928 or 79.28% ——————(i)
  From Fig. 3 (for SMO), authors have true-positives (TPs) as 2654 and false-positives (FPs) as 1506 and total number of instances is 4601, so accuracy in terms of confusion matrix is
- ACCURACY = (2654 + 1506)/4601 = 0.9041 or 90.41% ——————(ii)
  From Fig. 4 (for J48) authors have true-positives (TPs) as 2632 and false-positives (FPs) as 1646 and total number of instances is 4601, so accuracy in terms of confusion matrix is
- ACCURACY = (2632 + 1646)/4601 = 0.9297 or 92.97% ——————(iii)

Table 2 Statistical analysis of classifier using cross validation

Metrics	Class	Used algorithm			
		Naive Bayes	SMO	J48	Random forest
TP rate	0	0.69	0.952	0.944	0.974
	1	0.951	0.831	0.908	0.906
FP rate	0	0.049	0.169	0.092	0.094
	1	0.31	0.048	0.056	0.026
Precision	0	0.956	0.896	0.94	0.941
	1	0.666	0.918	0.913	0.958
Recall	0	0.69	0.952	0.944	0.974
	1	0.951	0.831	0.908	0.906
F-measure	0	0.801	0.923	0.942	0.957
	1	0.784	0.872	0.911	0.931
ROC area	0	0.936	0.891	0.939	0.982
	1	0.939	0.891	0.939	0.982

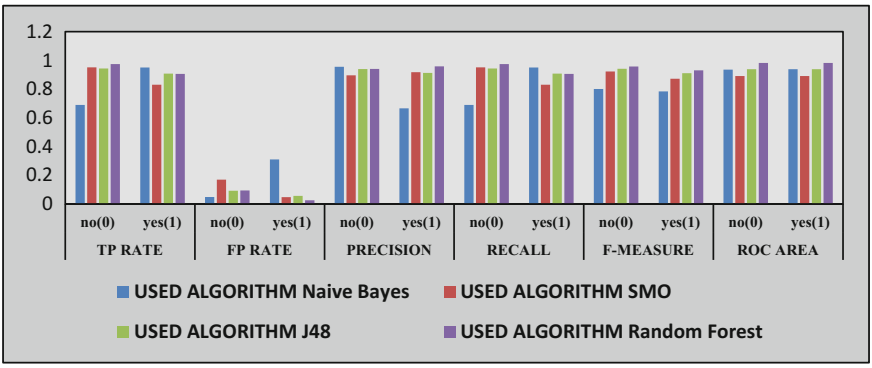


Fig. 6 Graphical representation of statistical analysis of classifier using cross validation (X-axis represents various performance metrics, and Y-axis represents values)

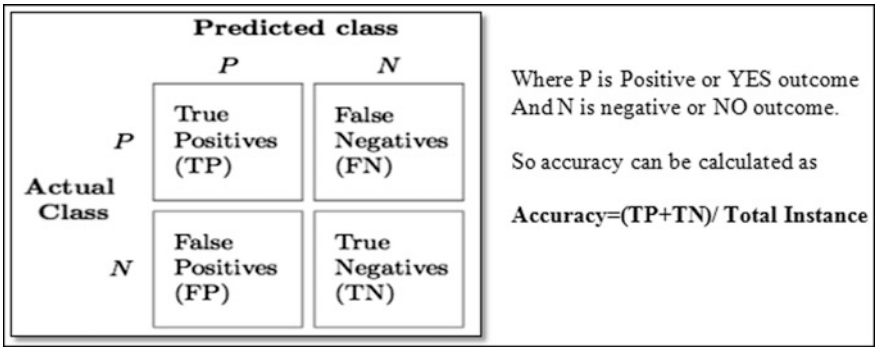


Fig. 7 Representation of confusion matrix

MINING TECHNIQUE	ACCURACY in%
Naive	79.28
SMO	90.41
J48	92.97
Random Forest	95.5

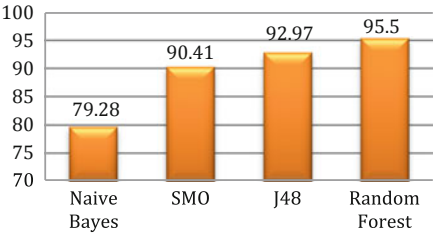


Fig. 8 Graphical representation of accuracy of all classifiers (X-axis represents various classifiers, and Y-axis represents accuracy in percentage)

From Fig. 5(for random forest), authors have true-positives (TPs) as 2709 and false-positives (FPs) as 1642 and total number of instances is 4601, so accuracy in terms of confusion matrix is

- ACCURACY = (2709 + 1685)/4601 = 0.9550 or 95.50% ——————(iv)

From (i), (ii), (iii), (iv), authors have compared accuracy (in %). Figure 8 shows graphical representation of the same.

From this comparison, it has observed that random forest performs best in terms of accuracy and is having accuracy of 95.50%. After executing the comparison of all classifiers on the basis of statistical analysis and accuracy, authors would compare them on the basis of time requirement (in seconds) to build the classifier model. It has found that Naive Bayes algorithm performs best as compared to other classifiers, and it takes 0.25 s to build the classifier model (Fig. 9).

After having all the comparison of classifiers on the basis of various performance metrics, a knowledge flow model is shown in Fig. 10, which shows membership

Algorithm	Naive Bayes	SMO	J48	Random Forest
Time in Sec	0.25	0.66	1.59	3.11

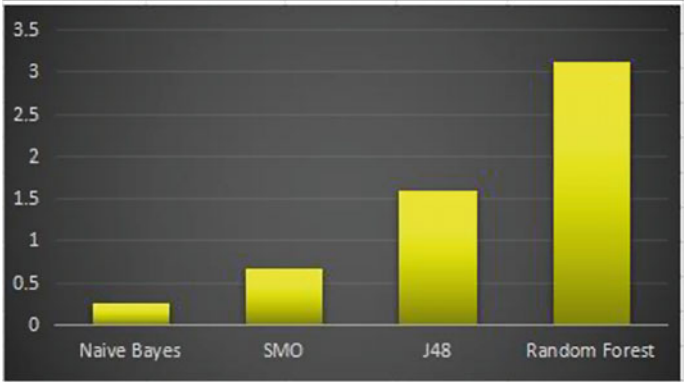


Fig. 9 Graphical representation of all classifiers on the basis of time requirements (X-axis represents used classifiers and Y-axis represents time taken to build classifiers)

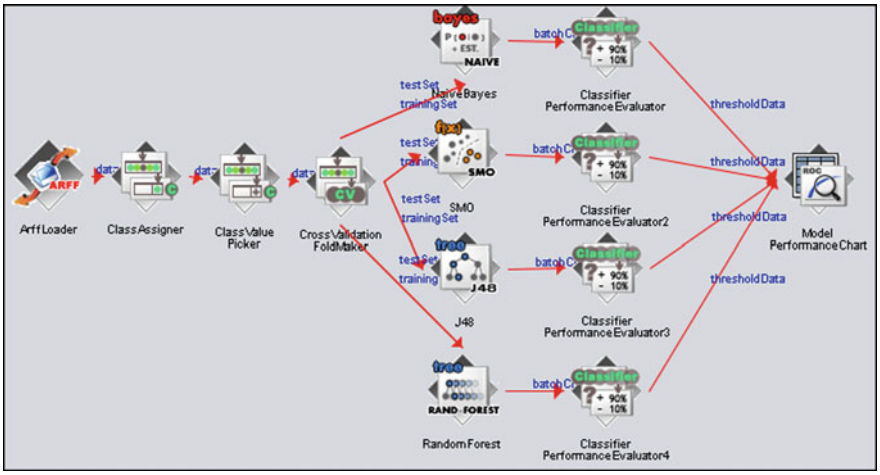


Fig. 10 Knowledge flow layout

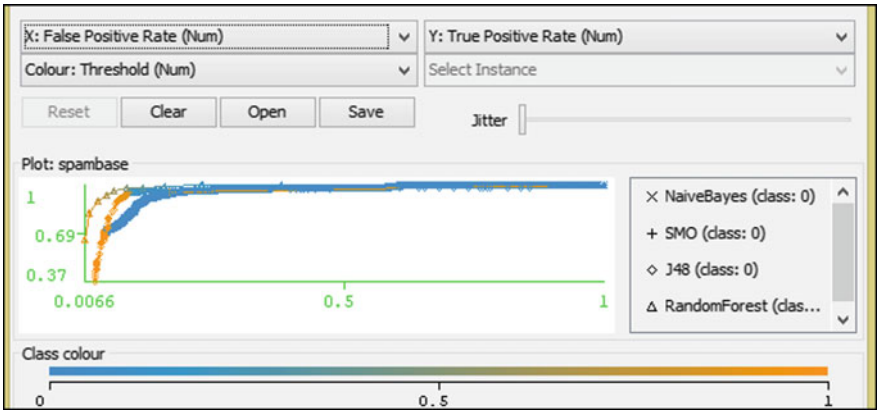


Fig. 11 Model performance chart: ROC curve of classifiers

trees structure. Authors have loaded the data set (email-spam.arff file) using Arff Loader functionality, and after running the entire model, authors have obtained a model performance chart. The obtained model performance chart is shown in Fig. 11 which shows the region of convergence (ROC) curve for all the classifiers that are used in our work, and it has identified that random forest classifier covers more region with its curve that is close to value 1.

5 Conclusion

In this paper, mail spam is classified using four sorting techniques, namely; Naive Bayes, SMO, J48, and random forest in the data set contain 4601 instances. The system architecture has been proposed to classify spam mail. Then, each classifier has analyzed and compared based on their performance. It has been observed that the random forest technique has the highest accuracy, highest weighted precision, highest weighted recall, and highest weighted F-measure of 95.50%, with high region under curve that is ROC value close to 1, but if authors consider the performance on the basis of time or execution time metrics in that case Naive Bayes performs best. Boosting the accuracy further close to 100% and minimizing the time requirements to build the classifier model of each classifier and implementing this technique in Gmail, Yahoo mail, and other email domain is put under future scope.

References

1. Ducheneaut, N., Bellotti, V.: E-mail as habitat: an exploration of embedded personal information management. Interactions ACM **5**, 30–38 (2001)
2. Sahami, M., Dumasi, S., Heckerman, D., Horvitz, E.: A Bayesian approach to filtering junk e-mail. In: Learning for Text Categorization. Workshop, Madison, Wisconsin (1998)
3. Perkins, A.: The classification of search engine spam (2001). http://www.ebrandmanagement.com/whitepapers/spam-classification
4. Paulo, C., Clotilde, L., Pedro, S.: Symniotic data mining for personalized spam filtering. In: The Web Intelligence and Intelligent Agent Technology, pp. 149–156 (2009)
5. Rasim, M.A., Ramiz, M.A., Saadat, A.N.: Classification of textual e-mail spam using data mining techniques. J. Appl. Comput. Intell. Soft Comput. (2011)
6. Patil, S.S., Deepak Kapgate D., "A Review on Detection of Web Based Attacks using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, No.12, December 2013
7. Ahmed, I., Guan D., "SMS Classification Based on Naïve Bayes Classifier and Apriori Algorithm Frequent Item Set", International Journal of Machine Learning and Computing, Volume 4, No.2, April 2014
8. Patil, T.R., Sherekar, S.S.: Performance analysis of Naive Bayes and J48 classification algorithm for data classification. Int. J. Comput. Sci. Appl. **6**(2) (2014)
9. Paulo, C., Clotilde, L., Pedro, S.: Symbiotic data mining for personalized spam filtering. In: Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology. (IEEE/WIC/ACM), pp. 149–156 (2009)
10. Yang, Y., Elfayoumy, S.: Anti-spam filtering using neural networks and Bayesian classifiers. In: Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation, Jacksonville, FL, USA, 20–23 June 2007
11. Erosheva, E.A., Fienberg, S.E.: Bayesian mixed membership models for soft clustering and classification. Manuscript (2004)
12. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/spambase. Accessed 10 Sep 2017
13. Erosheva, E.A., Fienberg, S.E.: Bayesian mixed membership models for soft clustering and classification. Proc. Natl. Acad. Sci. **97**(22), 11885–11892 (2004)

Selection of Battery Size by Using Power Flow Decision Program for Microgrids

Ch. Padmanabha Raju and D. Dhanalakshmi

Abstract This paper concentrates on determining the size of energy storage system (ESS) for microgrid (MG). Renewable energies are intermittent, and their productions are stochastic and can cause insufficiency in supply of electrical systems. Applying an ESS can alleviate impact of renewable energies. ESS acting a main function in the MG, it is advantageous to shave the high demand and store the excess energy. Sizing of ESS can be viewed as primary while considering ESS in the MG. The majority of MGs work in grid-connected mode. ESS is used to store additional power created by sustainable power sources for the duration of low demand period and redispatching during high demand period. To determine size of an ESS, the aim is to minimize total annual operating cost, and this can be accomplished by using power flow decision.

1 Introduction

In future, microgrids will assume a primary part in the power systems because of energy and environmental crises, for example, exhaustion of non-renewable energy source assets, the growth of power demand, and emission of greenhouse gases. To answer such crises, an idea of microgrid, which is formed by distributed power resources, e.g., PV microsource, wind turbine lower scale source, fuel cells, and energy storage space system, should be considered one of the solution. Since renewable power sources are discontinuous, their productions are stochastic and can cause the deficiency in supply of electrical systems. Applying an energy storage system can alleviate the effect of renewable power sources, forecast error on power system performance, and increase system tolerance against the inadequacy of

Ch. P. Raju · D. Dhanalakshmi (✉)
Department of Electrical and Electronics Engineering,
P.V.P. Siddhartha Institute of Technology, Kanuru, Vijayawada 520007, India
e-mail: dhana.le203@gmail.com

Ch.P. Raju
e-mail: pnraju78@yahoo.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_57

supply. Battery banks (BBs) play an important role since they are efficient to shave the peak demand, store the surplus energy, and increase the reliability and security of power systems. Because of the effective role of BBs, their sizing is necessary for enhancing the efficiency of the MG performance. The utility network-connected photovoltaic system is a power producing system, with the intention of using a PV as the basis of power generation, along with working synchronously with the utility grid. Such a system could incorporate ESS and other distributed power resources in support of providing supply to nearby loads for the period of network blackouts and peak load hours [1]. There are three techniques for operation of an energy storage system.

1. ESS charge condition: At what time the demand is small, PV as well as utility grid will charge the energy system.
2. ESS idle condition: PV supply nearby loads at a specific hour at what time demand is low, both photovoltaic (PV) and load requirement be high, but PV is not sufficient to meet the load at the time ESS discharge is up to its maximum after its idle state.
3. ESS discharging condition: Both photovoltaic and ESS provide power during high demand periods at particular hours; energy storage system supplies power to the high demand period if PV is insufficient and utility grid power charge is high.

The real difficulties for photovoltaic system stay here in coordinating irregular power generation with the dynamic power demand. An answer is to put ESS component to these nonconventional and irregular power sources. For this situation, the hybrid system, made out of a PV generator, nearby loads, ESS, and the grid, can do numerous applications. In a utility-related photovoltaic system with ESS, if the energy is not adequate to take care of the load demand, the rest of the power will be provided by the grid. Then again, when the energy is produced more than the demand, the surplus power is sent to the grid [2, 3].

Various potential advantages can be obtained from the utilization of ESS in systems. They change the power purchase from high demand periods toward low demand periods, decrease the power charge, delay the investment on placing new transmission lines, prevent blackouts when demand goes beyond a basic point, decrease the losses, prevent need for continuous and adaptable supply, and efficiently shave the peak demand [4].

1.1 Basic Model Design

In this, ESS is connected with the AC bus through the inverter/charger unit, and the PV output is also sent through the inverter to change the DC into AC as shown in Fig. 1. The PV modules, utility grid, ESS, and the load are connected with AC bus bar. A PV inverter has been used to change the DC of the PV modules into AC

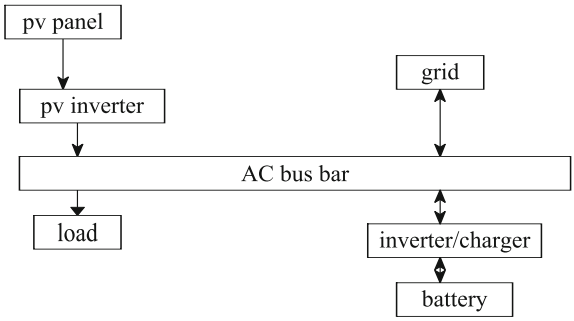


Fig. 1 Basic system configuration

output at the AC bus bar. ESS works on DC conditions, thusly an inverter/charger unit has been utilized among battery at AC bust bar to change AC to DC and DC to AC, while charge and discharge of the battery [3].

1.2 Photovoltaic (PV) System Configuration

Photovoltaic (PV) be a specific name utilized for power produced from the sunlight. A PV system converts the sunlight into electricity [5–10]. The DC output produced from PV an area Apvg (m2) at a solar radiation on tilted module plane Gt (W/m2) that can be given by Eq. (1).

$$E_{pv\_dc} = \eta_{pvg} \times A_{pvg} \times G_t \times f_{man} \times f_{dirt} \tag{1}$$

G_t Represents radiation of solar panel at tilted module plane at hour t, f_{man}, f_{dirt} represents derating factors, and η_{pvg} represents the PV generator efficiency. The AC output from the PV can be written as shown in Eq. (2)

$$E_{pv\_ac}(t) = E_{pv\_dc}(t) \times \eta_{inv} \tag{2}$$

1.3 Energy Storage System Modeling

By means of continuously growing energy requirement, in future, more expansion of renewable power, circulated power, and smart grid and the request for raising power storage systems will keep on increase. The usage of energy storage system reduces cost of electricity purchase and improves reliability of the power system. The state of charge of the battery is updated each hour, and charge and discharge equations for battery banks (BBs) are written in the following equations [7,8, 9].

The charge and discharge power transfer to and from the battery during specific sampling hour can be written as given in Eqs. (3)–(6)

$$E_{DC,\,bat}(d.t) = \frac{E_{bat}(d,t) - E_{bat}(d,t - \text{Sampling Time Interval})}{\text{Sampling Time Interval}} \tag{3}$$

The battery DC power is converted into AC power when it discharges through the inverter/charger unit to the AC side. AC power at the AC bus is converted into DC when it passes through the inverter/charger unit. The conversion efficiency of the inverter η_{bat} is assumed to be constant while charging and discharging the battery.

$$E_{AC,\,bat}(d.t) = \begin{cases} \eta\,\text{bat}\ \text{EDC, bat}(d,t) & E_{DC,\,bat}(d,t) < 0 \\ \dfrac{E_{DC,\,bat}(d,t)}{\eta\,\text{bat}} & E_{DC,\,bat}(d,t) > 0 \end{cases} \tag{4}$$

Charging

$$\text{SOC}(d,t) = \text{SOC}(d,t - \text{sampling time interval}) \times (1\,a) + \frac{\eta_{char} E_{DC,\,bat}(d,t)}{\text{E\_battery\_present\_kwh}(d,t)} \tag{5}$$

Discharging

$$\text{SOC}(d,t) = \text{SOC}(d,t - \text{sampling time interval}) \times (1\,a)\ \frac{E_{DC,\,bat}(d,t)}{\eta\text{disch} \times \text{E\_battery\_present\_kwh(day, t)}} \tag{6}$$

During charge and discharge process of battery storage system, it can lose some amount of energy, i.e., cumulative battery capacity loss, and it can be calculated as given in Eq. (7)

$$E_{cumi,\,bat,\,loss}(d,t) = \begin{cases} E_{cumi,\,bat,\,loss}(d,t-1) - Z \times E_{DC,\,bat}(d,t\ -1) & E_{DC,\,bat}(d,t) < 0 \\ E_{cumi,\,bat,\,loss}(d,t-1) & E_{DC,\,bat}(d,t) > 0 \end{cases} \tag{7}$$

ESS capacity loss during any sampling time can be found by using Eq. (8)

$$\text{BCL}(d,t) = E_{cumi,\,bat,\,loss}(t) - E_{cumi,\,bat,\,loss}(t-1) \tag{8}$$

Usable battery capacity for the next hour in KWh is given in Eq. (9)

$$C_{bat}(d,t) = E_{bat}, \text{nominal KWh} - E_{cumi,\,bat,\,loss}(d,t) \tag{9}$$

2 Optimal Selection and Sizing of Battery Storage System

The main purpose of optimal selection of battery storage system is to minimize the annual operation cost (AOC) as minimum as possible, and it is defined as in Eq. (10)

$$Q_{obj} = \min(AOC) \tag{10}$$

Where AOC includes the net electricity cost, ESS capacity loss cost, and annualized inverter cost. The cost of electricity can be given as in Eq. (11)

$$\text{Electricity\_cost}(d,t) = \text{E\_price}(d,t) \times \text{E\_grid}(d,t) \times \text{sampling time interval} \tag{11}$$

Where E_price (t) is the instantaneous electricity tariff in ($/KWH) and E_grid (t) is the power transfer to and from the grid. Electricity is purchased from grid when E_grid (t) > 0 and sold back to the grid when E_grid (t) < 0.

ESS capacity loss cost during any particular hour can be calculated as in Eq. (12)

$$BCL_{cost}(t) = \frac{BCL(t) \times B_{invest\_cost}}{1 \text{ minimum state of health of the battery}} \tag{12}$$

$$\text{Annualized inverter cost} = \text{inverter capital cost} \times CRF \tag{13}$$

Where CRF can be given as in Eqs. (14) and (15)

$$CRF(\text{Real interest rate}, \text{Inverter life time}) = \frac{\text{Real interest rate}(1+\text{Real interest rate})^{\text{inverterlifetime}}}{(1+\text{Real interest rate})^{\text{inverterlifetime}} 1} \tag{14}$$

$$\therefore AOC = \left[\sum_{d=1}^{365}\sum_{t=}^{24}(\text{Electricity\_cost}(d,t) + BCL_{cost}(d,t))\right] + \text{Annualized inverter cost} \tag{15}$$

2.1 Flowchart for Deciding the Battery Sizing

Flowchart of process for sizing of energy storage system is shown in Fig. 2.

Various input parameters considered in the proposed approach are listed in Table 1.

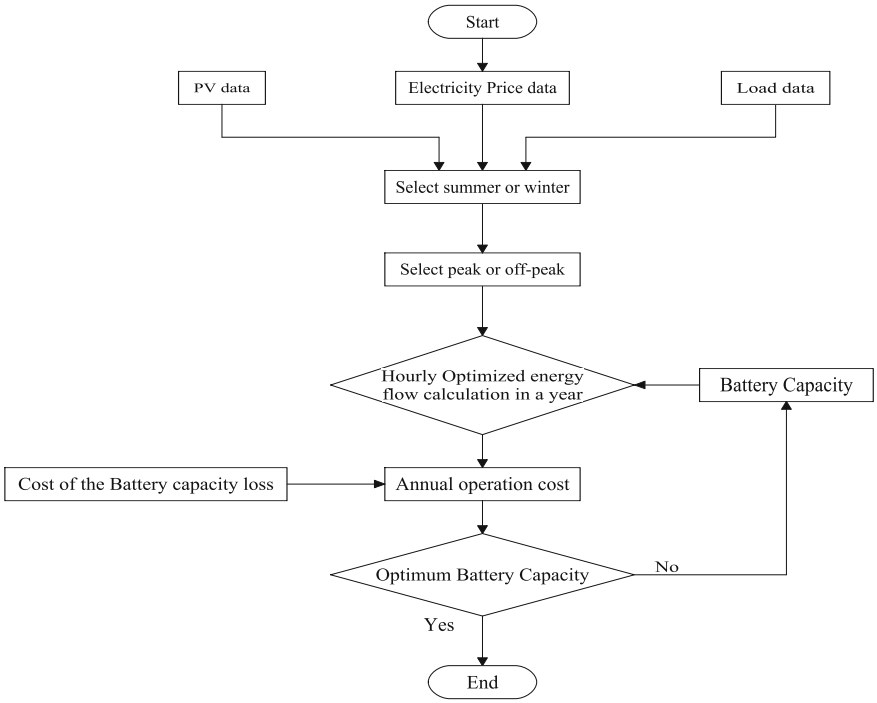


Fig. 2 Process for sizing of energy storage system

Table 1 Input parameter values for MATLAB optimization

Sl. no	Symbol	Variable	Value
1		PV system size	100kWdc
2	Z	Aging coefficient	3e-4
3	Soc_{min}	Minimum state of charge	30%
4	Soc_{max}	Maximum state of charge	90%
5	a	Self-discharging factor	2.5%(per month)
6	t_{min}	Minimum charging/discharging time	10 h
7	η_{inv}	PV inverter efficiency	97%
8	η_{bat}	Battery inverter efficiency	94%
9	η_{ch}	Efficiency of the battery charge	90%
10	η_{disch}	Efficiency of the battery discharge	90%
11	*V*	Nominal voltage of the battery	12v
12	Δt	Sampling time interval	1 h
13	CRF	Capital recovery factor	0.1233
14	i	Real interest rate	4%
15	N	Inverter lifetime	10 years
16		Battery investment cost rate	200$/kwh
17		Battery inverter cost rate	606$/kwh

3 Results

In order to obtain the program results, hourly PV panel output, load data, and electricity price data given as input, use of the energy flow decision based on cost of energy for a particular hour is subjected to the objective function of total annual operating cost (AOC) as minimum.

From Fig. 3 during off-peak time (from 10 pm to 7 am and 1 pm to 4 pm), the amount of energy required to feed the load will be taken from grid and ESS will also charge during this period; during peak times (from 7am to 1 pm and 4 pm to 10 pm), the amount of energy required to load will be taken from PV; if PV is not able to meet the load demand, the preference will be given to ESS.

It is observed from Fig. 4 that battery is well bounded between SOC_{min} and SOC_{max} and we ensured that battery will be working in good condition. During this peak hour period, customer sells all his battery energy to utility, until charging state of the battery reaches minimum state of the SOC. During off-peak time, the required power to feed the load will be taken from grid; in this condition, battery will be charging up to its maximum value of the SOC.

From Fig. 5, at an atypical day starting, cumulative battery capacity loss (CBCL) will be 9.289, it will remain until the battery starts to discharge its energy from 7 am to 1 p.m.; battery capacity gets reduced due to the aging effect.

Figure 6 shows the variation in annual operating cost with different battery capacities for 100KWdc system. The critical size of the battery which gives the minimum annual operating cost of the system is selected; here 18,500 Ah battery is selected. The annual operating cost of the system using this battery is 11421$.

Figure 7 shows the power transfer sequence of the grid-connected PV system without battery storage; during the day, the load is fed by the PV; if PV is unable to reach the load, the energy is taken from the utility grid. If excess PV is present, it

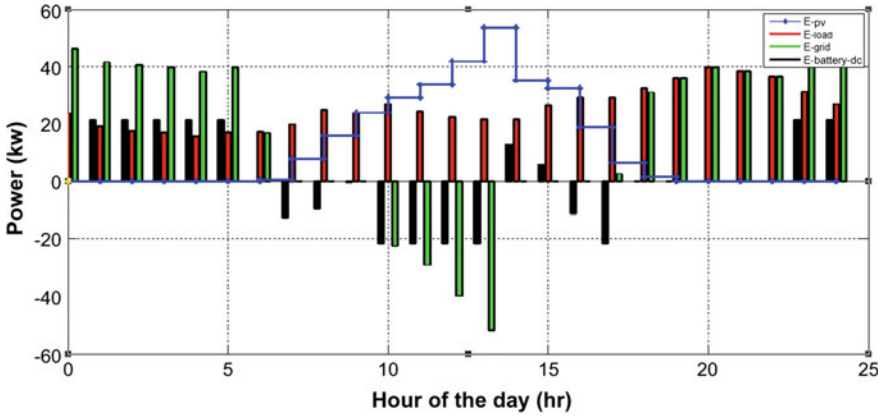


Fig. 3 Variation in active power in PV, load, grid, and ESS

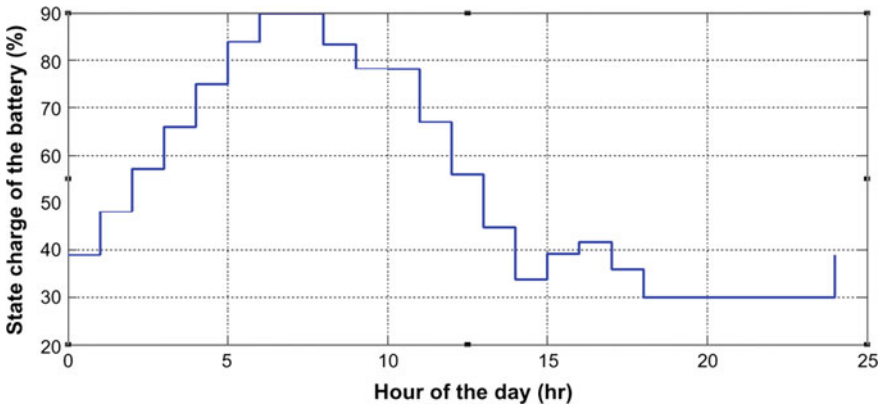


Fig. 4 Charging state of battery

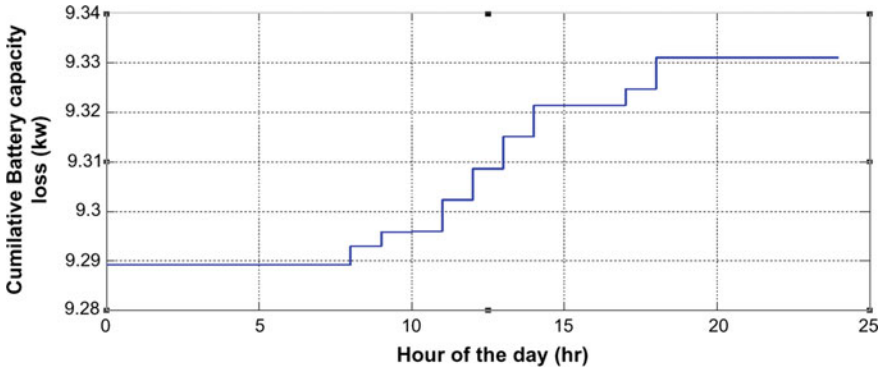


Fig. 5 Cumulative battery capacity loss (kWh)

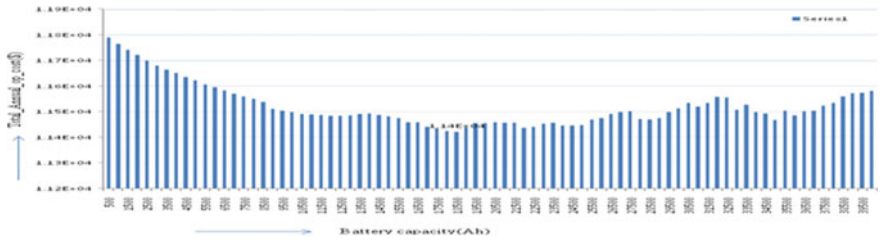


Fig. 6 Optimum value of the energy storage system

will be sent to the utility grid. Table 2 shows the results for selection of battery size and total annual operating cost with and without battery storage system.

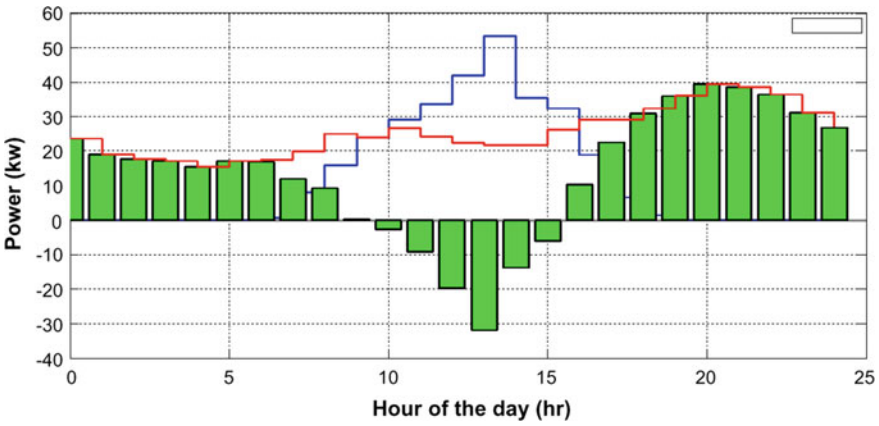


Fig. 7 Without battery storage system during the day

Table 2 Optimal battery capacity results

Name of the parameter	Value
Optimal battery size (Ah)	18,500
Total annual operating cost with battery storage system BSS($)	11,421
Total annual operating cost without battery storage system BSS($)	11,818

4 Conclusion

In this paper, the optimal value of battery storage system for grid-connected PV systems can be found using the power flow program choice on a particular hour subjected to the objective of total annual operating cost as minimum. Without an energy storage system, the total annual operating cost will be more compared to that with energy storage system of same capacity rating, given that some varying parameters were kept constant; for example, power transmission loss and maintenance cost of the battery will be negligible.

References

1. Riffonneau, Y., Bacha, S., Barruel, F., Ploix, S.: Optimal power flow management for grid connected PV systems with batteries. IEEE Trans. Sustain. Energy **2**(3), 309–320 (2011)
2. Lu, B., Shahidehpour, M.: Short-term scheduling of battery in a grid-connected PV/battery system. IEEE Trans. Power Syst. **20**(2) (2005)
3. Nayak, C.K., Nayak, M.R.: Optimal battery energy storage sizing for grid connected PV system using IHSA. In: IEEE Conference Publications (2016)

4. Zhang, T., Cialdea, S., Emanuel, A.E., Orr, J.A.: Electric energy cost reduction by shifting energy purchases from on-peak times. In: 2013 IEEE Electrical Power and Energy Conference
5. Diaf, S., Belhamelb, M., Haddaic, M., Louchea, A.: Technical and assessment of hybrid photovoltaic/wind system with battery storage in Corsica Island. Energy Policy **36**(2), 743–754 (2008)
6. Yang, H.X., Burnett, J., Lu, L.: Wether data and probability analysis of hybrid photovoltaic/wind power generation systems in Hong Kong. Renew Energy **28**(1813), 1824 (2003)
7. Yang, H.X., Lu, L., Zhou, W.: A novel optimization sizing model for hybrid solar-wind power generation system. Solar Energy **81**(1), 76–84 (2007)
8. Kaabeche, A., Belhamel, M., Ibtiouen, R.: Sizing optimization of grid independent hybrid photovoltaic/wind power generation system. Energy, **36**, 1214–1222 (2011)
9. Kusakana, K.: Optimal scheduled power flow for distributed photovoltaic/wind/diesel generators with battery storage system. IET Renew. Power Gener. **9**, 1–9 (2015)
10. Ru, Y., Kleissl, J., Martinez, S.: Storage size determination for grid-connected photovoltaic systems. IEEE Trans. Sustain. Energy **4**(1), 68–81 (2012)

An Effective Multi-faceted Cost Model for Auto-scaling of Servers in Cloud

S. Phani Praveen and K. Thirupathi Rao

Abstract The main aim of this work is to present a multi-faceted cost model depending on server auto-scaling which is based on task requests waiting in queues. To survive for long period in the market, cloud service providers can adopt task—server consolidation. The reason for heavy carbon emanation is power used up on cooling infrastructures of cloud. The present work proposes a dynamic cost model with incentive, fine based on task request time and effect of resource sharing between various tasks to achieve maximum usage of server. Task requests are described at the time of registering with cloud service provider. Cloud clients decide about their cost for service and incentives for the task. Various features are recognized in this profile that includes many weights for the service charge. Possibility of users getting back is also estimated.

1 Introduction

In cloud computing, every client can access services, store, compute, and get back large volumes of information for commercial and domestic use without possessing any resources by pay for usage method. The characteristics such as virtualization, no pre-investment cost, ease of implementation, and fault tolerance takes cloud computing to a great extent. The extreme demand of cloud results gigantic datacenters and various infrastructures. Huge amount of power used in cooling datacenters results in heavy carbon emanation. Huge amount of power is spent in cooling the network systems of datacenters leads to high carbon release. Power consumed for cooling the infrastructure resulting negative assumptions on cloud being green cloud computing. Green cloud is a conscious computing of cloud that is power efficient

S. P. Praveen (✉)
Department of Computer Science, Bharathiar University, Coimbatore, India
e-mail: phani.0713@gmail.com

K. T. Rao
Department of Computer Science and Engineering, KL University, Vijayawada, India
e-mail: profktrao@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_58

which uses resources for computing along with management of waste [1]. The transformation to green cloud can be very useful in long time to environment. Frequent server breakdown is a crucial concern, when servers are underutilized with unfair workload. The major cause of server outage is the necessity of the clients that their requests need to run in separation. Server usage varies from 15 to 30% in many of the cases. Several cloud computing providers with similar approaches like pay less when you use more, pay less when you reserve, pay as you go.

It is assessed that the usage of cloud in US companies could diminish carbon emanation and retrieve up to 13 billion dollar in costs and the corresponding of 300 million tubs of oil by 2020 [2]. Cloud customers can be usually categorized as small to large organizations. AWS is one step forward of its challengers like Century Link, Microsoft, Rackspace. AWS has commenced free layer utilization with a higher control of computing energy and memory since 2010. AWS rules are no fine, no compensation, and no association charges [2]. Business approach has made some amazing progress from making high benefits to support in advertise for more. Till now, cloud computing is concerned only with business. Cloud has appropriated design where they have data centers and servers in couple of areas as per its use per client. It is almost like having Web suppliers, users having diverse necessities for various purposes having distinctive spending plans. Fundamentally, cloud computing have three diverse services, i.e., software as a service (SaaS), infrastructure as a service (IaaS), and platform as a service (PaaS).

2 Related Research

Subcontracting many applications to Cloud can save energy and reduce carbon footmark of companies [3].

Cloud service provider should use a multi-faceted cost model that should be eco-friendly, beneficial and survive in the market for many more years. When amount of work on the server is very less, cloud customer should do their work by using huge job request. Cleer method [4] provides knowledge of carbon release and saving of power in industry move to cloud demonstrated in Fig. 1, and this model has operated on various types of servers globally. This model makes an original code available for scholars to identify what will be saved and how much it will be saved if a business uses cloud.

As user pays for the cloud services, higher priority needs to be given to user. Till now, very less is accomplished near dynamic concession of service-level agreements between customer and services for automatic allocation of resources to various challenging job requirements [5]. In the present state, cloud provider has constraint and stubborn billing on customer to exchange one provider to other provider.

The gap between customer utilization pattern and market policy of cloud service provider is a crucial concern. Review on cost model shows supple price conditions, original setup charges, and amount of discount which are omitted in various cloud providers [6]. The main goal is to report about gap in the pricing model and resource provisioning.

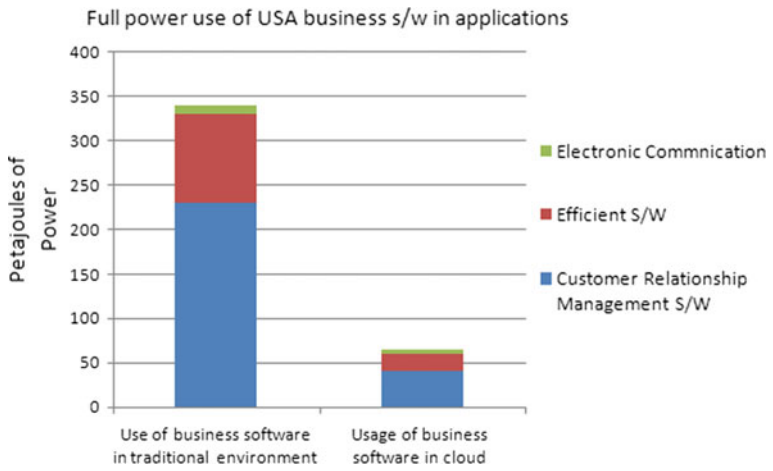


Fig. 1 Power saving on transfer to cloud

Utilization of power of job request is comparative to job profile [7]. Level of multi-tenancy and virtualization [8] should be considered and executed effectively. In many of the cases, it is identified that server usage varies from 10 to 50%.

While balancing the load using heterogeneous server, the work [9] initiated on power restriction leads to performance optimization. Its goals are reducing of response time and best server speed calculation and best load dissemination between the servers.

We can calculate total number of users waiting in the queues by using service rate, arrival rate, upper threshold/lower threshold which leads to addition/deletion of server. A recursive method is utilized to attain stable state [10] where queue size remains stable.

A queue is maintained with on-demand higher threshold and lesser threshold displaying the arrival rate and users waiting in it [11].

In [1], the least dynamic physical mechanism is resolved by observing total resource weight ratio of virtual machine; it contains virtual machines immigration in case of extreme load on dynamic physical mechanism.

Scheduling method and amount of server usage are the two main metrics of cloud computing. Suppose all tasks are incoming at a time, M-M D R R A was introduced by [12]. It is a higher version of round robin algorithm where time period is variance among minimum burst time and maximum burst time of job requests. Less waiting time and minimum amount of context switches lead to higher throughput. Every chromosome is denoted as series of genes. Gene is signified as list or slots allotted to tasks prepared randomly. Each job request is classified into two or more tasks and using delay scheduler, SGO-based scheduler, and SJFS scheduler measured the task/server association. The goal of SGO algorithm is the most recent execution time of all the tasks [13]. By using DVSM and load balancing, energy of slightly loaded processor is decreased whereas in case of overfilled by transferring to slighter loaded system. Survivability of cloud provider is based on its quality of service and

estimating strategy [14]. Potential is the variance of time period between entire execution and onset of the job requests. Cloud providers utilize various types of cost strategies. Cost is generally classified into three types:

(1) On-demand rate,
(2) Reserved prices,
(3) Spot services prices.

As soon as the market rate beats a certain price, the service is removed from the provider without any earlier notice in on-demand services. It is really very useful for mini businesses or individual use. Risk analysis for on-demand services is provided Amazon Web service which benefits in bidding. Drawbacks of spot services are no assurance of whole implementation of job request, inconsistent prices, traversing regular interruption.

In [15], authors focus on the on-demand pricing of services, arbitrage effect, and latency across markets of USA. From CloudSleuth.com, latency information is gathered. Hence, the present work identified several issues which affect the profile of task request, service charge and proposes the dynamic pricing method taking Nash equilibrium into concern.

3 Model Description

1. Cloud customer—Cloud customer can appeal services from anyplace across world through Web. Customers are characterized into four types:
   (i) Domestic customer or commercial customer,
   (ii) Time (off-peak time, less peak time, peak time),
   (iii) Quantity usage (high, medium, low),
   (iv) Time-based elements (old user, new user, bestowed user).
2. Task request Profiler—Gathers precise features and selections of customer and allot weightage to every feature while profiling every task request. All requests are executed by scheduler based on the selections at task request profiler; it directs to the particular queue.
3. Centralized admin server—Several similar physical devices are used as server admin, where every admin system creates many VMs to execute various task requests in parallel.
4. Service cost—The cost for a service is based on service-level agreement between customer and cloud service provider.
5. Incentive—In our method, every task request is recorded and allotted a distinctive task identity, along with the predicted service cost, incentive. The incentive is based on several factors and every factor has weight.
6. Fine—Task request during the high peak hour leads to congestion. This cost is called fine.

Let us consider a load balancer of several queuing systems with queue-based different virtual machines. Number of task requests is placed in load balancer, and after recording with task request profiler, it directs into its specified queue followed by server for execution.

The load balancer divides the arrival requests into three sub-types in Eq. (1):

$$\$ = \$1 + \$2 + \$3 \tag{1}$$

Requests in queue are executed in first come first serve depending on their selection of server. Now, auto-scaling of server helps to avoid over server usage. If task requests waiting in the queue surpasses higher threshold, then a new server is added to execute the requests. Likewise, task requests waiting in the queue are below the lesser threshold then server is decreased by 1. Our algorithms of task scheduling are knowingly dissimilar in their cost-effectiveness. Here, removing server will be completed only after complete execution of task requests. Relocation of task requests from one server to other server leads to overhead of storage.

In our model, multi-faceted cost model is represented in Fig. 2; we have considered three queues (blue queue, yellow queue, and red queue). Every task request is allotted an incentive. Identified parameters are allocated separate weight which

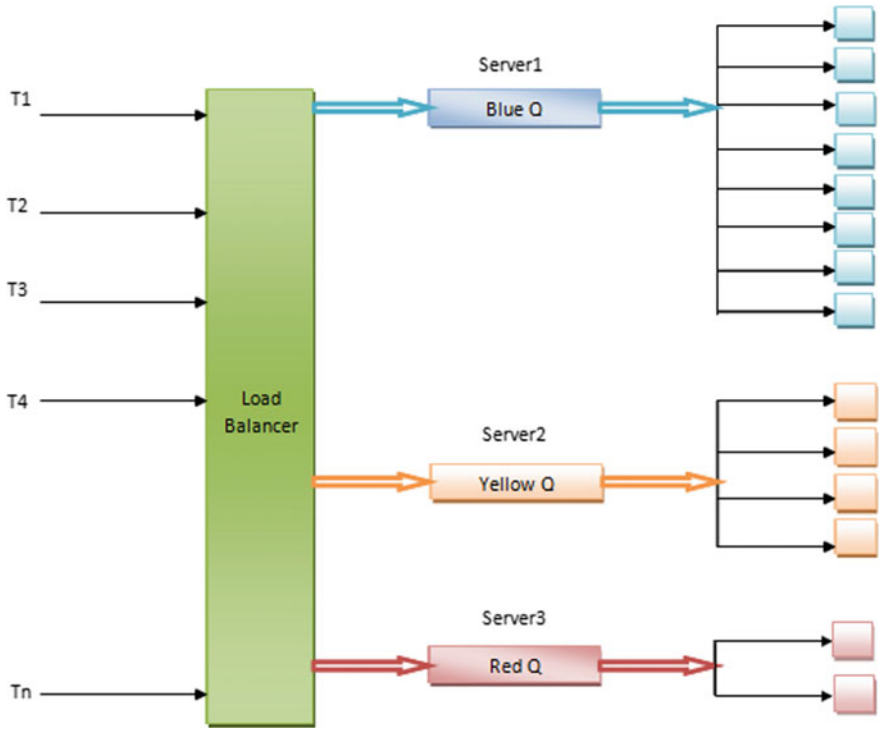


Fig. 2 Multi-faceted cost model

adds the pricing amount. Task requests in every batch are organized in rising order. Time quantum (TQ) is measured by using RR algorithm [12] for virtualization of the server. Variance between the least and extreme capacity of task requests in the batch is taken as the TQ. The rate of TQ is calculated iteratively.

Blue queue contains eight requests, which are reserved in single batch and executed in parallel and server is shared with all assigned task requests. In yellow queue, four task requests are executed in parallel in a batch. Red queue processes two task requests serially. The demand of task requests is controlled by using incentive and fine.

Giving incentive off-peak hour utilization and fine for congestion at peak time will decrease the carbon emanation to a large extent.

3.1 Multi-faceted Cost Model Algorithm

- Step 1: Task requests are recorded with load balancer; all the external tasks are kept in form of rows (task id, customer id, commercial or domestic, type of client, type of server, task request time), Number of Active Servers = 3, Maximum Active Servers = 10.
- Step 2: Based on customer selections of server and task profile, service cost and reward are assessed in accord to pricing strategy.
- Step 3: If server selects blue, tasks enters into blue queue. Task requests form a batch of 8, host creates eight virtual machines and tasks are processed by using RR scheduling algorithm.
- Step 4: If server selects yellow, tasks enter into yellow queue. Task requests form a batch of 4, host creates four virtual machines and tasks are processed by using RR scheduling algorithm.
- Step 5: If server choice is red, requests are processed serially depending on FCFS.
- Step 6: If sum of tasks waiting in any two queues is more than upper threshold, then active servers is increased by 1. Active Server = Active Server + 1.
- Step 7: If the sum of tasks waiting in any two queues is lower than lower threshold, then the active server is decreased by 1. Active Server = Active Server - 1.
- Step 8: Calculate total income, across blue server, yellow server, and red server, incentive for every task request, fine (if practical).

The total incentive is calculated by using Eq. (2)

$$\text{Total Incentive} = \sum 3i = 1 \ \text{Ti*Se*Vi*Ui*S.} \tag{2}$$

T = Time period of task request, peak time task request (t1 = 0.3), less peak hour job- request (t2 = 0.6), off-peak-hour (t3 = 0.8); Se = Extent of

multi-tenancy, red server (s1 = 0.3), yellow server (s2 = 0.6), blue server (s3 = 0.8), V = Volume of job requests, when low (v1 = 0.3), medium (v2 = 0.6), high (v3 = 0.8) U = type of user, when new (u1 = 0.3), old (u2 = 0.6), frequent (u3 = 0.8), and S = size of job requests in MIPS.

4 Experimental Evaluation

Our multi-faceted cost model contains three fragments: task request profile features, CPU consumption [16], and incentive/fine connected with every task. The task requests arrival from 5 to 25. Service charge of every server is based on the execution speed of the server and size of task requests. Results are then obtainable depending on latency time of task request, power usage, and waiting time in the queue. Power usage is being direct with CPU consumption [17]. We have taken average CPU use as the measure for power used. In our simulation, several similar servers are considered for executing task requests; in all circumstances, three servers are active, whereas extra servers are added on initiate of sensor when task requests waiting in the queue cross its higher threshold. Likewise, servers get decreased when tasks waiting in the queue are lower than lower threshold such that a best trade in can be formed between addition and removal of server and optimal server utilization. On completion of allotted task request, next task which is on list is taken by the CPU. The virtual CPU of a VM,

Vcpu can be calculated in Eq. (3) as:

$$Vcpu_i = \frac{VMmips_i}{HOSTmips_i}. \tag{3}$$

$$\text{VirtualCPU}_\text{i} = \text{Vitual machine}(\text{mips}_\text{i})/\text{Host}(\text{mips}_\text{i})$$

In every simulation, 250–300 tasks are generated from 50 customers.

Every task capability ranges from 10 to 40 MIPS which is created randomly. Million instructions per second (MIPS) are used to describe the allocation speed of host device, virtual machine, and size of every request. The number of requests in the buffer, waiting in the particular queues, completed is read every 5 s. Total waiting time, latency time, and processing time are calculated on completion of processing of task request.

Figure 3 illustrates various task requests processed in total three queues and task requests waiting to be processed. The x-axis illustrates waiting task requests in red queue (in red *), yellow queue (in yellow *), and blue queue (in blue *). Correspondingly processed red job requests in red square, processed yellow job requests in yellow, and processed blue job are designed in blue square.

The requests waiting are high in red queue (expressed in red color), whereas requests waiting in yellow queue (expressed in yellow). Number of requests waiting in blue queue is least. Blue server performs task requests in group. The job requests

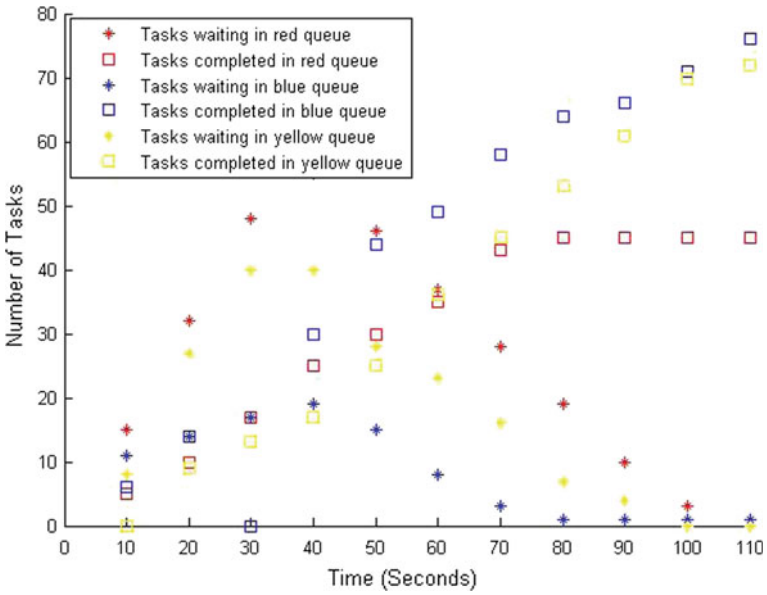


Fig. 3 Number of task requests waiting for processing and processed

processed entirely are maximum in blue and minimum in yellow server. Every element is allotted with few weights (varying from 0.2 to 1.0). In this experiment, 15 host devices and 25 VMs are taken. Every host favors up to 4 VMs. We can attain higher usage of servers by encouraging multi-tenancy and virtualization.

Figure 4 shows the performance of servers. Blue server processes task requests in a batch of 8, yellow server processes requests in a batch of 4, and red server processes requests successively. More MIPS is processed by blue server, whereas least in red server (expressed in red color) and yellow server (expressed in yellow) performance lies between blue and red. In order to achieve green model, CPU usage rate must be greater than 60%. In terms of power performance, the more VMs can be placed on the host device to get maximum rate of usage of CPU and minimum carbon emanation.

According to approach time of task request, time is categorized into three time slots, off-peak hours is from 12 am to 8 am, instances of peak hour mention task requests located between 8 am and 4 pm, and less peak hour denotes 4 pm–12 am. The clearness of the promoting rate, the effect of the elements on estimation of fine and incentive are considered in the resulting

Incentive estimation procedure is displayed in Table 1.

Profile of some task requests is displayed across row.

Task request with serial number 3 is a domestic request where customer is divided into dedicated server type: 2 for red, 4 for yellow, and 8 for blue. Incentive considered in case of task request 1 is 0.3*0.3*0.6*0.6*V, incentive for task request 2 is 0.6*0.6*0.6*0.6, and 0.8*0.3*0.6*0.8*V for task request 3.

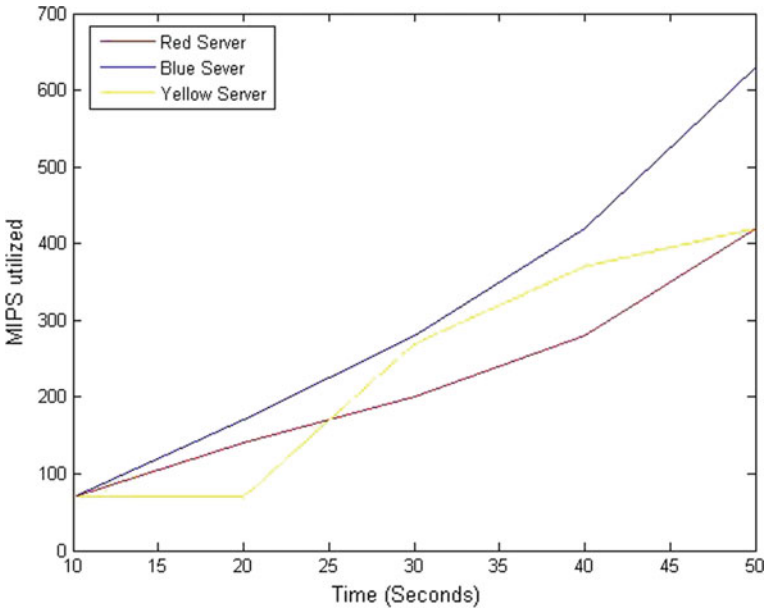


Fig. 4 MIPS consumed per 10 s

Table 1 Incentive industry regulation

Sl No	D/C	N/O/B	Server	Time	Usage
1	2	2	2	3	3
2	2	3	3	2	2
3	2	4	4	4	2
4	3	4	2	3	4
5	3	2	4	3	3

In our experiment, different costs for commercial and domestic users are considered. Based on the volume utilized, it is categorized into three parts, i.e., high, moderate, and low. Domestic clients get maximum profit for less usage, whereas commercial clients get best profit when using high. Maximum CPU usage reduces carbon emanation and waiting time of task requests. In Fig. 5, experiment result is plotted in the form of candle chart. It is observed that when high number of tasks selecting blue server, then money invested toward incentive is maximum than otherwise. Income invested toward incentive varies from 0.2 to 8%.

The rate of customer returning is defined in Eq. (4).

$$\text{Customer\_returning}_t = I_i * \text{Incentive}_{t-1}. \tag{4}$$

Here, I_i is a constant whose value lies between 2 and 3.

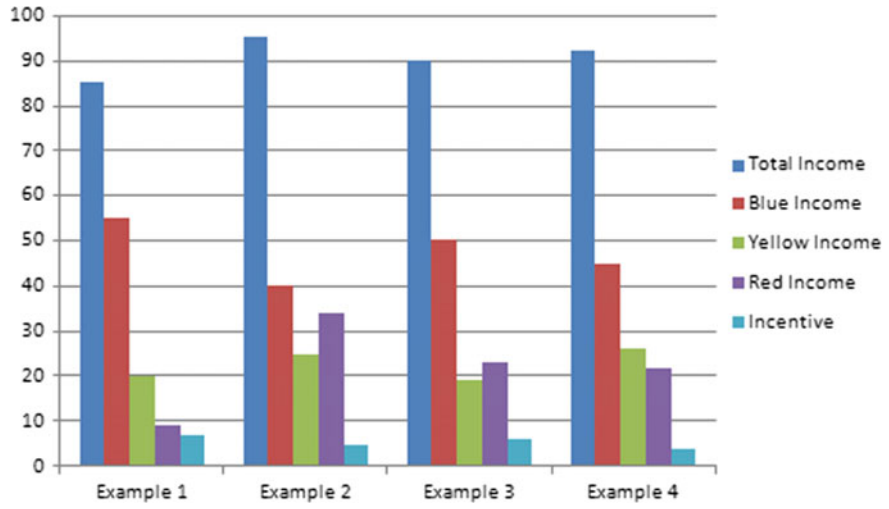


Fig. 5 Experiment result

$$\text{Total Income} = \text{Total Income of service cost} - \text{Income spend for Incentive} + \text{Blocking Charge} \tag{5}$$

The total income is measured with formula defined in Eq. (5).

Although incentive is negative income for provider, it improves survivability of the provider in market for longer than its opponents. Higher incentive improves the possibility of returning a user to the provider by 20% than with no incentive. Drawback of our work is that information locality is not considered.

5 Conclusion

Long-time existence of cloud server provider is based on the planned survey of its cost model, effective resource usage, optimization of energy utilization and clearness with the end customer. Cloud federation need to include reusable source of energy, widespread virtualization, energy effective hardware, expansion of computerization, effective software, energy proficient disk type of memory, liable utilization of internet traffic flow etc., to attain Green Cloud Computing and lessen carbon footprint. Including Nash equilibrium both at cost model and end customer request can collectively achieve environment supportable green cloud, the best influential parallel computation policy which incorporates retail, logistics production, finance, advertising, etc. Model of this perception can be completed at a higher capacity to get enhanced transparency on its merits and demerits. Data locality

should be taken for improved vision. Virtualization should be integrated effectively in order to achieve resource optimization having lesser carbon emanation.

References

1. Mell, P., Grance, T.: The NIST Definition of Cloud Computing, vol. 1, pp. 26–36. National Institute of Standards and Technology (2009)
2. Iveroth, E., Westelius, A., Petri, C.-J., Olve, N.-G., Cster, M., Nilsson, F.: How to differentiate by price: proposal for a five-dimensional mode. Future Gener. Comput. Syst. **2**, 16–26 (2012)
3. Garg, S.K., Buyya, R.: Green cloud computing and environmental sustainability. In: S. M. a. G. G. (eds) Harnessing Green IT: Principles and Practices, pp. 315–340. Wiley, New York (2012)
4. Masanet, F., Shehabi, A., Liang, J., Ramakrishnan, L., Ma, X., Hendrix, V., Walker, B., Mantha, P.: CLEER model (2013), Lawrence Berkeley National Laboratory, http://crd.lbl.gov/news-and-publications/news/2013/study-moving-computer-services-to-the-cloud-promises-significant-energy-savings
5. Buyya, R., Yeo, C., Venugopal, S.: Market-oriented cloud computing: vision, hype, and reality for delivering it services as computing utilities. In: Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, pp. 5–13 (2008)
6. Huang, J.: Pricing strategy for cloud computing services. In: The Proceedings of the Pacific Asia Conference on Information Systems (PACIS) 2013, 279 pp
7. Aggar, R.B.M., Ajenstat, F.: Cloud computing and sustainability: the environmental benefits of moving to the cloud. Accenture Microsoft Report (2010), http://www.wspenvironmental.com/media/docs/newsroom/cloudcomputingandSustainability
8. Vogels, W.: Beyond server consolidation. ACM Appl. **6**, 20–26 (2008)
9. Cao, J., Li, K., Stojmenovic, I.: Optimal power allocation and load distribution for multiple heterogeneous multicore server processors across clouds and data centers. IEEE Trans. Comput. **63**, 45–58 (2014)
10. Goswami, V., Patra, S.S., Mund, G.B.: Performance analysis of cloud with queue-dependent virtual machines. In: Proceedings of the 1st International Conference on Recent Advances in Information Technology, pp. 357–362 (2012)
11. Aljohani, A.M.D., Holton, D.R.W., Awan, I.: Modeling and performance analysis of scalable web servers deployed on the cloud. National Institute of Standards and Technology, pp 238–242 (2009)
12. Panda, S.K., Bhoi, S.K.: An effective round Robin algorithm using min-max dispersion measure. Int. J. Comput. Sci. Eng. (IJCSE) **4**, 45–53 (2012)
13. Praveen, S.P., Rao, K.T., Janakiramaiah, B.: Arab. J. Sci. Eng. (2017). https://doi.org/10.1007/s13369-017-2926-z
14. Pal, R., Hui, P.: Economic models for cloud service markets pricing and capacity planning. Distrib. Comput. Netw. (ICDCN) 113–124 (2012)
15. Cheng, H.K., Li, Z., Naranjo, A.: Cloud computing spot pricing dynamics latency and limits to arbitrage (2013)
16. Yang, C.T., Liu, J.C., Huang, K.L., Jiang, F.C.: A method for managing green power of a virtual machine clustering cloud. Future Gener. Comput. Syst. **37**, 26–36 (2014)
17. Yamini, R.: Power management in cloud computing using green algorithm. In: IEEE International Conference on Advances in Engineering, Science and Management, 30, 31 March 2012, pp. 128–133

Sign Language Translator Using LabVIEW Enabled with Internet of Things

M. Pradeep Kumar, M. Thilagaraj, S. Sakthivel, C. Maduraiveeran, M. Pallikonda Rajasekaran and S. Rama

Abstract The sign language translator is an aiding device which uses a hand glove embedded with sensors for converting sign language into the English language. This aiding device targets the speech and hearing impaired community. The speech and hearing impaired people are having problem in communicating with normal people. The current aiding devices in the market fail to evacuate their communication problem. This project focuses to resolve their problem by giving a gesture glove as aiding device using NI platform. We have enabled Internet of Things to establish better learning skills which in turn improves their literacy rate.

1 Introduction

In today's busy world, people are so busy minding their own problems and not able to pay attention to others' problems. One such problem that we saw in the society was the deaf and dumb people were unable to communicate with the normal people

M. Thilagaraj (✉) · M. P. Rajasekaran
KLU-NI Technology Innovation Centre, Kalasalingam University, Anand Nagar, Krishnankoil, India
e-mail: thilagaraj@klu.ac.in

M. P. Rajasekaran
e-mail: m.p.raja@klu.ac.in

M. P. Kumar · S. Sakthivel · C. Maduraiveeran · S. Rama
Department of Biomedical Engineering, Kalasalingam University, Anand Nagar, Krishnankoil, India
e-mail: predeepkumar30@gmail.com

S. Sakthivel
e-mail: sakthivelsankaran92@gmail.com

C. Maduraiveeran
e-mail: maduraiveerancmv@gmail.com

S. Rama
e-mail: ramasolai123@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_59

and every time they need to use written communication to normally talk with people. So we tried to work on giving a solution to those people by designing a glove with sensors attached to translate the motions into letters and words, which will be displayed in the computer.

In the survey of literature, various articles have been published for sign language translation where either only letters are displayed or only word is displayed. Here in our paper, we have designed such a way that we are able to take both letters and words so that this can be used for educational purpose. The target group of speech and hearing impaired people ranges about 26% of the total population of disabled people according to 2011 census statistics of India [1] as shown in Fig. 1. These persons usually meet with a problem of communicating with normal persons in regular circumstances [2].

Many authors have tried to develop building words with only flex sensors and use LCD display for displaying words. One such article was designed for displaying letters individually and then framing combined letters into words [3]. Mohammed Elmahgiubi et al. developed a data acquisition system for sign language translator with 96% accuracy [4]. Jamal Hayder et al. built a glove with flex sensors and an accelerometer for American Sign Language finger spelling [5]. Yellapu Madhuri et al. developed a vision-based sign language translation device for converting sign to speech output [6]. Carlos et al. proposed a deaf-mute communication device to translate the words to audio signal [7]. Rajat Sharma et al. have proposed a portable embedded device for accurate message communication [8].

The flow of the paper is in such a way that it introduces the various system modules in Sect. 2 and then the working with flow diagram in Sect. 3 with the values listed in table for letters and words. Finally in the Sect. 4, we discuss the results and conclusions on how this project has been helpful. In Sect. 5, we discuss the future work that we are carrying to extend the project.

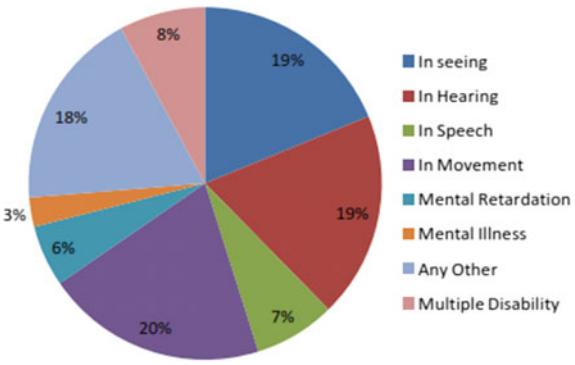


Fig. 1 Disabled population by type of disability in India

2 Finger spelling

Finger spelling is a form of borrowing, a linguistic process wherein words from one language are incorporated into another. American Sign Language (ASL) is the standard language that serves for SHIP community to communicate with the normal people. ASL consists of a set of 26 signs which can be used to spell out words from the English language. The ASL manual alphabet which is used for all alphabets is shown in Fig. 2. For the clarity of letters to be displayed, the symbols are shown in different views like side view and front view. Each sign makes a specific impartment in the SHIP community which makes them so fluent to guess the symbols. This finger spelling is similar to the blind people's reading where they are trained through with different recognitions.

3 System Description

The device uses five flex sensors, accelerometer from NI myRIO, another accelerometer, NI myRIO controller, NI LabVIEW software as a whole setup. The flex sensors which are fixed with the gloves on each finger part read the data from each

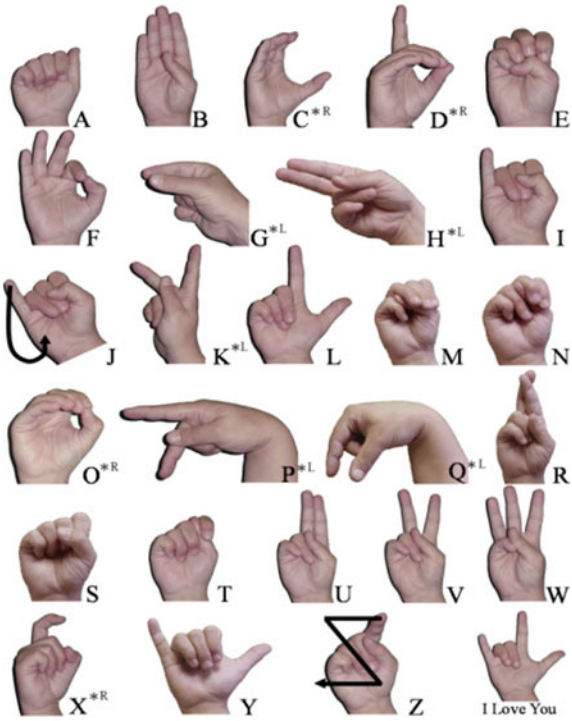


Fig. 2 The ASL manual alphabet

finger when it bends for a particular word. The NI myRIO which has a built-in accelerometer reads the value of gestures along with another accelerometer which has been fixed perpendicular to that myRIO's accelerometer. The NI myRIO has built-in Wi-Fi transceiver which transmits the data from the sensor part to the computer. The controller and the computer are interfaced using NI LabVIEW software. As detailed in the title, IoT has also been enabled using single-server multi-client access over LAN coding in LabVIEW software. The various hardware components used in the design of our prototype have been shown in Figs. 3, 4, and 5.

Fig. 3 Flex sensor

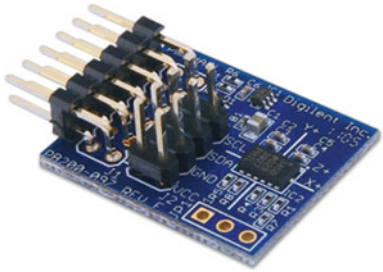


Fig. 4 Pmod accelerometer

Fig. 5 NI myRIO

3.1 Flex Sensors

They are also called bend sensors which measure the amount of deflection or bending. The variation in bend is measured through the change in variation of resistance. The resistance of the flex sensor changes when the metal pads are on the outside of the bend changes.

3.2 NI myRIO

This is a reconfigurable embedded device that features input/output on both sides of the device. The NI myRIO provides analog input (AI), analog output (AO), digital input and output (DIO), audio and power output in a compact embedded device. This device acts as a microcontroller in the whole project.

An accelerometer which is available in the myRIO is used to measure the position of the chip. The accelerometer measures both static acceleration for drop sensor applications and dynamic acceleration for shock and vibration data acquisition.

3.3 Pmod Accelerometer

The Pmod accelerometer is a three-axis digital accelerometer used to provide high-resolution acceleration changes including inclination changes of less than 1.0°. The Pmod accelerometer communicates with the host board via the SPI protocol or the I^2C protocol.

4 Working Description

The working of the device is a flow as a step-by-step process execution which has been achieved such that the output will be prompt. The first step involves acquiring data from fingers when a sign is shown by the user. The flex sensors when it bends changes the resistance and generates analog values together with the two accelerometers also gives corresponding values for a particular sign. These sensors which are connected to the NI myRIO controller send the data to myRIO. The myRIO acquires the data and transfers it wirelessly to the receiving end computer using the built-in Wi-Fi transceiver.

The acquired data to the system has been analyzed using NI LabVIEW coding. Each sensor value is assigned to a particular range using “In Range and coerce” functional palette. After setting the parameters to the sensors, the values are equated

with the respective string output using case structure. The true case involves, if all the values are in range as specified then the respective string output will be retrieved.

The received string output is transferred to the host processor, i.e., the computer processor using shared variable. In this VI, the text output has been converted into voice output using Microsoft text-to-speech convertor. As it is IoT enabled with single-server multi-client access over LAN, which is achieved by using TCP protocol in Data communication in NI LabVIEW software. This enables that all the interconnected systems can receive the same data at the same time without any delay which enhances the literacy rate of the deaf and mute people. The whole flow process of the model is shown in Fig. 6.

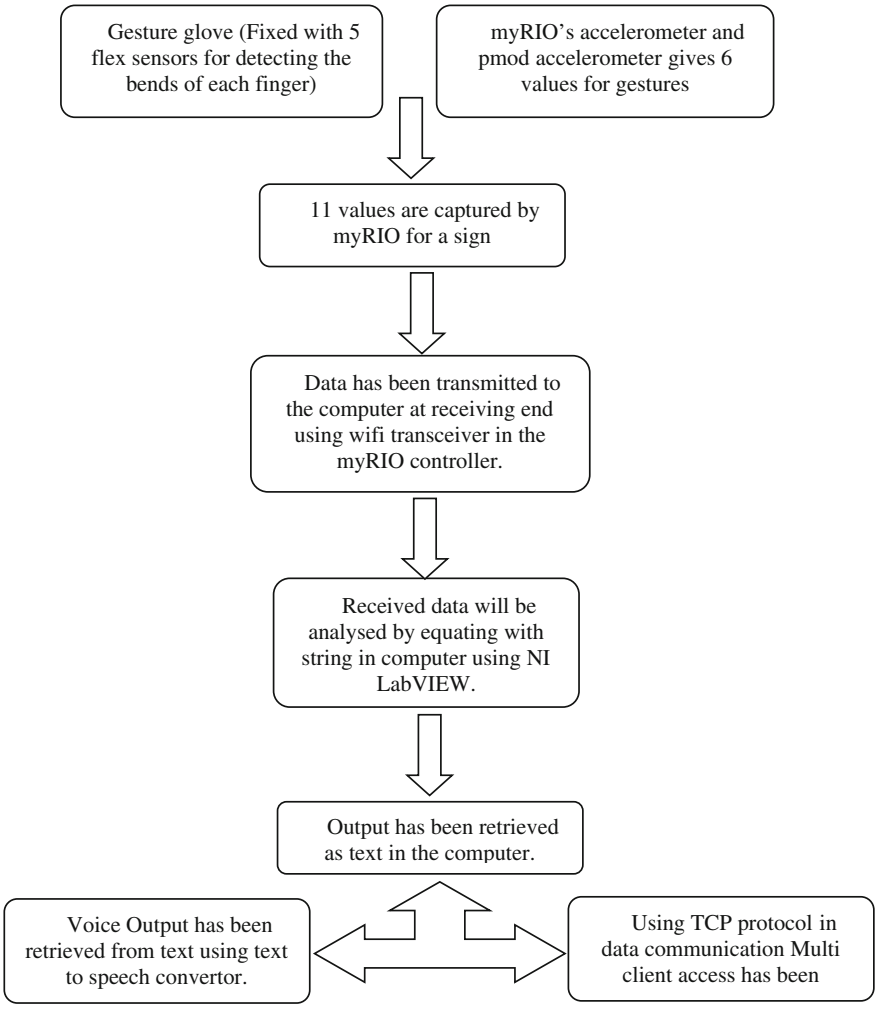


Fig. 6 Flowchart of the working description

5 Conclusion and Future Work

It is concluded that the sign language translator glove helps the disabled as an aiding device and also it lifts the communication barrier they have in regular scenario. It is also clear that this device can replace the other aiding devices in the market which are available for the speech and hearing impaired people. Since all the other devices fail to establish communication, this as an aiding device kicks their communication barrier. The accuracy is high as such the particular sign or gesture is deliberately translated into the English language. This accuracy is achieved because of two accelerometers which are placed perpendicular to each other. The sample values for letters and words are listed in Tables 1 and 2 and 3 simultaneously. As an additional feature of multi-client access using TCP protocol, this enables accessing the same data in real time in all the systems which are interconnected. Thus, it also ensures establishing a sound knowledge in English equivalent to the sign or gesture. The front panel and block diagram of the system which shows the output are mentioned in Figs. 7 and 8, respectively.

Table 1 Analog values of flex sensors for signs

Word/letter	Thumb	Index	Middle	Ring	Little
A	40–100	20–60	40–120	310–350	750–820
B	0–100	0–2	170–200	1–10	60–100
C	4–10	4–19	180–186	5–15	81–120
D	0–100	0–2	180–200	4–25	110–150

Table 2 Analog values of myRIO accelerometer for gestures

Word	X-axis (myRIO)	Y-axis (myRIO)	Z-axis (myRIO)
MY	30 to 35	–70 to 35	–140 to 230
NAME	–230 to –180	–140 to –110	0 to 60
HI	–210 to –170	–130 to –160	10 to 70
THANK YOU	270 to 290	–140 to –110	0 to 60
SLOW	40 to 110	–260 to –220	–50 to 0
FAST	–100 to 260	–20 to 25	–255 to –20
HOW	270 to 290	50 to 170	0 to 60
NICE	–180 to –150	–260 to –220	–50 to 0
MEET	50 to 135	–70 to 35	150 to 260
SORRY	300 to 435	–140 to –110	5 to 50
UNDERSTAND	–210 to –170	–130 to –160	15 to 56

Table 3 Analog values of Pmod accelerometer for gestures

Word	X-axis	Y-axis	Z-axis
MY	1000 to 150	–970 to 500	–240 to 180
NAME	–45 to 20	–110 to –60	–710 to –500
HI	–165 to –120	–10 to 30	–265 to –245
THANK YOU	90–130	– 110 to –60	–710 to –500
SLOW	30–60	–50 to 100	–980 to –930
FAST	–485 to 1030	70 to 970	–20 to –200
HOW	90 to 130	–135 to –120	–710 to –500
NICE	–250 to –220	–50 to 100	–980 to –930
MEET	226 to 341	–970 to 500	–195 to –175
SORRY	160 to 170	–110 to –60	–700 to –500
UNDERSTAND	–150 to –172	–265 to –245	–10 to 30

In near future, the entire setup will be converted into a single portable device which can be used in real time by the disabled. The cloud data transfer will also be enabled as an additional feature of IoT.

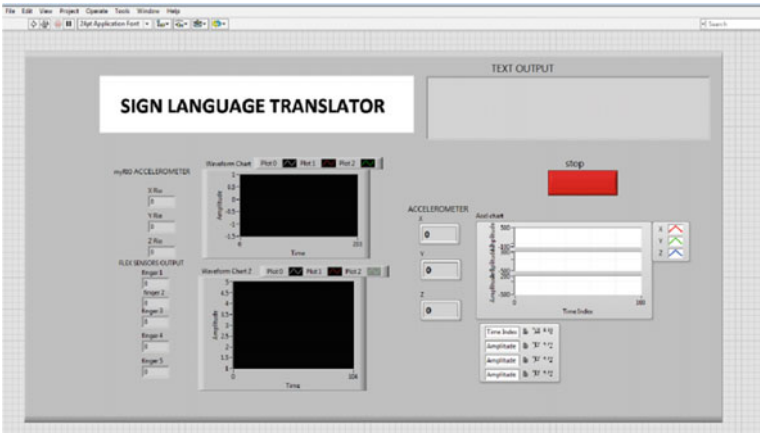


Fig. 7 Front panel diagram

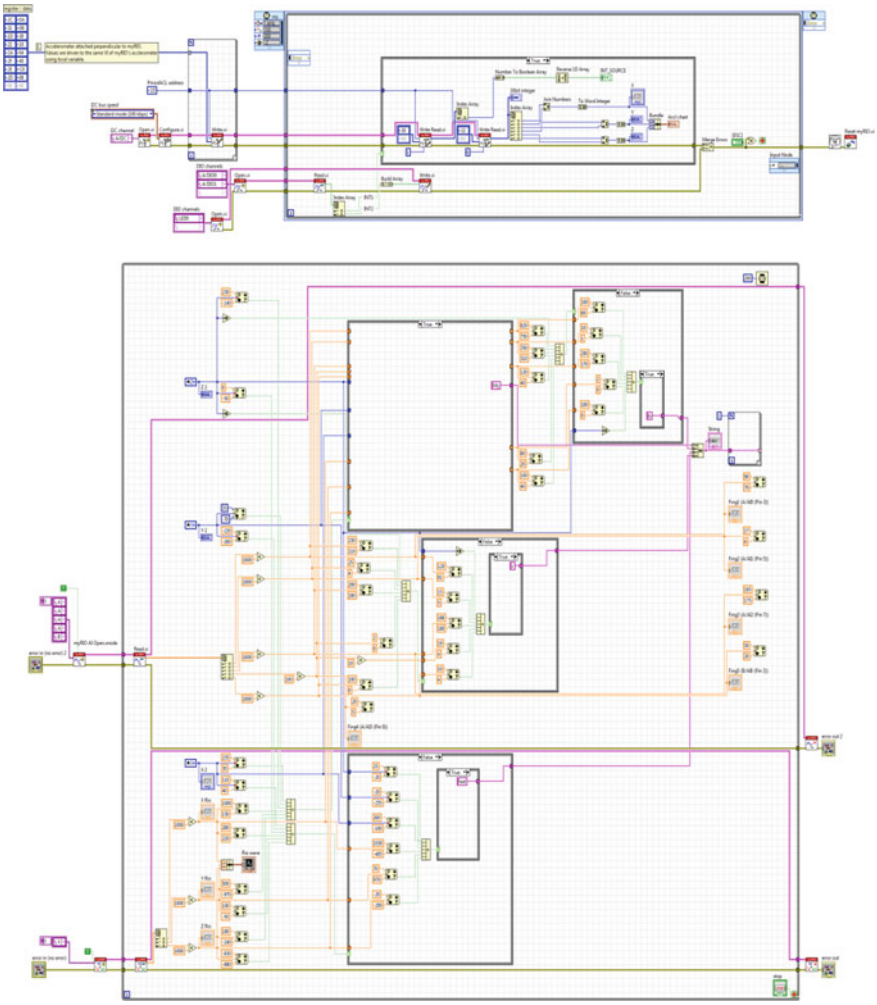


Fig. 8 Block diagram

Acknowledgements The authors would like to thank the Management of Kalasalingam University. Also, we thank the KLU-NI Technology Innovation Centre of Kalasalingam University, Tamil Nadu, India, for permitting to use the computational and hardware facilities of the center for my research work.

References

1. Disabled persons in India a statistical profile 2016, Social statistics division Ministry of Statistics and Programme Implementation, Government of India

2. Stenberg, M.: The American Sign Language Dictionary, Multicorn (1996)
3. Solanki Krunal, M.: Indian sign language using flex sensor. Int. J. Eng. Trends Technol. **4**(6) (2013)
4. Elmahgiubi, M., Ennajar, M., Drawil, N., Elbuni, M.S.: Sign language translator and gesture recognition. In: IEEE International conference on Computer and Information Technology (2015). https://doi.org/10.1109/gscit.2015.7353332
5. Hayder, J., Dalal, B., Hussainy, S., El Khansa, L., Fahs, W.: ASL fingerspelling translator glove. Int. J. Comput. Sci. **9**(6), 1 (2012)
6. Madhuri, Y., Anitha, G., Anburajan, M.: Vision based sign language translation device. In: IEEE International Conference on Information, Communication and Embedded Systems (2013). https://doi.org/10.1109/icices.2013.6508395
7. Fiel, C.P., Castro, C.C., Pitalua-Diaz, N., Jimenez, D.S.: Design of translator glove for deaf-mute alphabet. In: IEEE International Conference on Electric and Electronics (2013)
8. Sharma, R., Bhateja, V., Satapathy, S.C., Gupta, S.: Communication device for differently abled people. In: Proceedings of the International Conference on Data Engineering and Communication Technology. Advances in Intelligent Systems and Computing, vol. 469. https://doi.org/10.1007/978-981-10-1678-3_54

Study of Galaxy Evolution Through GA-SVM

M. Raja Sekar and N. Sandhya

Abstract Galaxies' development and progress are the major problems in cosmology. Evolution of galaxies plays a vital role in cosmographic senses. Several researchers found that galaxies' evolution takes place by merging different subsystems. These subsystems form the basis for the galaxy evolution. Also, cosmic time depends on lens statistics which is associated with the quality of lens statistics. Galaxy development depends on the factors like masses surrounding the galaxies. In this study, we formulated models which help in understating the formulation of galaxies with the help of support vector machines embedded with genetic algorithms. The models which are formulated depend on the distribution function of galaxies. Datasets are analysed with the help of HST rules. The major contribution of the anti-evolutionary model is to identify the small angle of separation of galaxies more precisely. This study is carried out on the publicly available EDGE database. This paper describes the analysis of galaxies with genetic algorithms.

1 Introduction

For the optical astronomer, the most striking features of the universe are the fact that stars appear in the discrete groups known as galaxies. Astronomers are forever examining the observational properties of galaxies in search of clues about how they may have formed [1–4].

Galaxies come in several clear types, and it is a challenge to account for their distinct properties. At the crudest level, galaxies form a two-parameter family in which the characteristics that vary are the total amount of light and how this is divided between two components: the bulge and the disc. The family of galaxies formed by adding bulges and discs in different proportions is called the Hubble

M. Raja Sekar · N. Sandhya (✉)
Department of CSE, VNRVJIET, Hyderabad, India
e-mail: sandhyanadela@gmail.com

M. Raja Sekar
e-mail: rajasekar_m@vnrvjiet.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_60

sequence. This simple sequence of varying bulge-to-disc ratio has spawned a variety of historical terminology, because Hubble originally thought of his sequence as being an evolutionary path, elliptical and so can be called early-type galaxies and spiral are late-type galaxies. The bulge component is sometimes called spheroid (especially in Milky Way) and so elliptical may be referred to as spheroidal [5–9].

Any interpretation of the spectral data on galaxies requires some means of predicting how galaxies should appear as a function their star-formation histories. Over a number of years, a semi-empirical method has arisen in order to meet this need. Stellar evolution evolutionary spectral synthesis which means that the theory of stellar evolution is used to predict the population of HR diagram as a function of time, starting from a main sequence with a known probability distribution of stellar mass, the IMF (initial mass density function). The predicted mix of stellar types present becomes more complex than the simple initial range from blue supergiant to red dwarf [10].

Given a set of galaxies drawn from a magnitude-limited survey for which redshifts are known [12–14], the simplest estimator of the luminosity function is giving by binning up the as a function of absolute magnitude to investigate the evolution with redshift, this binning can be done separately for shells covering different distance ranges. The correct estimator for density in a given bin is the commonsense sum of the density defined by each object [15–17].

$$\varphi = \sum \varphi i + \sum [V(Zmax) - V(Zmin)i - 1 \tag{1}$$

where Zmax is the smaller of the maximum redshift within which a given object could have been seen and the upper limit of the redshift band under consideration; Zmin is the lowest limit of the band, and V(Z) is the commoving volume out to give redshift over the relevant are of sky.

Image quasars, lets and gravitational arcs are useful in analysis of galaxy evolution. These concepts form the basis for designing the tools for the analysis of issues related to cosmology. Gotter designed a tool to calculate lensing probability based on separation of images related to cosmology. The of this tool is improved by Ostriker in which normal distribution function is utilized.

Major statistical properties of galaxies are initiated by Mao. All these statistical properties are analysed and enhanced with the help of gravitational statistics. Redshift methodology is designed based on these statistical properties and analysed by Mao and Ostriker. Redshift model is used to enhance the lensing probability for the images associated with cosmology. Gravitational lenses are used in the improvement of mean separation cosmological images large mean separation of images [18].

The evolutionary methodology is proposed by Rix et al. in which all the mathematical properties of cosmological images are observed with high accuracy compared to the conventional model. This model can be treated as more practical model in all evolutionary models. Major concept behind all evolutionary models is the lens population based on the merging procedure of cosmological images. There is a strong correlation between source redshift and lens image separation.

This correlation is observed by Park. These observations are described by Golt. Jain et al. also studied the effect of galaxy evolution. Brian et al. studied thoroughly the formation of the galaxies and distribution properties of cosmological images. In this, they studied cosmological images with decaying and constant images [15].

In this paper, lens quasars are used to study the image distributions. The proposed tool is used to evaluate different models associated with galaxy formulation [12].

2 Evolution of Galaxies

Astrophysics-related principles and theorems contributed in the analysis of unsolved problems associated with cosmology. Galaxy formation takes place with complex interaction of materials. There are researchers who believe that galaxies are formed with well-defined principles. In the study of galaxy evolution, formation of elliptical galaxies is still a mysterious problem to handle which is shown in Fig. 1.

A major breakthrough in the analysis of cosmological images is achieved with the help of datasets obtained from deep Hubble space telescope (HST). All the characteristics of HST datasets are analysed for different values of Z. This analysis is carried out by considering $Z < 1$ and $Z > 1$. All HST images are observed by considering redshift analysis. This paper concludes that all the cosmological images which we are observing today are formed by merging of different subsystems shown in Fig. 2.

This suggests that the galaxies we see today could have been assembled from the merging of smaller systems sometime before $Z \sim 1$ which is shown in Fig. 2.

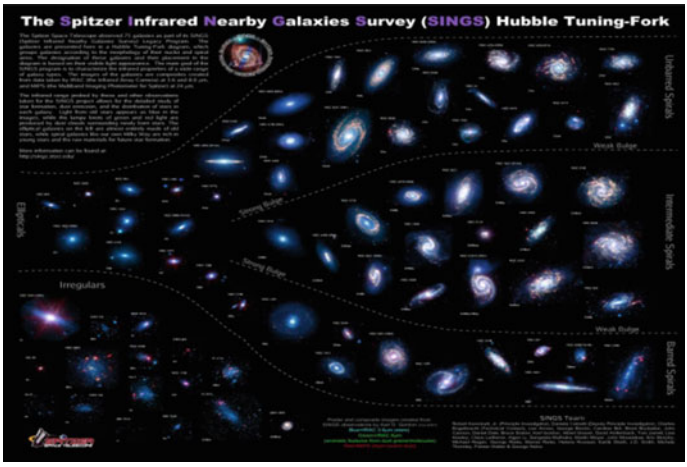


Fig. 1 Evolution of galaxies

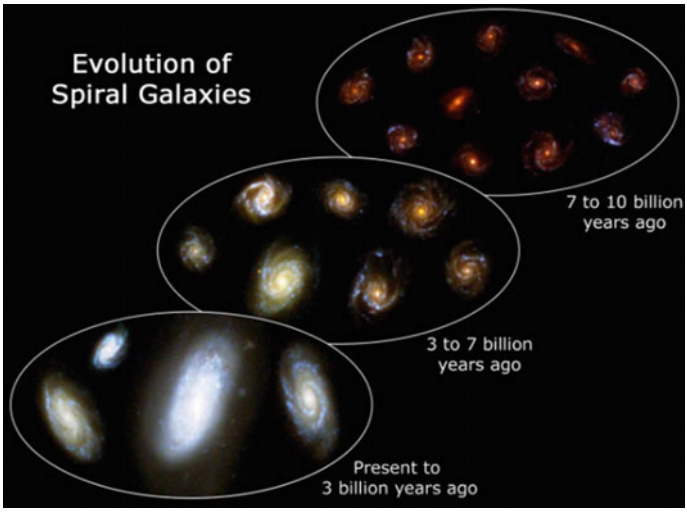


Fig. 2 Evolution of spiral galaxy

3 Basic Equations

The probability of cosmological light rays can be evaluated by the well-known formula given by

$$d\varphi = n(Z)\beta\, cdz/dt \tag{2}$$

Here n(Z) is the parameter to estimate TOG.

All the properties including lensing characteristics of real cosmological images are evaluated with suitable intensity levels by SIS methodology. In this regard, the parameter velocity and accelerated dispersion are used. The velocity and accelerated deflection angle for impact variables are calculated by

$$\Omega = 4\pi\omega^2/C^2 \tag{3}$$

where Ω, W and C are lensing parameters. Cosmological images are used to evaluate angular position and βc (critical angle). Deflection of cosmological image is enumerated by the following formula through SIS.

$$\alpha_c = \alpha\, \mathrm{E}_{MT}/E_{PT} \tag{4}$$

The parameter $\mathrm{E_{MT}}$ and $\mathrm{E_{PT}}$ are calculated with the help of ratio formulas.

$$E_{MT} = d(O, EL), E_{PT} = d(E_L, E_S)$$

is the angular diameter distance between the redshifts Y_1 and Y_2. Then, the critical impact parameter is defined by $a_c = EP + L\beta c$, and the cross section can be calculated with the following formula

$$\gamma = \pi a_c^2 = 16\pi^3 (\mu/\vartheta)^4 (E_{PT} E_{MT}/E_{PT}) \tag{5}$$

4 Proposed Evolutionary Method

The joint probability $d\tau$ of a cosmological lens in the proposed evolutionary model is evaluated by

$$d\tau = \frac{16\pi^3}{CH_0^3} \phi v^4 \tau \left(\alpha + \frac{4}{v+1}\right) f(\delta t)^{\left(1-\frac{4}{v}\right)} (1+Z_l)^3 (D_{OL} D_{LS}/R_0 D_{os})^2 \frac{1}{Ro} \frac{Cdt}{dZl} \tag{6}$$

$$d\tau = \frac{16\pi^3}{CH_0^3} \phi v^4 \tau \left(\alpha + \frac{4}{v+1}\right) f(\delta t)^{\left(1-\frac{4}{v}\right)} (1+Z_l)^3 \tag{7}$$

$$d\tau = \frac{16\pi^3}{GH_0^3} \phi v^4 \tau \left(\alpha + \frac{4}{v+1}\right) G(\delta t)^{\left(1-\frac{4}{v}\right)} (1+X_l)^3 \tag{8}$$

where f(δt) in the proposed model can be treated as the functional form.

The evolutionary model uses the function as f(δt), fast merging uses the form Exp(QH0), and VG model uses (1 + z)1.5 as f(δt). Models like slow merging and mass accretion use different kinds of functions.

Using the relation luminosity–velocity relation

$$\varphi = \frac{\beta\theta}{8\pi (v/c)^2} \tag{9}$$

$$\frac{L}{L^*} = \left(\frac{v}{v^*}\right)^v \tag{10}$$

We have

$$\frac{L}{L^*} = \left(\frac{D_{OS}}{D_{LS}} \phi\right)^{v/2} \tag{11}$$

Equation (11) represents the angular separation.

Angular separation can be measured with the help of φ and $\varphi + d\varphi$. The angle between these two parameters is measured with dZl value.

$$\frac{d^2\tau}{\partial x d\phi} dYl = g(1+Y)^3 \tag{12}$$

$$d\beta = \frac{16\pi^3}{fg_0^3} \chi v^4 \tau \left(\chi + \frac{4}{v+1}\right) F(\delta t)^{\left(1-\frac{4}{v}\right)} (1+\delta_l) \tag{13}$$

mass acceleration model basically depends on Eq. (13), whereas the proposed evolutionary model is a special case which deals with parameter F in Eq. (13). Redshift theorem is based on angular separation of cosmological images. It makes use of the phase shifting and scalar invariance properties of the evolutionary model. Angular distribution can be enhanced by integrating Eq. (12).

$$\frac{\partial^2 \Omega}{\partial v^2} = <B> \int_0^{Zs} \frac{\partial^2 \tau}{\partial x d\phi} dZl \tag{14}$$

Quasar samples of cosmological images are obtained by integrating Eq. (14) with respect to evolution parameter. Same methodology is applicable in the case of HST quasar sample acquisition procedure. HST quasar is used in the evaluation of average bias which in this case happens to be 9.34. All the cosmological images associated with average bias are shown in Fig. 3.

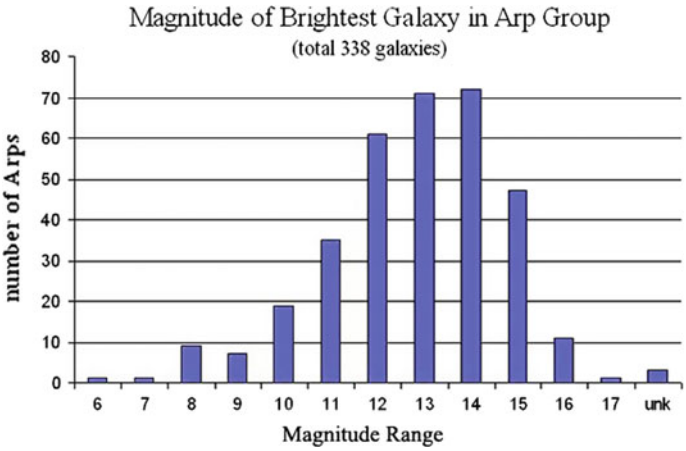


Fig. 3 Relation between magnitude range and Arp groups

4.1 The Non-evolutionary Model

In non-evolutionary methodology, optical depth is calculated with the following equation.

$$d\tau = \frac{16\pi^3}{fH_0^3}\phi v^4 \tau\left(\beta + \frac{4}{v+1}\right) G(\delta t)^{\left(1-\frac{4}{v}\right)}(1+\delta_l)^3 + \frac{1}{Ro}\frac{cdt}{dZl} \tag{15}$$

Equation (15) gives the relation between calculation procedure of optical depth and redshift non-evolutionary model. If v = 4, then we can easily establish relation between redshift and optical depth. Optical depth lensing is measured with first-, second- and third-order partial derivatives. All these derivatives are obtained by partially differentiating with respect to evaluation parameters under consideration. Non-evolutionary model-based angular separation can be calculated by measuring the angle between lens-related parameters Φ and Φ + dΦ.

$$\frac{d^2\tau}{dZcd\phi} = \frac{16\pi^3}{fH_0^3}\phi v^4 \tau\left(\beta + \frac{4}{v+1}\right) G(\delta t)^{\left(1-\frac{4}{v}\right)}(1+\delta_l)^3 \tag{16}$$

$$\frac{d^2\tau}{dZcd\phi} = \frac{16\pi^3}{fH_0^3}\phi v^4 \tau\left(\beta + \frac{4}{v+1}\right) G(\delta t)^{\left(1-\frac{4}{v}\right)}(1+\delta_l)^3 + \frac{\beta\theta}{8\pi(v/c)^2} \tag{17}$$

Equations (16) and (17) give the relation between dispersion velocity and dispersion acceleration.

5 Results and Discussion

When the Gibbson-Melinda index v cannot be obtained from 4 and 6 but this index can be derived from Eqs. (16) and (17).

Cosmological variables Ω = 0.3 and Λ = 0.7 are evaluated to obtain cosmological parameters.

Snapshot survey of cosmological HST images is used to understand the significant image separation. Four lensed quasar cosmological images are used to construct histograms.

The k96 set of parameters shift the peak of the image separation distribution function (dn/dt) for all the evolving models towards smaller angle of separations. The larger angle separations are not explained by these parameters. With LPEM parameters the mass accretion model and to a lesser extent, the non-evolution model seems to have reasonable success in explaining the observations as shown in Figs. 4 and 5.

All the experiments are carried with various classes such as spiral, elliptical, irregular and hybrid combinations. Various parameters like accuracy, sensitivity,

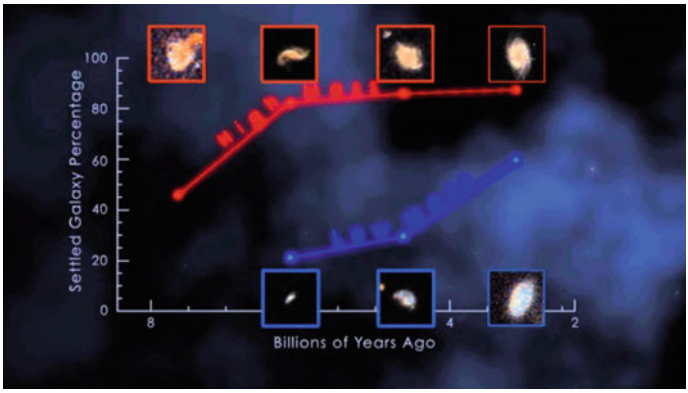


Fig. 4 History of galaxy evaluation

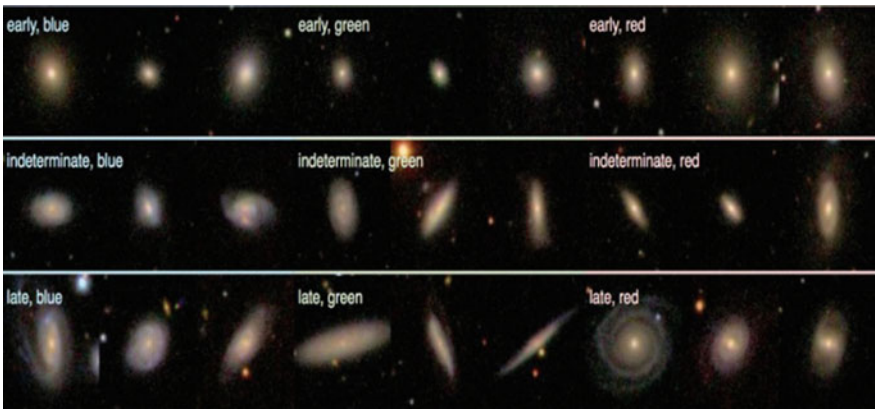


Fig. 5 Collection of galaxies in different forms

specificity, time and number of iterations are calculated and shown in Tables 1, 2, 3, 4, 5 and 6. All the results are also analysed with respect to hybrid tier, irregular first tier, spiral first tier and elliptical tier. (Fig. 6)

Table 1 Average of two-class GA-SVM: Experiments 1–3

Classes	Accuracy	Sensitivity	Specificity	Time	Iterations
Spiral	96.8	97.8	95.6	4.0	123
Elliptical	71.6	71.4	71.8	10.6	118
Regular		84.5	93.6	6.9	134
Spiral elliptical	88.5	68.4	82.4	9.5	120
Spiral regular	79.5	69.5	76.2	9.4	132
Spiral regular	82.6	73.5	68.4	10	132

Table 2 Average of three-class GA-SVM: Experiments 4–6

Classes	Accuracy	Sensitivity	Specificity	Time	Iterations
Spiral	73.1	94	86.2	5.0	132
Elliptical	74.3	75	87.3	11.2	102
Regular		85	54.6	7.3	141
Spiral elliptical	67.2	71	90.2	10.4	132
Spiral regular	74.5	73	88.1	11.5	123
Spiral regular	63.5	75	72	12	126

Table 3 Confusion matrix, spiral first tier

Classes	Accuracy	Sensitivity	Specificity	Time	Iterations
Spiral	73.1	89.2	75.6	5.1	130
Elliptical	74.3	79.1	66.8	11.4	120
Regular	67.2	82.1	83.6	7.9	143
Spiral elliptical	71.5	73.5	72.4	8.5	132
Spiral regular	81.3	49.5	86.2	11.4	142
Spiral regular	78.1	63.5	78.4	12	162

Table 4 Confusion matrix, elliptical first tier

Classes	Accuracy	Sensitivity	Specificity	Time	Iterations
Spiral	67.1	88.2	65.6	9	120
Elliptical	84.3	89	76.8	12	180
Regular	77.2	72	63	8.2	123
Spiral elliptical	51.5	77.6	62	9.2	162
Spiral regular	71.3	59.4	76	10.1	192
Spiral regular	68	83.8	73	11.3	122

Table 5 Confusion matrix irregular first tier

Classes	Accuracy	Sensitivity	Specificity	Time	Iterations
Spiral	63.1	84	91	6.0	122
Elliptical	64.3	65	97	12.2	112
Irregular		86	64	8.4	121
Spiral elliptical	57	82	89	11.4	12
Spiral irregular	64.5	63	78	12.1	113
Spiral irregular	73.5	89	53	13.1	126

Table 6 Confusion matrix hybrid tier

Classes	Accuracy	Sensitivity	Specificity	Time	Iterations
Spiral	77.1	84	78.2	6	132
Elliptical	65.3	95	97.3	1.2	102
Irregular	65.1	85	55.6	7.3	141
Spiral elliptical	49.2	77	11.2	10.4	132
Spiral irregular	72.5	65	88.1	11.5	123
Spiral irregular	67	65	62	12	126

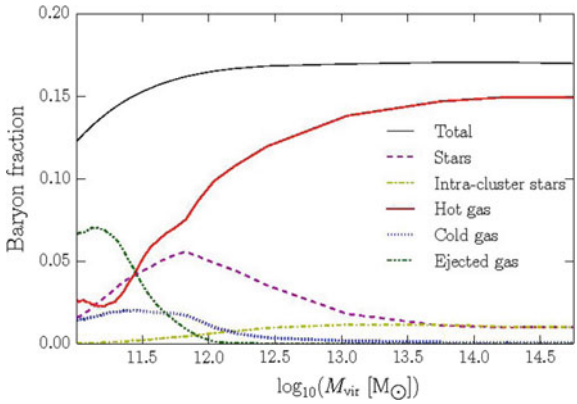


Fig. 6 Application of GA-SVM on database of galaxy dataset

6 Conclusion

It is clear that to reach at a firm conclusion about galaxy evolution, more reliable and valid set of schechter parameters are badly needed. The X-ray observations by CXO (Chandra X-ray Observatory) should be especially helpful. An increase in the overall number of lensed objects of all wavelengths is also expected in the next few years. The knowledge of cosmological parameters and galaxy distribution(schechter) parameter will mutually refine each other as more results becomes available and then it will be possible to say with some certainity which of the various galaxy evolution models are favoured by observations and which are ruled out.

References

1. Raja Sekar, M., et al.: Tongue image analysis for hepatitis detection using GA-SVM. Indian J. Comput. Sci. Eng. **8**(4) (2017)
2. Raja Sekar, M., et al.: Mammogram images detection using support vector machines. Int. J. Adv. Res. Comput. Sci. **8**(7), 329–334 (2017)

3. Raja Sekar, M., et al.: Areas categorization by operating support vector machines. ARPN J. Eng. Appl. Sci. **12**(15), 4639–4647 (2017)
4. Raja Sekar, M.: Diseases identification by GA-SVMs. Int. J. Innov. Res. Sci. Eng. Technol. **6** (8), 15696–15704 (2017)
5. Raja Sekar, M.: Classification of synthetic aperture radar images using fuzzy SVMs. Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET) **5**(8), 289–296, **45** (2017)
6. Raja Sekar, M.: Breast cancer detection using fuzzy SVMs. Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET) **5**(8), 525–533 (2017)
7. Raja Sekar, M.: Software metrics in fuzzy environment. Int. J. Comput. Math. Sci. (IJCMS) **6** (9) (2017)
8. Raja Sekar, M.: Interactive fuzzy mathematical modeling in design of multi-objective cellular manufacturing systems. Int. J. Eng. Technol. Comput. Res. (IJETCR) **5**(5), 74–79 (2017)
9. Raja Sekar, M.: Optimization of the mixed model processor scheduling. Int. J. Eng. Technol. Comput. Res. (IJETCR) **5**(5), 74–79 (2017)
10. Raja Sekar, M.: Fuzzy approach to productivity improvement. Int. J. Comput. Math. Sci. **6**(9), 145–149 (2017)
11. Raja Sekar, M., et al.: An effective atlas-guided brain image identification using X-rays. Int. J. Sci. Eng. Res. **7**(12), 249–258 (2016)
12. Raja Sekar, M.: Fractional programming with joint probability constraints. Int. J. Innov. Adv. Comput. Sci. **6**(9), 338–342 (2017)
13. Raja Sekar, M.: Solving mathematical problems by parallel processors. Current Trends Technol. Sci. **6**(4), 734–738 (2017)
14. Raja Sekar, M., et al.: Bullet classification system based on image processing and support vector machines **5**(3), 1006–1009 (2017)
15. Aggarwal, K.K., Gupta, J.S., Misra, K.B.: A new heuristic criterion for solving a redundancy optimization problem. IEEE Trans. Reliab. **24**(1), 86–87 (1975)
16. Beraldi, P., Bruni, M.: An exact approach for solving integer problem under probabilistic constraints with random technological matrix. Ann. Oper. Res. **177**(1), 127–137 (2010)
17. Bulfin, R.L., Liu, C.Y.: Optimal allocation of redundant components for large systems. IEEE Trans. Reliab. **34**(3), 241–247 (1985)
18. Kamgar-Parsi, B., Kamgar-Parsi, B., Jain, A., Dayhoff, J.: Aircraft detection: a case study in using human similarity measure. IEEE Trans. Pattern Anal. Mach. Intell. **23**(12) (2001), 1404–1414 (2016)

Energy Distribution and Coherence-Based Changes in Normal and Epileptic Electroencephalogram

Revati Shriram, V. Vijaya Baskar, Betty Martin, M. Sundhararajan and Nivedita Daimiwal

Abstract In endeavor toward better understanding of brain functions, the analysis of information transfer between the various brain lobes plays a crucial role. Electroencephalogram (EEG) is an electrical brain signal in microvolts, which provides unique and important information about the brain dynamics. Frequency of EEG signal lies between 0 and 100 Hz. In epileptic or seizure related studies, decomposition of EEG signal into various frequency sub-bands such as α, β, δ, θ, and γ is essential. EEG plays a key role in diagnosis of neurological disorders such as epilepsy. In this paper, we explore decomposition of EEG by db18 wavelet, power spectral density, coherence, energy distribution, and empirical cumulative distribution function of EEGs. This work was carried out to study the changes in the normal and epileptic EEGs with respect to PSD, coherence, energy, and ECDF to check the suitability of these parameters as an input to the classifiers. The methodology was applied mainly to three groups consisting of male and females between the age group of 01–107 years: (1) healthy subjects (normal), (2) subjects with focal seizures, (3) subjects with generalized seizures. The work was carried out on the signals obtained from real subjects to study the EEG-based brain connectivity analysis. It was observed that PSD and coherence study related to the sub-bands reveal more accurate information than the study of complete EEG with or without the seizures.

R. Shriram (✉)
Department of ECE, Sathyabama Institute of Science & Technology, Chennai, India
e-mail: revatishriram@yahoo.com

V. V. Baskar
Sathyabama Institute of Science & Technology, Chennai, India

B. Martin
Shanmugha Arts, Science, Technology & Research Academy, Thanjavur, India

M. Sundhararajan
Bharath Institute of Higher Education & Research, Chennai, India

R. Shriram · N. Daimiwal
Cummins College of Engineering for Women, Pune, India

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_61

1 Introduction

EEG is postsynaptic potentials formed from a complex process of firing of cortical pyramidal cells in a nonlinear brain system. An electrical signal, EEG captured from surface of the head represents the entire brain dynamics. Electroencephalogram is a superimposition of various processes in the brain and is a very important state indicator of a brain such as epilepsy, sleep, awake, tired. Amplitude of this electrical signal is in few 'microvolts' (~200 mV). The amplitude, frequencies, and characteristics of an EEG signal change from one state to the other and with age. Mainly EEG signal is decomposed into five sub-bands: δ (0–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30 Hz and above) [1].

1.1 Types of Seizures

Seizure is abnormal type of EEG signal with, high frequency and high or low magnitude. They can be classified into two main types: generalized seizures and partial or focal seizures [2]. The work was carried out by the authors on generalized epileptic, focal epileptic, and normal subjects.

2 Methods

Traditional Fourier transform (FT) works well for the stationary signal. Authors have used wavelet filter for the sub-band decomposition for its various advantages like flexibility, time–frequency localization, and scale-space analysis [3].

2.1 Wavelet Transform

In the wavelet transform (WT), variable window size can be applied over a length of a signal. This is the reason why wavelet transform is widely used for the feature extraction technique in the non-stationary signals, such as EEG. Two types of WT are continuous and discrete wavelet transforms [4]. Equations. 1 and 2 show the CWT and DWT equations, respectively.

Continuous Wavelet Transform (CWT)

$$X(x,y) = \frac{1}{\sqrt{b}} \int_{-\infty}^{\infty} X(t)\varphi\left(\frac{t-x}{y}\right)dt \tag{1}$$

where 'ψ' is the mother wavelet, 'x' is a time shift, and 'y' modulates the width.

Discrete Wavelet Transform (DWT)

$$\Psi_{(x,y)}(t) = 2^{x/2}\psi\left(2^{-x/2}(t-y)\right) \quad (2)$$

Selecting a subset of scales 'a' and positions 'b' of the mother wavelet 'ψ(t)' plays very important role in DWT [5]. Bior3.7 is a biorthogonal wavelet used for the EEG signal decomposition. The property of perfect reconstruction/synthesis of a signal using biorthogonal family makes it a very popular wavelet. General criteria for selection of wavelet are the shape of mother wavelet and ability of a wavelet to analyze the signal [3].

2.2 EEG Signal Decomposition

Analysis and synthesis of a signal are known as decomposition and reconstruction of a signal, respectively. EEG signal analysis is carried out with the help of multi-resolution analysis using complementary low-pass (approximate component) and complementary high-pass (detailed component) filters. Discrete wavelet transform (DWT) is used for the analysis/decomposition. After first level analysis/decomposition of EEG signal gives the high- and low-level components. Again, low-level component (i.e., approximate component) is further analyzed to obtain high- and low-level components. Sampling frequency of the signal and bandwidth decides the number of decomposition/analysis levels. db 18 wavelet is used for the decomposition of EEG signals. Table 1 shows the various levels of decomposition with LPF and HPF.

2.3 Power Spectral Density (PSD)

It is the strength of the variations in energy as a function of frequency. Weak variations and strong variations with respect to the frequencies are reflected in PSD

Table 1 EEG sub-band decomposition and frequency bands

Decomposition levels	Approximate/ detailed component	Filter	Band of frequency
First	A11/D12	LP & HP	0–125 Hz/ gamma-three 125–250 Hz
Second	A21/D22	LP & HP	0–62.5 Hz/ gamma-two 125–62.5 Hz
Third	A31/D32	LP & HP	0–31.25 Hz/ gamma-one 31.25–62.5 Hz
Fourth	A41/D42	LP & HP	0–15.62 Hz/ beta 15.62–31.25 Hz
Fifth	A51/D52	LP & HP	0–7.81 Hz/ alpha 7.81–15.62 Hz
Sixth	A61/D62	LP & HP	delta 0–3.90 Hz/ theta 3.90–7.81 Hz

[6]. In PSD of a signal, Sx(f) (as shown in Eq. 3) is the Fourier transform (FT) of the autocorrelation function R(τ), of the signal assuming the signal in a wide-sense stationary random process.

$$Sx(f) = \int_{-\infty}^{\infty} Rx(\tau) e^{-2\pi i f \tau} dt \tag{3}$$

The DTFT of a covariance matrix is also the power spectral density given by Eqs. 4 and 5.

$$\varnothing(w) = \sum_{k=-\infty}^{\infty} r(k) e^{-iwk} \tag{4}$$

$$\varnothing(w) > 0 \quad \text{for all } w \tag{5}$$

Integrating over positive and negative frequencies can be used to get the power of the signal in specific frequency band. Equation 6 shows the PSD estimation by periodogram method.

$$\widehat{\varnothing_p}(w) = \frac{1}{N} \left| \sum_{t=1}^{N} y(t) e^{-iwt} \right|^2 \quad \text{Periodogram} \tag{6}$$

Welch method (as shown in Eq. 7) is the refined Bartlett method of PSD estimation. 'P' (as shown in Eq. 8) denotes the power of temporal window v(t) [3].

$$\widehat{\varnothing_j}(w) = \frac{1}{MP} \left| \sum_{t=1}^{M} v(t) y_j(t) e^{-iwt} \right|^2 \quad \text{Welch} \tag{7}$$

$$P = \frac{1}{M} \sum_{t=1}^{M} |v(t)|^2 \tag{8}$$

Spectral representation is useful for describing random stationary processes in frequency domain [7–9].

2.4 Coherence

It is a measure of phase related stability between two different time series. It is the logical connection or neural consistency [10, 11]. Coherence value becomes '1' for the two signals with constant phase difference and for the two signals with random phase difference, coherence value becomes '0'. Auto-frequency coherence or coherence means estimation is carried out in the same frequency band [12]. The degree of

relationship/coupling between two systems is completely understood with the help of coherence and with knowledge of its frequency structure over a longer period of time [6, 13, 14]. Coherence (as shown in Eq. 9) reveals the details of temporal relationship between the coupled systems sensitively and immediately [15, 16].

$$C_{ab}(w) = \frac{|G_{ab}(w)|^2}{G_{aa}(w) * G_{bb}(w)} \tag{9}$$

where $G_{ab}(w)$: cross-PSD of signals 'a' and 'b', $G_{aa}(w)$ and $G_{bb}(w)$: auto PSD of 'a' and 'b' (as shown in Eq. 10), $C_{ab}(w)$: MS coherence between signals 'a' and 'b'. When $G_{ab} = 0$; $G_{aa} = 0$; $G_{bb} = 0$, then C_{ab} becomes zero.

$$G_{aa}(w) = |\widehat{a}(w)|^2 = \widehat{a}(w)\overline{\widehat{a}(w)} \;\&\; G_{ab}(w) = \widehat{a}(w)\overline{\widehat{b}(w)} \tag{10}$$

When signals 'a' and 'b' are coherent then the value of C_{ab} becomes closer to 1 or one. For non-coherent 'a' and 'b' signals, the C_{ab} becomes zero [17]. Magnitude-squared coherence (MSC) is implemented by the authors in this work.

2.5 Cumulative Distributive Function (CDF)

The D(x) a distribution function is also known as cumulative frequency function (CFF) or cumulative distribution function [18]. It is the probability that an X variate takes on a value ≤ number x. D(x) (as shown in Eq. 11) is related to a P(x)-probability density function (PDF) by the following equation [19].

$$D(x) = P(X \le x) = \int_{-\infty}^{\infty} P(\xi) d(\xi) \tag{11}$$

So P(x) is the derivative of the distribution function, as shown in Eq. 12.

$$P(x) = D'(x); \quad D'(x) = [P(x)]^x = P(x) - P(-\infty) = P(x) \tag{12}$$

Empirical cumulative distribution function (ECDF) $\hat{F}_n(x)$ for the sub-band related coherence data was applied (as shown in Eq. 13). 'ECDF' estimations provide more accurate results for the larger data set [19].

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \le x\} \tag{13}$$

where $1\{X_i \le x\} = \{1, if\ X_i \le x, \qquad 1\{X_i \le x\} = \{0, otherwise$

The plot is useful in examination of distribution of sampled data. It helps us in understanding the change in coherence values for the normal and abnormal/ epileptic EEGs [20, 21].

2.6 Energy Distribution

As per the 'Parseval's Theorem,' Fourier transform is unitary (i.e., unitary operator operating on Hilbert's space) and the integral of square of its transforms is equal to the integral of square of function.

$$EDi = \sum_{j=1}^{N} |Dij|^2, \quad i = 1, \ldots, l \tag{14}$$

$$EAl = \sum_{j=1}^{N} |Alj|^2 \tag{15}$$

where i = 1, …, l are the levels of decomposition by the wavelet. A number of coefficients of approximate and detail component are denoted by 'N'. Energy of the detail component of the signal is denoted by EDi (as shown in Eq. 14), whereas energy of the approximate component of the signal is denoted by EAl (as shown in Eq. 15) [22].

3 Details of Database Used

Freely available database of EEG signals published by Karunya University, Coimbatore, India, is used by the authors. Database consists of two types of epileptic EEGs, such as with focal seizures and with generalized seizures. Age group of patients varies between 1 and 107 years. It is an eighteen-channel EEG signal recording with sampling frequency of 256 Hz [23]. Data collection of normal EEG signal was carried at Cummins College of Engineering, Pune, India. Twenty subject volunteers participated in data collection, of both genders, aged between 18 and 50 years. Study was carried out on total 174 subjects. Out of them, 80 subjects had focal seizures, 80 subjects had generalized seizures, and 14 subjects with normal EEG.

4 Results

EEG signal was decomposed into its sub-bands using db18 wavelet. Decomposition of signal was carried out to study the estimation of PSD and coherence in each sub-band. Figure 1 shows the decomposed EEG signal into various sub-bands. PSD

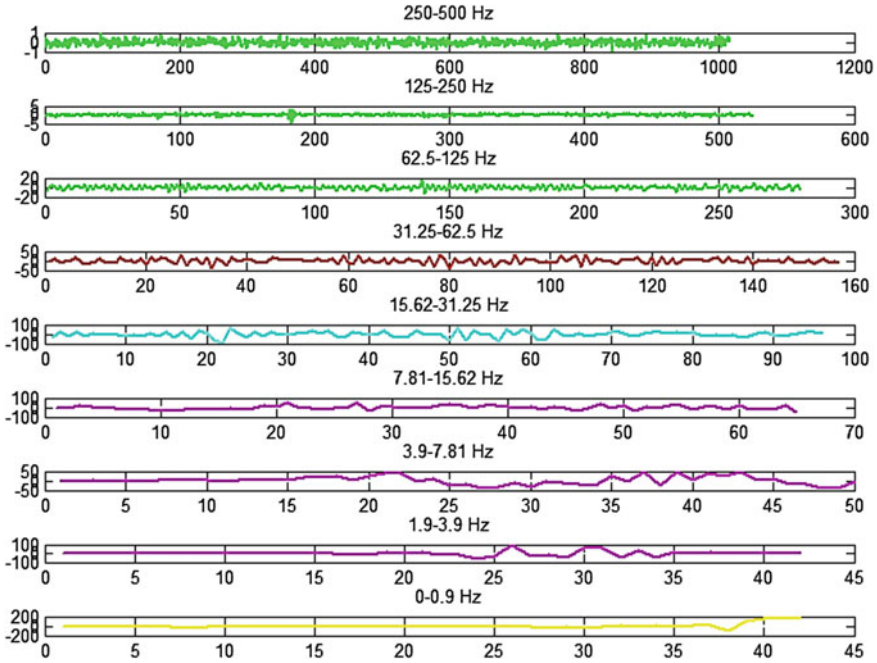


Fig. 1 Decomposed EEG signal by using db18

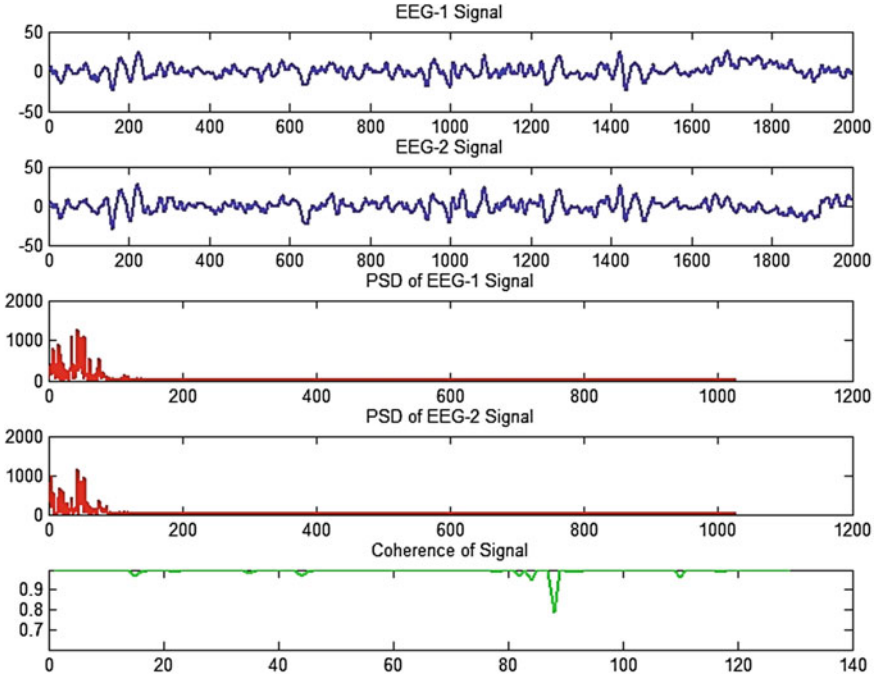
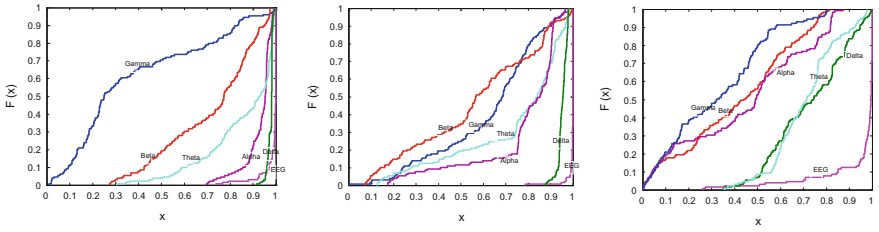


Fig. 2 PSD and coherence of EEG with focal seizures

Table 2 Mean PSD and mean coherence for various types of EEG signals

Parameters	EEG waves	Normal	Focal epileptic	Generalized epileptic
Mean PSD	δ Wave	195881.67	441.42	150.22
	θ Wave	81171.80	428.37	340.89
	α Wave	20322.49	343.24	244.88
	β Wave	2681.47	117.77	116.06
	γ Wave	333	10.35	10.25
	EEG signal	6496.5	42.18	28.54
Mean coherence	δ Wave	0.8725	0.7734	0.79
	θ Wave	0.8632	0.7913	0.78
	α Wave	0.6984	0.7881	0.77
	β Wave	0.4686	0.7086	0.75
	γ Wave	0.3112	0.6909	0.71
	EEG signal	0.9678	0.9795	0.97



(a) ECDF Plot for sub -bands of Normal EEG Signal

(b) ECDF Plot for sub - bands of Focal Seizures in EEG

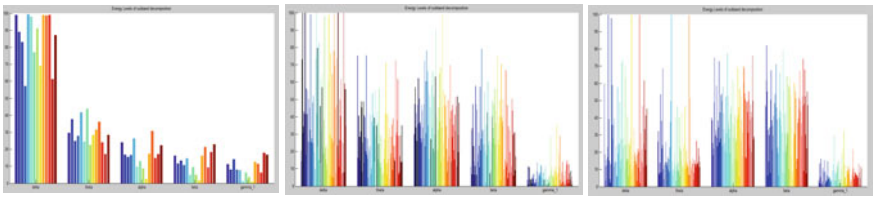
(c) ECDF Plot for sub - bands of Generalized Seizures in EEG

Fig. 3 Empirical CDF plot for normal and epileptic EEG signals

and coherence were calculated for normal and diseased EEGs. Figure 2 shows the plot of PSD and coherence.

Table 2 shows the mean power spectral density and coherence for each sub-band of EEG.

Delta band (1–3 Hz): Mean value of delta band power and coherence for normal EEG are observed to be very higher than the epileptic EEGs. Theta band (4–7 Hz): Mean value of PSD and coherence is too higher in normal EEG than the EEG with epileptic seizures. Alpha band (8–12 Hz): Mean of PSD and coherence is almost same in EEG with seizures, while the value of PSD is much lower as compared to normal EEGs, maintaining the values of coherence almost same. Beta band (13–20 Hz): Mean value of PSD for normal EEG is higher, and coherence is lower than epileptic EEGs, whereas PSD and coherence for epileptic EEGs are almost same. EEG Signal: Mean value of PSD for complete normal EEG signal (without decomposition) is way higher than the epileptic EEGs. Coherence was found to be 0.9 (highly coherent) in normal and deceased EEGs. Figure 3 shows the nature of



(a) Energy Plot for sub - bands of Normal EEG Signal

(b) Energy Plot for sub -bands of Focal Seizures in EEG

(c) Energy Plot for sub - bands of Generalized Seizures in EEG

Fig. 4 Energy plot for normal and epileptic EEG signals

empirical cumulative distributive function for EEG signal (with and without seizures) and five sub-bands. Normal and epileptic EEG is decomposed by using wavelet into the above-mentioned sub-bands, and then percent energy in each sub-band is calculated.

Figure 4 shows the energy distribution in the various sub-bands of normal and epileptic EEG. Mean coherence in delta and theta bands is in upper higher band (~ 0.8) and for alpha, beta and gamma bands, it was found to be weak (~ 0.5) for normal EEG, whereas in epileptic EEGs (focal or generalized) mean coherence in all the sub-bands is almost same.

From Fig. 3, it can be seen that the nature of ECDF in various sub-bands of normal EEG is very much different from the nature of ECDF in various sub-bands of epileptic EEGs. From Fig. 4, we can see that the energy level in delta sub-band is much higher than any other sub-band for the normal EEG, whereas, for epileptic EEGs, energy levels in all the sub-bands are almost same. No prominent variation in the energy w.r.t. the sub-bands was observed in EEG with focal or generalized seizures.

5 Discussion

According to World Health Organization (WHO), approximately 1% of the world population and out of them 15 million people affected by epilepsy in southeast Asia region. It is most common and one of the serious chronic neurological disorders, which causes electrical disturbance (loss of synchronization) in the brain leading to some of the symptoms listed in Table 1. Authors have carried out the work in the multivariate environment. In multivariate analysis, more than one/single channel of EEG is taken into account simultaneously. This is very useful in studies related to the interactions/ synchronizations between various brain regions leading to seizures. It was observed that EEG sub-band related studies reveal more important information than the study of complete EEG signal. Empirical cumulative distribution function indicates the amount of fluctuations in coherence related data. These

feature vectors are calculated in each sub-band of normal and epileptic EEGs. Changes in the PSD and coherence value were observed in each sub-band for normal and deceased EEG signal.

6 Conclusion

Epilepsy is a very widely found neuronal disease. Epileptic seizures can be of various types, and these are the transient electrical disturbance in the brain, denoted by larger amplitude and frequency EEG signal. This disorder is more common in young children and the adults over the age of 60 years. The work was carried by the authors to check whether epileptic EEG differs from the normal EEGs based on the features such as coherence, energy distribution and the empirical cumulative distribution function. It was observed that the coherence, energy, and ECDF-related changes in complete EEG signal (without decomposition) do not vary with the presence of epilepsy. So, during the epileptic studies instead of complete EEG, various sub-bands of EEGs should be studied. It was observed that the mean coherence in all the sub-bands of epileptic EEG is almost same, whereas in normal EEG signals mean coherence in delta, theta, and alpha is higher than rest of the sub-bands. ECDF-related changes are more prominent in higher frequency band (gamma) of epileptic and normal EEG signal, whereas theta and beta bands do not reveal any change in ECDF nature with respect to any type of seizure. It can be seen from Fig. 4 that the percent energy in the delta band is highest, followed by theta and alpha bands in the normal EEG, whereas percent energy in all the sub-bands is almost same in the EEG with seizures. It was observed that percent energy and mean coherence in the higher frequency band (gamma) are always lower as compared to any other sub-bands with or without seizures. This shows that these three parameters can be used for the classification of EEG signals with or without seizures.

Declaration: Authors have taken the consent from the concerned authority to use the materials, etc., in the paper. Authors will be solely responsible if any issues arise in the future with regard to this.

References

1. Sanei, S., Chambers, J.: EEG signal processing. In: Centre of Digital Signal Processing, Cardiff University, Wiley, UK (2007)
2. Gauthamdas: A Neuro-Behavioral Medicine Clinic. Dr. Gauthamadas's Neuro Centre www.docgautham.com (2017)
3. Olkkonen, H.: Discrete Wavelet Transforms—Biomedical Applications. ISBN 978-953-307-654-6. www.intechopen.com (2011)
4. Rafiee, J., Rafiee, M., Prause, N., Schoen, M.: Wavelet basis functions in biomedical signal processing. Expert Syst. Appl. **38**, 6190–6201 (2011)

5. Shaker, M.: EEG waves classifier using wavelet transform and fourier transform. World Academy of Science, Engineering and Technology, Int. J. Med. Health, Pharm. Biomed. Eng. **1**(3) (2007)
6. Mihandoust, S., Amirani, M.: Epileptic seizure detection using GARCH model on EEG signals. In: 1st International eConference on Computer and Knowledge Engineering (ICCKE 2011), Mashhad, Iran (2011)
7. Sulaiman, N., Taib, M., Lias, S., Murat, Z.: Novel methods for stress features identification using EEG signals. IJSSST. a. 12. 01.04, ISSN: 1473-804x online (2004)
8. Malar, E., Gauthaam, M., Chakravarthy, D.: A novel approach for the detection of drunken driving using the power spectral density analysis of EEG. Int. J. Comput. Appl. **21**(7) 0975–8887 (2011)
9. Li, X.: Multiple significance tests and their relation to P-values. College of Graduate Studies and Research, University of Saskatchewan, Saskatoon (2008) (Thesis Submitted)
10. Zavaglia, M., Astolfi, L., Babiloni, F., Ursino, M.: The effect of connectivity on EEG rhythms, power spectral density and coherence among coupled neural populations: analysis with a neural mass model. IEEE Trans. Biomed. Eng. **55**(1) (2008)
11. Adeli, H., Ghosh-Dastidar, S., Dadmehr, N.: A wavelet-chao methodology for analysis of EEGs and EEG sub-bands to detect seizure and epilepsy. IEEE Trans. Biomed. Eng. **54**(2) (2007)
12. Benesty, J., Chen, J., Huang, Y.: Estimation of the coherence function with the MVDR approach. IEEE ICASSP (2006)
13. Ombao, H., Bellegem, S.: Evolutionary coherence of nonstationary signals. IEEE Trans. Signal Process. **56**(6) (2008)
14. Stoica, P., Moses, R.: Spectral Analysis of Signals, p. 07458. Prentice Hall, Upper Saddle River (2004)
15. Correa, A., Leber, E.: An automatic detector of drowsiness based on spectral analysis and wavelet decomposition of EEG records. In: 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina (2010)
16. Huang, R., Jung, T., Makeig, S.: Tonic Changes in EEG Power Spectra During Simulated Driving. HCI International, San Diego (2009)
17. Golińska, A.: Coherence function in biomedical signal processing: a short review of applications in neurology, cardiology and gynecology. Stud. Logic Grammar Rhetoric **25**(38) (2011)
18. Umbgen, L., Rufibach, K.: Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. Bernoulli **15**(1) 40–68 (2009)
19. Bonnery, D., Breidt, F., Coquet, F.: Uniform convergence of the empirical cumulative distribution function under informative selection from a finite population. Bernoulli (2012)
20. Signal Processing Toolbox™User's Guide, The MathWorks, Inc. 3 Apple Hill Drive, Natick, 01760-2098
21. Schaffoner, M., Andelic, E., Katz, M., Kruger, S., Wendemuth, A.: Memory-Efficient Orthogonal Least Squares Kernel Density Estimation using Enhanced Empirical Cumulative Distribution Functions. Otto- von-Guericke-University, Magdeburg, Germany (2007)
22. Omerhodzic, I., Avdakovic, S., Nuhanovic, A., Dizdarevic, K.: Energy distribution of EEG signals: EEG signal wavelet-neural network classifier. Int. J. Biol. Life Sci. **6**(4), 210–216 (2010)
23. Selvaraj, T., Ramsamy, B., Jeyaraj, S., Suviseshamuthu, E.: EEG database of seizure disorders for experts and application developers. Clin. EEG Neurosci. **45**(4), 304–309 (2014)

Welch's Power Spectral Density of Cranial PPG Signal Using AVR ATmega 8535 Microcontroller

Nivedita Daimiwal, S. Poornapushpakala, Betty Martin, M. Sundararajan and Revati Shriram

Abstract Functional near-infrared spectroscopy is an optical non-invasive technique for measurement of neural activity and hemodynamic response and has a potential for brain mapping. In this work, we aimed to develop a system to capture the cranial photoplethysmogram (CPPG) using IR source (860 nm) and detector (OPT 101). AC excitation for the IR source in the range of 1−2 MHz plays a major role in the CPPG sensor. Brain functional activity in prefrontal lobe is detected by placing the sensor on the forehead. The CPPG signal is captured with eyes open (EO) and eye blinking (EB) activity for various emotions (happy and sad) on prefrontal lobe. Data acquisition is done using AVR ATmega 8535 microcontroller at a sampling rate of 500 Hz. The data are acquired from subjects in the age group of 20−60 years. Using Daubechies 9, six-level wavelet decomposition of CPPG signal is performed, and spectral analysis of each level is done using Welch's method. From the spectral analysis, it is found the centre frequency of A_6 is at 1.953 Hz. The higher frequency part of the signal is found in D_4 scale and the centre frequency is 17.58 Hz. All other scales D_3−D_1 show predominantly the noise part of the signal. Statistical features of A_6 and D_6 decomposition levels are important for analysis and classification of activity.

N. Daimiwal (✉) · S. Poornapushpakala
Sathyabama Institute of Science and Technology, Chennai, India
e-mail: nivedita.daimiwal@gmail.com

B. Martin
Sastra University, Thanjavar, India

M. Sundararajan
Bharat University, Chennai, India

R. Shriram
Cummins College of Engineering for Women, Pune, India

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_62

1 Introduction

Optical CPPG technique is a optical sensor with an infrared source and photodiode as a detector. The detector senses the intensity of transmitted or reflected light associated with changes in blood volume. For different optical wavelength, the sensitivity of blood and tissue component is different. The selection of source wavelength is important, and it should have greater absorption for blood as compared to tissue. IR and red IR are used as light sources because absorptivity for oxyhemoglobin and deoxyhaemoglobin.

During the systole heart pumps the blood, the blood volume increases resulting in the more absorption of light. The blood travels back to the heart through venous network leading to decrease in blood volume. The recorded PPG signal is a pulsatile physiological waveform that shows cardiac synchronous changes in the blood volume with each heart beat. PPG is widely used for real-time monitoring of heart rate (HR), blood pressure, respiration, vascular assessment and autonomic function assessment like HR variability/PRV pulse variability.

There are two modes of sensor configuration: transmission and reflection modes. In transmission mode, source and detector are placed opposite to each other, and in reflection mode, source and detector are on same side. The transmission mode is limited for finger tips and earlobes. Reflection mode detects light that is back-scattered, so it is used for bone and blood vessels and tissues [1–4].

Neural activity is measured non-invasively by NIRS system. During the functional activity of the brain, the blood flow is increased. NIRS detects the neural activity in a specific brain area by sensing the light absorption for the oxygenated and deoxygenated blood. Advantages of NIRS system are freedom of movement, better temporal resolution, easy to use, test young infant and cheaper cost of operation. NIRs and fMRI are closely aligned techniques as both techniques detect functional brain activity due to increase in blood flow. The neural activity in the brain is closely linked with the changes in blood flow and blood oxygenation. The consumption of energy is increased during the neural cell activity. The response is to increase blood flow to regions of increased neural activity, which occurs after 1–2 s delay. This leads to changes in local cerebral blood volume and the concentration of oxyhemoglobin that are detected using plethysmography effects. CPPG offers several advantages over other in vivo optical methods. CPPG is measured by using optical sensors. Optical sensors consume less power and require little maintenance [5]. Neuroimaging techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) provide functional brain mapping information. PET uses radioactive substances, and fMRI requires motion restriction. So it is difficult to fully examine brain of children, elderly and psychoneurological patients using fMRI and PET. Neural activity is measured non-invasively by NIRS system. Advantages of NIRS system are freedom of movement, better temporal resolution, easy to use, test young infant and the cost of operation is cheaper [5–7].

2 Materials and Methods

Transformations are used to detect hidden information in the raw data. Fourier transform (FT) represents the signal in frequency domain, and no time information is available. Fourier transform of the signal is used for stationary signals. Biomedical signals are usually non-stationary signals. The frequency content of biomedical signal changes over time. Therefore, for biomedical signal analysis, time and frequency information is needed [8, 9].

Spectral and statistical analysis (mean, variance, standard deviation, energy and entropy) was done on the captured CPPG signal for various activities.

2.1 System Block Diagram

CPPG sensor is designed with IR source (860 nm) and detector (OPT 101). AC sine wave in the range of 1–2 MHz frequency is used as source excitation. Light is penetrated through forehead using IR source. Signal is acquired during eye open and eye blinking activities with sampling frequency of 500 Hz. Using Daubechies 9, six-level wavelet decomposition of CPPG signal is performed. Figure 1 shows the system block diagram [10–12].

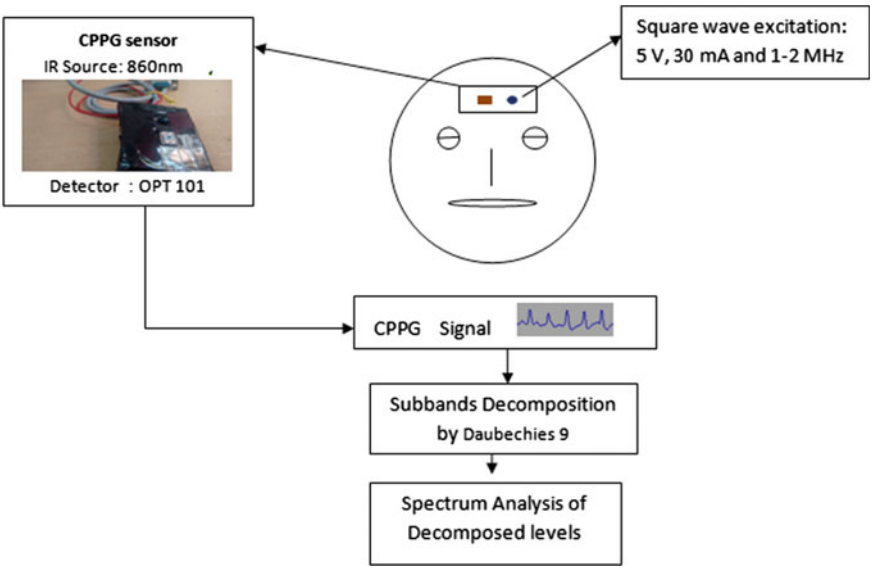


Fig. 1 System block diagram

2.2 Data Collection of CPPG

We recruited few volunteers to participate in the trail: the study was carried out in Cummins College of Engineering, Pune, India, as shown in Fig. 2. The following parameters were recorded before capturing the signal: subject name, age, height, body temperature, blood pressure, heart rate, sugar and SPO2. All the volunteers were healthy for the duration of the experiment as shown in Table 1.

The signal is captured for various activities like eyes open and eyes blinking as well as for various auditory emotions. The CPPG sensor was placed on prefrontal lobes and the signal was recorded and analysed before and after the activity.

The CCPG signal is acquired using AD620 instrumentation amplifier, and AVR ATmega 8535 microcontroller is used for digitizing the signal with sampling frequency of 500 samples per second and transmitting the data onto the computer as shown in Fig. 3. It is on this platform the signal is further processed and analysed.

Fig. 2 Subject under study

Table 1 Parameters recorded before CPPG acquisition

Specimen	Age	Gender	Height (cm)	Weight (Kg)	BP mm Hg	Sugar mg/dL	HR BPM	Body temperature (°F)	Spo2 (%)
1	47	Female	164	65.2	176/93	120	104	93.7	99
2	51	Female	149	59	140/80	101	73	92.2	98
3	53	Male	149	40.8	127/60	110	104	91.8	99
4	49	Male	170	76.5	127/80	167	106	93.2	99
5	31	Male	169	75.4	138/91	98	94	91.2	99
6	23	Female	163	50.3	136/79	100	81	92.2	98
7	46	Female	153	53.7	127/69	120	100	93	99
8	33	Female	156	51	108/71	107	89	92.2	99
9	39	Female	170	39	128/66	115	91	93.1	98
10	28	Male	165	67	146/101	96	93	93	99
11	43	Female	161	46.9	117/74	101	80	92.1	98

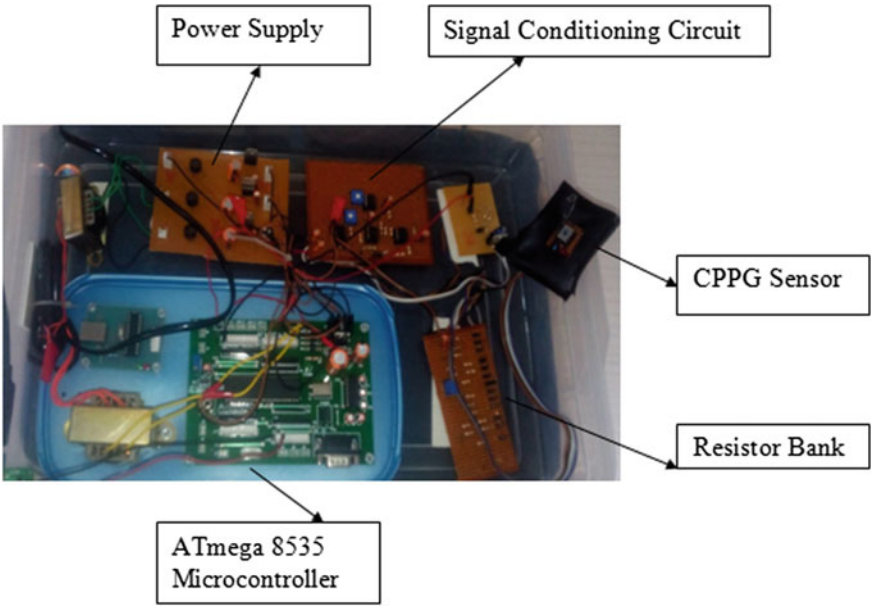


Fig. 3 Hardware set-up for CPPG acquisition

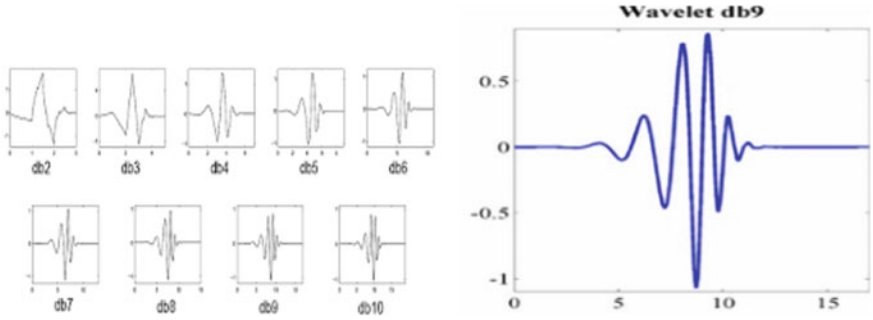


Fig. 4 Wavelet function of Daubechies 9

2.3 Daubechies Wavelet

The most commonly used set of discrete wavelet transforms was formulated by the Belgian mathematician Ingrid Daubechies in 1988. This formulation is based on the use of recurrence relations to generate progressively finer discrete samplings of an implicit mother wavelet function; each resolution is twice that of the previous scale. Db9 shows similarity with the shape of the normal PPG signal. Therefore, it is chosen for analysis of the CPPG signals. The wavelet function of Daubechies 9 is shown in Fig. 4 [13–15].

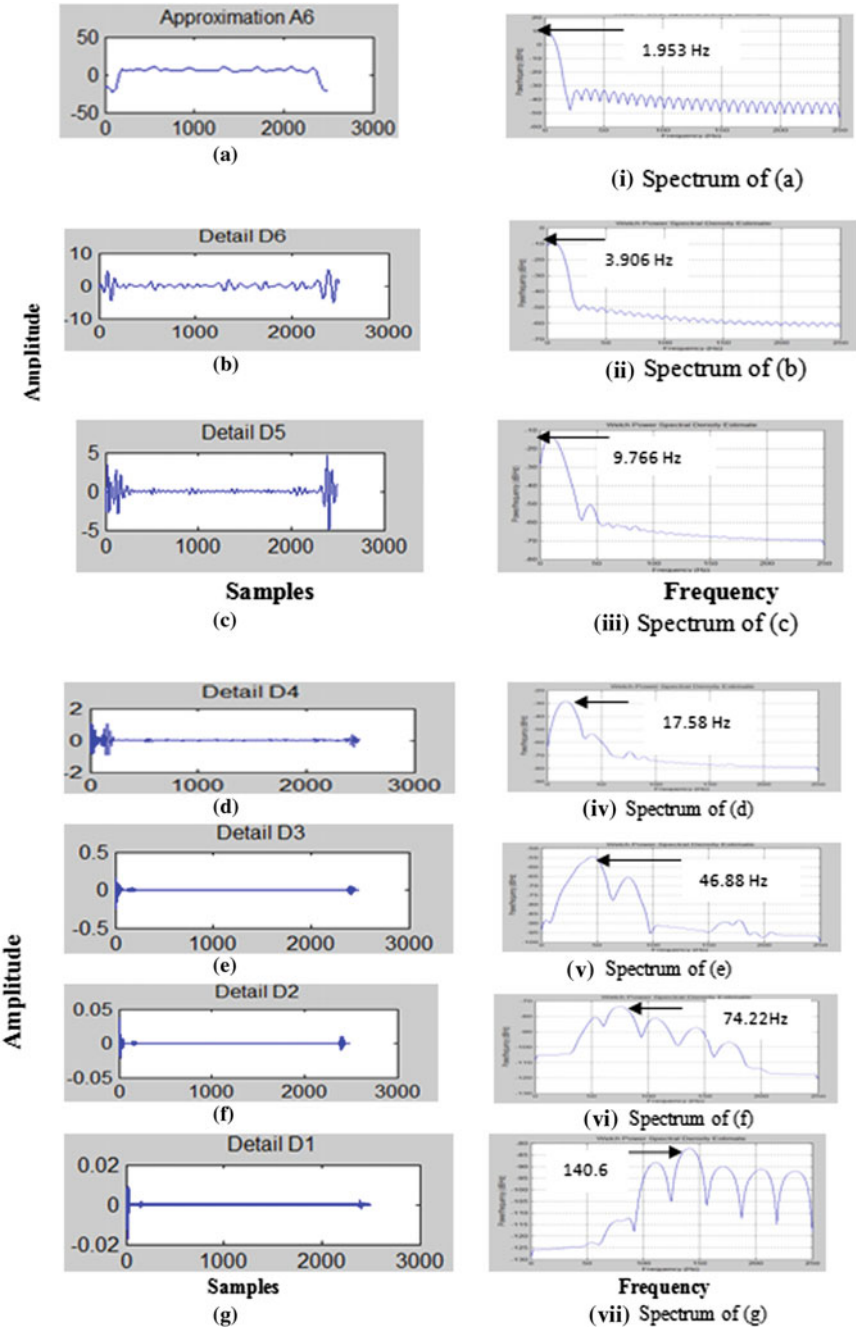


Fig. 5 **a** Reconstructed signal at scale A_6 with panel (i) showing its spectrum. **b** Reconstructed signal at scale D_6 with panel (ii) showing its spectrum. **c** Reconstructed signal at scale D_5 with panel (iii) showing its spectrum. **d** Reconstructed signal at scale D_4 with panel (iv) showing its spectrum. **e** Reconstructed signal at scale D_3 with panel (v) showing its spectrum. **f** Reconstructed signal at scale D_2 with panel (vi) showing its spectrum. **g** Reconstructed signal at scale D_1 with panel (vii) showing its spectrum

3 Result of Spectral Analysis

Using Daubechies 9 wavelet, a six-level wavelet decomposition of the signal is performed. Spectral analysis of all six levels is performed using Welch's method [16–18] shown in Fig. 5. The signals are reconstructed from sub-band levels. The reconstructed signals and its corresponding spectra at different scales are plotted in MATLAB [19].

The spectral analysis of reconstructed signals found that the centre frequency of A_6 is at 1.953 Hz. The higher frequency part of the signal is found in D_4-scale, and the centre frequency is 17.58 Hz, D_5-9.766 Hz and D_6-3.906 Hz. Centre frequency of D_1-140.6 Hz, D_2-74.22 and D_3-46.88 Hz consist of predominantly the noise part of the signal. The signal frequency is detected in D_6 and A_6 levels. Statistical feature of main signal and A_6 and D_6 are extracted.

CPPG signal is decomposed and features can discriminate between different activities. The features used include some wavelet-based features and some statistical features of CPPG signal [20]. Wavelet-based features: mean, variance, standard deviation, energy and entropy, are used as parameters after wavelet decomposition.

Interval plots of statistical parameters of signal CPPG, A_6, D_6 are plotted using Minitab 17. Interval plots show the graphical summary of the distribution of statistical parameters that show the central tendency and variability as shown in Fig. 6 and Table 2.

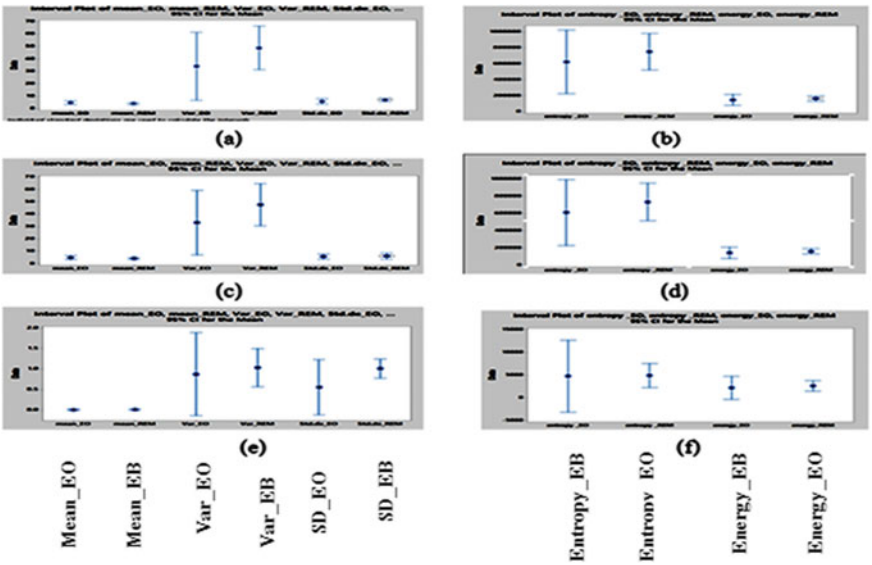


Fig. 6 **a** Mean, variance, standard deviation interval plot of CPPG for EO and EB, **b** entropy and energy plot of CPPG for EO and EB, **c** mean, variance, standard deviation interval plot of A_6 for EO and EB, **d** entropy and energy plot of A_6 for EO and EB, **e** mean, variance, standard deviation interval plot of D_6 for EO and EB, **f** entropy and energy plot of D_6 for EO and EB

Table 2 Central tendency of statistical parameters

	Mean		Variance		Standard deviation		Entropy		Energy	
CPPG signal	EO	EB	EO	EB	EO	EB	EO	EB	EO	EB
	4.8024	3.99968	34.1024	48.8540	5.71355	6.95395	623200	750437.5	144745	162338
A6 level	4.7979	3.99308	32.9877	47.5801	5.62155	5.90627	609455	731783	141853	159023
D6 level	0.00193741	0.00727529	0.87235	1.0318	0.553925	1.0077	4738.5	4878.5	2180	2577.5

4 Conclusion and Discussion

CPPG signal is captured to detect the functional activity in the prefrontal lobe. Functional activity in the prefrontal lobe is detected using analysis of cranial PPG. Data are collected from healthy persons. Six-level decomposition using Daubechies 9 is performed on the signals. Spectral and statistical analysis using Welch's method for different sub-bands is recorded. The spectral analysis shows the centre frequency of A_6 at 1.953 Hz. In D_4-scale, the centre frequency is 17.58 Hz, in which higher frequency part of the signal is found. D_3-D_1 scales consist of predominantly the noise part of the signal. The other details of sub-bands contain the noise part of the signal and hence not considered for calculation. These features of A_6 and D_6 can be used for classifying the activated and non-activated regions in prefrontal lobe. Statistical analysis shows that for A_6 level mean, variance, standard Deviation, entropy and energy values with eyes open are 4.7979, 32.9877, 5.62, 609455 and 141853, respectively, whereas mean, variance, standard deviation, entropy and energy values for eyes blinking are 3.99308, 47.5801, 5.90, 731783 and 159023, respectively. Statistical parameters differ for eye open and eye blinking activities. These features can be used for classifying the activated and non-activated regions in prefrontal lobe. Authors have taken the consent from the concerned person to use the picture and database in the paper. Authors will be solely responsible if any issues arises in future with regard to this.

References

1. Sun, Y., Thakor, N.: Photoplethysmography revisited: from contact to noncontact, from point to imaging. IEEE Trans. Biomed. Eng. **63**(3) (2016)
2. Feng, L., et al.: Motion –resistance remote imaging phoplethysmography based on the optical properties of skin. IEEE Trans. CircuitsSyst. Video Technol. **l25**(5), 879–891 (2015)
3. Sun, Y., et al.: Noncontact imaging photoplethysmography to effectively access pulse rate variability. J. Biomed. Opt. **18**(6), 061205-1−061205-9 (2013)
4. Stickerer, R., et al.: Non-contact video based pulse rate measurement on a mobile service robot. In: Proceedings of IEEE International Symposium on Robot Human Interactive Communication (Conference), pp 1056–1062 (2014)
5. Sheth, S.A., Yanamadala, V., Eskandar, E.N.: Intraoperative human functional brain mapping using optical intrinsic signal imaging. Advances in Brain Imaging. ISBN 978-953-307-955-4 (2012)
6. Korhonen, V.O., Myllyl¨a, T.S., Kirillin, M.Y., Popov, A.P., Bykov, A.V., Gorshkov, A.V., Sergeeva, E.A., Kinnunen, M., Kiviniemi, V.: Light propagation in NIR spectroscopy of the human brain. IEEE J. Sel. Top. Quantum Electron. **20**(2) (2014)
7. Hespos, S.J.: What is optical imaging? J. Cogn. Dev. **11**(1), 1–13, Copyright# 2010 Taylor & Francis Group, LLC
8. Hou, X., Ding, H., Teng, Y.: NIRS study of cerebral oxygenation and hemoglobin in neonate at birth
9. Ahammad, N., Fathima, T., Joseph, P.: Research Article Detection of Epileptic Seizure Event and Onset Using EEG. Hindawi Publishing Corporation BioMed Research International, Vol. 2014 (2014)

10. Cromwell, L., Weibell, F.J., Pfeiffer, E.A.: Biomedical and Measurements, 2nd edn. Eastern Economy Edition, New Delhi (2009)
11. OPT 101 Data sheet
12. Carr, J.J., Brown, J.M.: –Introduction to Biomedical Equipment Technology
13. Cvetkovic, D., Ubeyli, E.D., Cosic, I.:Wavelet transform feature extraction from PPG, ECG and EEG signal responses to ELF PEMF exposures. Digit. Signal Process. **18** (2008)
14. Mohd Tumari, S.Z., Sudirman, R., Ahmad, A.H.: Selection of a suitable wavelet for cognitive memory using electroencephalograph signal. Engineering **5**, 15–19 (2013). https://doi.org/10.4236/eng.2013.55b004
15. Jayachandran, E.S., et al.: Analysis of myocardial infarction using discrete wavelet transform. J. Med. Syst. **34**(6), 985–992 (2010)
16. Sun, L., et al.: ECG analysis using multiple instance learning for myocardial infarction detection. IEEE Trans. Biomed. Eng. **59**(12), 3348–3356 (2012)
17. Sharma, L.N., et al.: ECG signal denoising using higher order statistics in wavelet subbands. Biomed. Signal Process. Control **5**(3), 214–222 (2010)
18. Sharma, L.N., Tripathy, R.K., Dandapat, S.: Multiscale energy and eigenspace approach to detection and localization of myocardial infarction. IEEE Trans. Biomed. Eng. **62**(7) (2015)
19. Rangayyan, R.M.: Biomedical Signal Analysis: A Case –Study Approach. IEEE Press Series in Biomedical Engineering. IEEE Press, Piscataway
20. Ting, W., Guo-Zheng, Y., Bang-Hua, Y., Hong, S.: EEG feature extraction based on wavelet packet decomposition for brain computer interface. Measurement **41**, 618–625 (2008)

Multi-mode Routing Mechanism with Cryptographic Techniques and Reduction of Packet Drop Using 2ACK Scheme MANETs

V. Lakshman Narayana and C. R. Bharathi

Abstract A mobile ad hoc network (MANET) is an accumulation of portable hubs that are equipped for imparting information to one another by means of a remote connection. A MANET can be positioned wherever necessary without any fixed infrastructure. There are certain parameters which are important to the execution of a MANET, i.e., routing, throughput, packet drop, delay, and so forth. For secure data transmission cryptographic mechanisms are used and routing should be done dynamically so that nodes can easily find the shortest path for data delivery from source to destination. A private key generator (PKG) is used for generating keys for performing encryption and decryption of data. In this chapter, we propose a cryptographic strategy for generating keys and for dynamic routing a multi-mode routing algorithm is proposed. An ad hoc on-demand distance vector (AODV) protocol is used which identifies the shortest path and makes nodes available for data transfer. Another strategy is proposed which mitigates packet drop by presenting a 2ACK plan with a sorting calculation.

1 Introduction

Directing is choosing ways to send information bundles. A uniquely named directing tradition is a custom, or customary, that controls how focus guides pick which route toward manage course assigns selecting contraptions in a versatile

V. L. Narayana (✉)
Department of CSE, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India
e-mail: lakshmanv58@gmail.com

V. L. Narayana
Vignan's Nirula Institute of Technology and Science for Women, Guntur, Andhra Pradesh, India

C. R. Bharathi
Department of ECE, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India
e-mail: crbharathi@veltechuniv.edu.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*, Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_63

without any preparation structure. In remarkably named systems, focus don't begin acquainted with the topology of their structures; rather, they need to find it. Stale sections are territories in a network which have not been used for a long time. Such areas and furthermore the routes operating within them, are removed [1]. By then new target comes. This is the way by which it limits.

All instances where a packet has dropped, because of portability, blockage, transmission blunder, or assault, are collectively known as packet drop. Henceforth, packet drop = sent packet–received packet [2]. Other than those issues mentioned above, a mobile ad hoc network (MANET) experiences other significant issues like dark opening assault, malignant assault, and worm gap assault which are usually the main causes of packet drop. A few hubs deliberately drop packets, these hubs are called malevolent hubs. Indeed, even after the ideal determination of the course of data transfer, packet drop can reduce performance [3]. Hubs are in charge of packet drop in two ways [4]. Hubs can dispose of packets because of inadequate assets (clog) or can dispose of a packet for no apparent reason (a malevolent hub).

2 Related Work

There has been a considerable volume of work completed on tending to packet drop in a MANET. Yi Lu and Yuhui Zhong proposed a strategy to comprehend the primary issues which are specifically identified with packet drop. They demonstrated that an ad hoc on-demand distance vector (AODV) has more packet drop because of versatility in contrast with blockage; thus, an AODV is more delicate in terms of portability. Shiv Shakti et al., proposed a technique in which they utilized blended elements of static in addition to dynamic directing calculation. The proposed framework is sufficiently adept at discovering following hubs for conveyance of the packet. An ahead-of-time AODV has smaller packet drop compared with an AODV. Hemant Kumar et al., proposed a technique in which they used a strategy for identifying broken connections between two hubs, after which the connections were either repaired or disposed of in order to prevent packet drop. The proposed strategy keeps packet drop because of the movement thickness and distances itself from pernicious connections.

3 The Working of an Ad hoc on-Demand Distance Vector

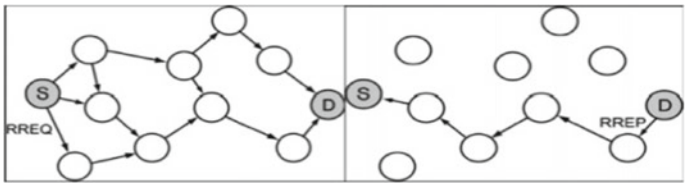
In an AODV, the system is tranquil until the point that a connection is required. At that point the system focus point that needs an association becomes connected. Other AODVs pass forward the message, and record the point they received it from, representing short-lived courses back to the focus point. When a point inside the network receives such a message, and has a course to its coveted focus point [5], it bestows something specific in a switch through a vaporous course to the required focus point.

Each adaptable host in the framework goes about its task acting as a particular switch and receiving its required course—in this way the framework becomes self-starting [6]. Each center point in the framework maintains a guiding table [7], with the guiding information sections being passed to its neighboring centers, as well as two separate counters: a center course of action number and an impart id. When a center (say, source center point "S") needs to talk with another (say, objective center "D"), it builds its convey id and begins path divulgence by imparting a "course ask distribute" to its neighbors. The route request (RREQ) contains fields including: originator-addr, originator-sequence, destination-addr, destination-sequence, and bounce count.

3.1 Path-Finding Process

When attempting to set up an association with a target focus point without knowing a dynamic route to it, the sending focus will start a disclosure process [8]. A "course ask for" message (RREQ) is passed on to all neighbors, which keep broadcasting the message [9] to their neighbors. To guarantee circle-free and latest course data, each inside point keeps two flags: design number and broadcast_id [10]. The broadcast_id and the address of the source focus astoundingly see a RREQ message.

Broadcast_id is extended for each RREQ the source focus point starts. A halfway focus point can receive different duplicates of a practically identical course ask for passing on from different neighbors [7, 11]. For this condition—if a middle point has beginning at now got a RREQ with a relative source address and broadcast_id [5]—it will dispose of the package without broadcasting it. Figure 1 outlines the path-finding process using an AODV protocol and receiving replies from path available nodes. In order to discover the path a RREQ message is forwarded to all nodes from the sender with the destination node sending back a route reply (RREP) message so that the path is discovered. The RREQ partitions two movement numbers: the source blueprint number and the last target gathering number known to the source. As the RREP is sent back to the source, each transitional focus point, in this



(a) Process of path finding started from source

(b) A RREP packet is sent back to the source.

Fig. 1 Path-finding method of AODV protocol

manner, consolidates a forward course segment to its coordinating table. In case the course is never again used, it will be deleted after a predefined length of time. Since the RREP package [12] is continually sent back via the pivot route developed by the coordinated request, an AODV just supports symmetric associations.

3.2 Proposed Algorithm for Routing and Using a Private Key Generator

After successfully establishing a MANET the main problem is linked to path finding and using a private key generator (PKG). This is because the data at the source side is encrypted with a public key and for decryption a PKG issues a private key to the authorized destination. The proposed algorithm focuses on where a PKG is used and where its use is relaxed. Data from source to destination should be transferred securely along the shortest path within a MANET.

Algorithm Multi-Mode Routing {
if(No. of Nodes = = 2) {
Perform End to End Encryption of data among them.
}
else if(No. of Nodes = = 3) {
Consider the node other than source and destination as PKG and perform the transmission process.
}
else if(No. of Nodes > 3) {
Calculate the shortest path among the manet and find master and slave nodes (at least 3 nodes) by selecting nodes which satisfy the properties of MASTER & SLAVE nodes and initiate the data transfer.
}
else {
End;
} }

The above algorithm explains the process of the data transfer method between source and destination nodes using a PKG if the MANET group has a sufficient number of nodes. Otherwise the MANET initiates an end-to-end cryptography method for secure data transfer. After establishing a path from source to destination, a node within that path is selected as a master node which acts as a PKG. This node has two slaves under it in case the PKG moves out of the network—the slave would then become the master and would as the PKG. This process continues until the communication is completed.

3.3 End-to-end Cryptography

End-to-end encryption (E2EE) is an arrangement of correspondence where just the imparting clients can read the messages. On a fundamental level, it forestalls potential spies—including telecom suppliers, Internet suppliers, and even the supplier of the person who benefits from the correspondence—from having the capacity to get to the cryptographic keys which could unscramble the conversation [11]. The frameworks are intended to vanquish any endeavors at reconnaissance or alteration of information, in light of the fact that no outsiders can decode the information being imparted or put away. In E2EE, the data is encoded on the sender's system or contraption and only the recipient can unravel it. Nobody amidst, be they an Internet pro-association, application master center, or software engineer, can read it or disturb it. The cryptographic keys used to scramble and unravel the messages are secured exclusively on the endpoints, a trap influenced possible utilizing to open key encryption.

3.4 Layers Causing Packet Drop

Essentially, a media access control (MAC) layer and system layer are in charge of all sorts of packet drops [13]. A Macintosh layer is in charge of versatility-related packet drop and in addition blockage-related packet drop, whereas a system layer is mindful just of portability-related packet drop in directing convention.

Organize layer. When a packet reaches the system layer, directing convention is depended upon to forward the packet, if a course to the goal is accessible. If a course is not accessible then the packet is cradled until a course is accessible.

Two instances of packet drop at the system layer occur if:

1. The cushion is flooded when an approaching packet should be cradled.
2. The put away packet in the support surpasses the termination time (for an AODV the NS2 code permits 30 s living time).

This sort of packet drop generally happens in exceptionally powerful topologies of organization. The carrier-sense multiple access/collision avoidance (CSMA/CA) convention is in charge of blockage-related packet drop in a MAC layer.

The reasons for this are:

1. The backoff time is high and it surpasses the point of confinement.
2. The buffer is full (space is not accessible for a new approaching packet).

Since a MANET is an exceptionally powerful tool because of its remarkable conduct and versatile attributes, it tends to suffer connection breakage. If there is a connection breakage [13] the existing courses end up inert and hubs have to find a

Table 1 Packet drop at a MAC layer and network layer

	MAC layer	Network layer
Mobility related	Yes	Yes
Congestion	No	Yes

new course via a "course ask for packet." In such an instance RREQ messages are sent when a sender hub needs to impart information and is finding another course in on request steering convention, for example, AODV evidential, portability may build the quantity of course demand packet on the system. Table 1 illustrates the packet drop layers and reasons for such losses.

$$\text{Total packet drop} = (\text{MAC layer drop}) + (\text{network layer drop})$$

3.5 Customized 2ACK—Based Method

Various ACK-based calculations are utilized to discover a vindictive hub in the system. In the current 2ACK-based plan every hub sends ACK to the next hop and the received hop sends ACK back to the sender. This affirmation-based plan delivers a greater number of ACK packets in the system. At the end of the day it expands overheads in the system and the mass measure of 2ACK is capable of debasing system execution. Presently, we changed the plan by presenting a sorting calculation at the sender's side. Each hub sends ACK to the next hop and that hop sends ACK to its sender, in the wake of sending packets to its neighbor. It may be conceivable that at least two successive hubs may have the same number of ACK. The sorting process at the sender's side reduces the overhead and removes the issue because of its substantial overhead.

Here is the rejected algorithm for packet loss reduction:

For every hub decide the drop packets If packet dropped > edge include the hubs into pernicious hub list if (every hub is suspected)
{
hub X checks ACK and sort them/X is the vindictive recorded hub if (any ACK is lost from neighbor hubs)
{
check the ACK from goal comp(ACKneighbor, ACKdestination)
{
on the off chance that((ACK neighbor – ACKdestination) >=7) Add hub into malignant hub list
else
break;
} } }

Table 2 Parameters used for simulation

Number of nodes	100
Routing protocol	AODV
Area	1200*1200
Packet size	512 bytes
Mobility speed	0–50 ms
Simulation time	300
Max no. of connections	60
Tool	NS2

4 Results

The suggested method developed a MANET using an NS2 simulator and the results proved that the suggested method is far better than the existing method for secure routing and data transfer across the shortest path of the MANET.

The proposed method uses an AODV protocol and Table 2 shows the parameters used in the NS2 simulator.

Figures 2 and 3 outline the process of creating a MANET and the performance of the multi-mode routing algorithm in which the PKG is used only when the number of nodes in the MANET is greater than three and end-to-end cryptography is if the nodes in the MANET are limited to two.

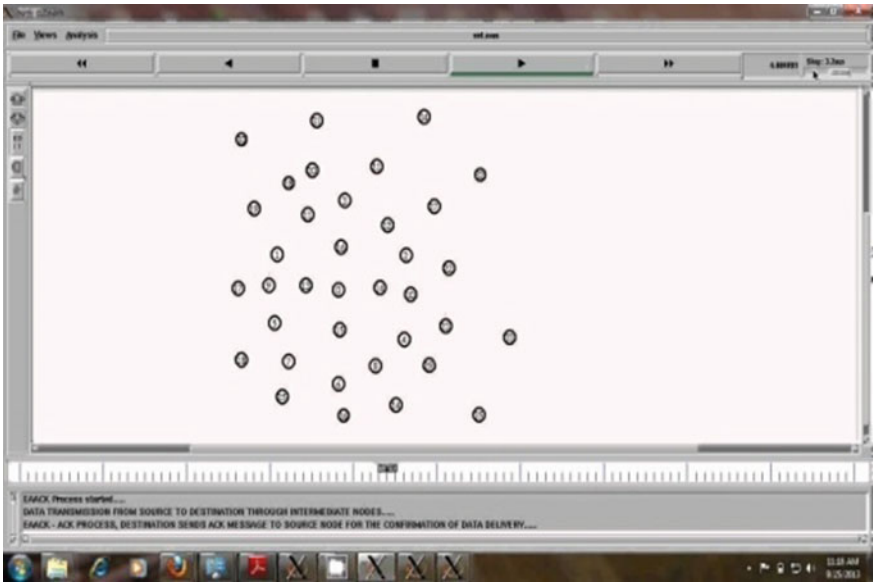


Fig. 2 Creating a MANET

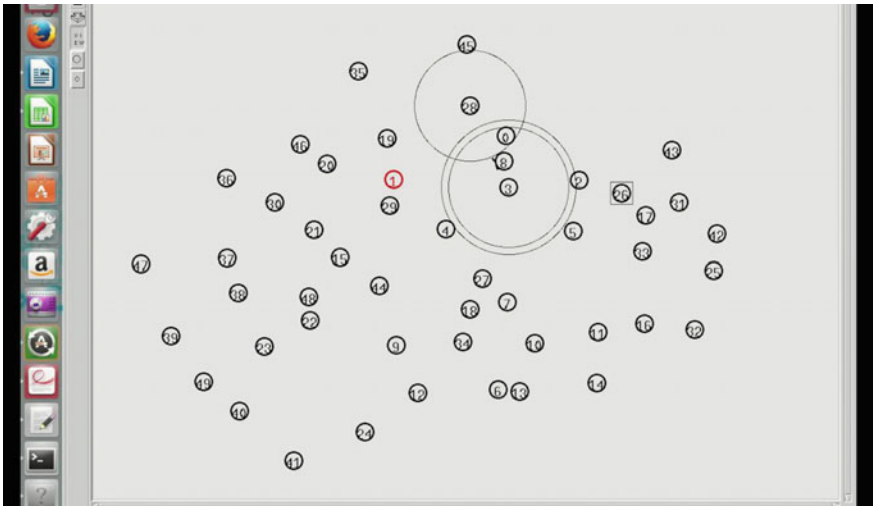


Fig. 3 Multi-mode routing mechanism

A comparison between the proposed mechanism and the existing method is illustrated in Fig. 4, which proves that the existing method reduces the packet drop ratio.

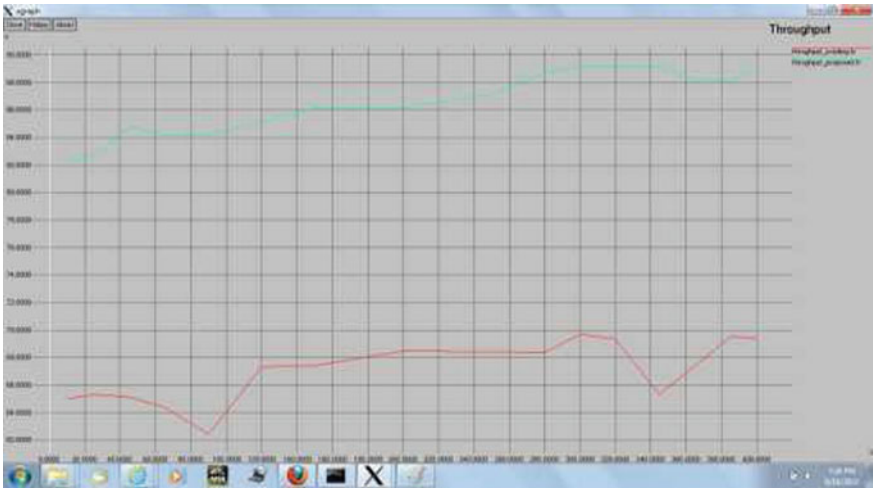


Fig. 4 Comparison of throughput between existing and proposed methods

5 Conclusion

In this chapter, the bad conduct of hubs is examined and another approach is proposed for finding and confining hubs which are acting up. Here a new multi-mode algorithm is proposed which is used for secure routing and finding the shortest path among the MANETs. A PKG is used for the encryption and decryption process and the data is securely transferred among MANETs. The recommended approach can be combined with any source-directing convention, for example, ALOHA and is in light of sending affirmation packets for gathering of information bundles and utilizing wanton mode for checking the quantity of information packet to such an extent that it conquers the issue of 2ACK plot.

Finally, we note that throughput is specifically affected by packet drop in a MANET. The 2ACK-based plan limits the overhead of the system and henceforth has a constructive outcome on the execution of the network. To demonstrate the viability and after effects of this proposed approach, usage deal with NS2 simulator. Future work will incorporate some confirmation instrument to ensure that the ACK packets are certifiable and also incorporate an instrument to rebuff uncontrollable hubs.

References

1. Zhu, Q.: A mobile ad hoc networks algorithm improved AODV protocol. In: Proceedings of International Conference on Power Electronics and Engineering Application (2011)
2. Garg, H.K., et al.: Minimization of average delay, routing load and packet drop rate in AODV routing protocol. Int. J. Comput. Appl. **44**(15) (2012)
3. Manish, V.M., et al.: Diminution of packet drop by efficient selection of network route in MANET. Int. J. Comput. Sci. Inf. Technol. **5**(2) (2014)
4. Lu, Y., et al.: Packet drop in mobile adhoc networks. Computer Science Technical Reports Paper, Purdue University, p. 1558 (2003)
5. ShaktiSrivastva, S., et al.: Minimization of the packet drops in MANETs based on both static and dynamic routing protocols. Int. J. Comput. Appl. (IJCA) (2011)
6. Bhanumati, V., et al.: RSS based energy efficient scheme for the reduction of overhearing and rebroadcast for MANET. Elsevier, Procedia Eng. **38**, 2463–2472 (2012)
7. Hassan, S.T.: Designing a new MANET environment using computer simulation. Int. J. Comput. Sci. Electron. Eng. (IJCSEE) **1**(3) (2013). ISSN2320-401x; EISSN 2320-4028
8. Gaikwad, S., Adane, D.S.: Reduction in routing overhead using 2-Ack scheme and novel routing algorithm. IJETT **4**(8)
9. Kaur, A., et al.: Efficiency enhancement in AOMDV to reduce the chances of packet drop in MANETs. Proc. Int. J. Sci. Res. (IJSR), 2319–7064 (2013). ISSN (on line)
10. Sharma, S., et al.: Reducing packet drop in MANET. Proc. Netw. Complex Syst. (2013). ISSN 2224-610X (paper); ISSN 2225-0603 (online)
11. Tyagi, S., et al.: A reliablity based variant of AODV in MANETs: proposal, analysis and comparison. In: Proceedings of the 7th International Conference on Communication, Computing and Virtualization (2016)

12. Ghosekar, P., et al.: Mobile ad hoc networking: imperatives and challenges. In: Proceedings of IJCA Special Issues on Mobile Ad-hoc Networks (MANET) (2010)
13. Lanjewar, A., et al.: Optimizing cost, delay, packet drop and network load in AODV routing protocol. Proc. Int. J. Comput. Sci. Inf. Secur. **11**(4) (2013)

Automatic Early Stage Glaucoma Detection Using Cascade Correlation Neural Network

Thresiamma Devasia, K. Poulose Jacob and Tessamma Thomas

Abstract The automatic detection of glaucoma with the fundus eye image analysis using novel techniques is presented in this paper. The classification of glaucoma is done using cascade correlation neural network. The advantage of this network is that it adds the hidden units one by one instead of using a predefined set of hidden units until the error is minimized. A robust pattern-based feature encoding method, local mesh pattern is used to generate feature maps for segmenting the optic disc and optic cup using an efficient segmentation technique modified fuzzy C-means clustering algorithm. The cascade correlation neural network classifier of the proposed method determines whether the input image is normal or glaucomatous.

1 Introduction

According to the World Health Organization, glaucoma is the second leading cause of blindness [1], and it is a disease due to the intraocular pressure (IOP) within the eye. The optic disc (OD) is the region of retina where the blood vessels pass through, and a depression within the OD is the optic cup (OC). As the IOP increases, the size of the OC increases. The evaluation of cup-to-disc ratio (CDR) is used in this paper for the detection of glaucoma [2]. Since thinning of neuroretinal

T. Devasia (✉)
Department of Computer Sceince, Assumption College,
Changanacherry, Kerala, India
e-mail: mailtocherukusumam@gmail.com

K. P. Jacob
Department of Computer Science, Cochin University of Science
and Technology, Kochi, Kerala, India
e-mail: kpj0101@gmail.com

T. Thomas
Department of Electronics, Cochin University of Science and Technology,
Kochi, Kerala, India
e-mail: tessamma1@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_64

rim area is a symptom of a suspect of glaucoma, ISNT rule analysis and rim-to-disc ratio (RDR) are also used for glaucoma detection in the proposed method.

Different techniques have been explored towards the detection of glaucoma in the literature. Marin et al. [3] detected OD using morphological operations. OD segmentation is regarded as an energy minimization problem, and a variational model was used by Baisheng Dai et al. [4]. Fulong et al. [5] localized OD using a line operator filter, and contour is detected using iterative thresholding and ellipse fitting. Julian Zilly et al. [6] detected OD and OC using the entropy sampling technique and ensemble learning. Mookiah et al. [7] developed a system based on higher-order spectral features, trace transforms and discrete wavelet transform features for glaucoma detection. Chen et al. [8] applied deep learning with deep convolutional neural network for feature learning for glaucoma detection. Noronha et al. [9] classified glaucoma using higher-order cumulant features. Anish Isacc et al. [10] detected and classified glaucoma using an adaptive threshold technique. Maheshwari et al. [11] used empirical wavelet transform for glaucoma detection. The proposed system is a novel approach for detecting glaucoma at the initial stage.

2 Materials and Methods

2.1 The Materials

The proposed method was tested on a real image data set which includes normal and abnormal retinal fundus images of right and left eye collected from screening centres. All the real images were obtained using Carl Zeiss fundus camera with the pixel size of each image is $576 \times 768 \times 3$ [12].

2.2 Methods

The implementation of the proposed glaucoma detection system is done using four modules such as pre-processing, feature map generation, feature extraction and finally classification. The graphical representation of the proposed system is given in Fig. 1. Automatic OD localization and ROI extraction are done using the circular Hough transform of the canny edge image of the intensity component of the *HSI* model of the input image. The red and green channels are selected from the ROI and removed blood vessels, and it generates the LMeP feature maps for further processing. The pre-processed ROI is only used for the segmentation of OD and OC by the existing methods. The proposed work uses textural and structural information to segment the optic disc and optic cup. The generated textural LMeP feature maps, which contain more edge information is used to segment OD and OC, using a powerful segmentation technique modified fuzzy C-means clustering algorithm, and it provides rich description of OD and OC. Using the segmented OD and OC,

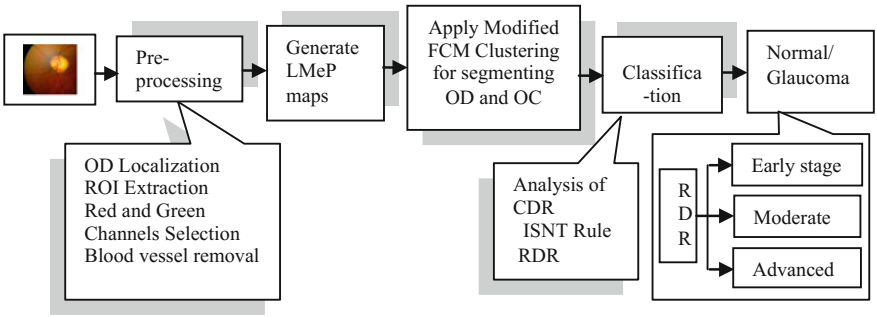


Fig. 1 Proposed glaucoma detection system

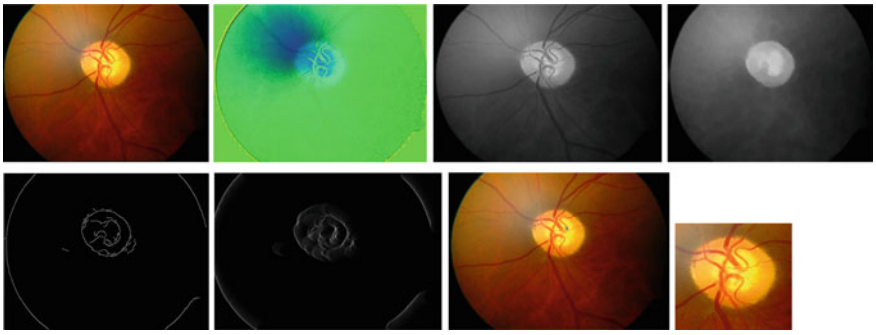


Fig. 2 From left to right and top to bottom: fundus image; HSI model; I component; vessel free I component; canny edge detection; circular Hough transform; region with highest blood vessel density and circular shape is marked as OD (blue x approximately at the centre of OD); region of interest (ROI) extracted from the localized OD

the glaucoma assessment parameters CDR, RDR and ISNT values are computed. The CCNN classifier [13] train and test the subjects using the extracted features CDR, RDR and ISNT rule values and differentiates the normal and glaucoma subjects using the ground truth obtained from the ophthalmologist. The stage of glaucoma of the glaucomatous images obtained from classification is identified using RDR as early/moderate/advanced.

(1) **Pre-processing**

The automatic localization of the OD is done using the circular Hough transform of the canny edge image of the intensity component of the *HSI* model [14] of the input image. From the peak of the intersection of the Hough circles, the OD is located and the region of interest (ROI) of size 250 × 300 is cropped. Figure 2 shows the OD localization and ROI extraction process.

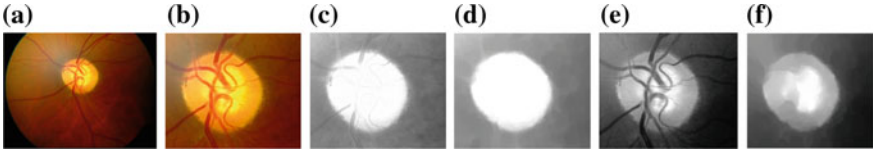


Fig. 3 Pre-processing steps. **a** input image. **b** ROI. **c** red channel. **d** vessel free red channel. **e** green channel. **f** vessel free green channel

The red channel [15] and green channel of the ROI are selected for OD and OC segmentation and removed blood vessels using morphological closing operation [16]. The pre-processing steps are given in Fig. 3.

(2) **Feature Maps Generation using Local Mesh Patterns (LMeP)**

To capture useful edge information for improving the segmentation accuracy, the local pattern LMeP inspired by [17] is used in the proposed method. LMeP is a local pattern in which an image is composed of micropatterns. For a given (T, R), the LMeP is computed using (1) [17]

$$\begin{gathered} LMeP^{i}_{T,R} = \sum_{j=1}^{T} 2^{(j-1)} \times F_{n2}\left(s_{\propto}|_{R} - s_{i}|_{R}\right). \\ \propto = 1 + mod((j+T+i-1), T). \\ \forall_i = 1, 2, \ldots.., \left(\frac{T}{2}\right). \end{gathered} \tag{1}$$

where i represents the LMeP index and mod(u, v) returns the reminder for u/v operation.

$$F_{n2}(x) = \begin{cases} 1 & if\ x > 0 \\ 0 & othrwise \end{cases}$$

(3) **Segmentation of Optic Disc and Optic Cup**

The optic disc is segmented using the LMeP feature map of the red channel of the ROI using an efficient segmentation technique modified FCM clustering algorithm [18]. The distance measurement dt_{ki} is modified as a combination of local and non-local information as shown in (2) [18].

$$dt_{ki}\left(a_j, b_i\right) = \left(1 - \lambda_j\right) dt_l^2\left(a_j, b_i\right) + \lambda_j dt_{nl}^2\left(a_j, b_i\right). \tag{2}$$

where dt_l and dt_{nl} represent the distance computation determined by local and non-local information. The distance computation dt_l is given in (3) [18].

$$dt_l^2\left(a_j, b_i\right) = \frac{\sum a_k \in N_j w_l\left(a_k, a_j\right) dt^2\left(x_k, b_j\right)}{\sum a_k \in N_i w_l\left(a_k, a_j\right)}. \tag{3}$$

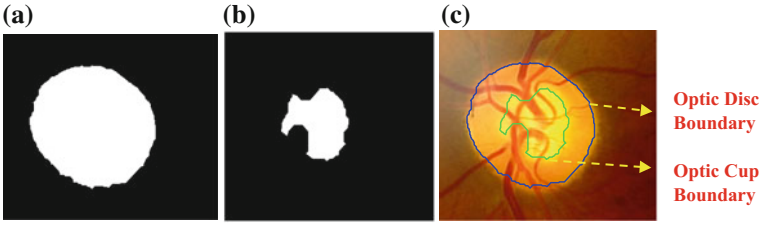


Fig. 4 **a** Extracted optic disc **b** extracted cup **c** boundary of the optic disc and cup overlaid on the ROI

where $dt_l^2(a_j, b_i)$ is the Euclidean distance computation and $w_l(a_k, a_j)$ is the weight of each pixel in N_i.

The distance computation dt_{nl} is given in (4) [18].

$$dt_{nl}^2(a_j, b_i) = \sum_{a_k \in I} w_{nl}(a_k, a_j) dt^2(a_k, b_i). \tag{4}$$

The membership value is evaluated by the distance measurement. Three clusters are generated from the feature map of red channel for OD segmentation [12, 19]. The cluster that holds the maximum membership grade M is found, and the corresponding grades are assigned with the same identification label [12, 19]. The process assigns the object $obj1$ to the class $Clust1$ grey levels of the obtained cluster with partition matrix M, as the highest membership. From the obtained cluster, the average of the maximum and minimum grey levels are found and used to set the level for thresholding [12, 19]. Now the image is converted to binary image using (5).

$$Clust1_{obj1} = arg_k \{\max(M_{k,obj1})\}, k = 1, 2, 3. \tag{5}$$

The largest connected component whose shape is approximately circular is selected using the compactness measure [12, 19]. This removes the unwanted objects from the binary image and is considered as the optic disc approximation [12, 19]. The same algorithm is applied to the LMeP feature map of green channel for the segmentation optic cup. Figure 4 shows the OD and OC segmentation.

(4) **Assessment of Glaucoma using Parameters**

From the segmented optic disc and optic cup, CDR area ratio, the ratio of the cup area to disc area is computed. Area ratio is chosen to achieve the comprehensive segmentation accuracy in all directions instead of the diameter ratio which gives the accuracy only in one direction. If CDR is less than 0.3, it appears to be a normal image; otherwise the image is glaucoma affected [20]. The OD damage affects the NRR and thinning of NRR, thus become a symptom of glaucoma [21]. The normal optic disc usually has a configuration which follows the ISNT rule (I > S > N > T), and violation of ISNT rule is the symptom of glaucoma. The calculation of RDR at superior and inferior poles is essential for the accurate diagnosis of glaucoma.

Fig. 5 **a** ISNT image. **b** optic rim and mask used for detecting the rim and the rim areas inside the mask: at **c** inferior side. **d** superior side. **e** nasal side. **f** temporal side

The value of RDR less than 0.2 in the superior and/or in the inferior region for glaucoma subjects, otherwise it refers to a normal subject. To filter one quadrant, a mask of the cropped-image size is used [22]. The ISNT image, extracted optic rim and the mask used for detecting the rim quadrants and the rim areas inside the quadrants are shown in Fig. 5.

3 Experimental Results

The algorithm has been designed in a framework of MATLAB R2012a, which aims to detect glaucoma at an early stage. The proposed approach is tested on a real image data set composed of 158 images composed of normal and glaucomatous of both left and right eyes which are collected from screening centres. The OD and OC are segmented after pre-processing, and from the segmented OD and OC, glaucoma detection parameters CDR, RDR and ISNT rule values are calculated for further processing. The CCNN classifier differentiates the input image as normal or glaucomatous using the ground truth obtained from the ophthalmologist. The performance of the proposed work is assessed using receiver operating characteristics (ROC) curve [23].

3.1 Classification Using Cascade-Correlation Neural Network

Artificial neural networks (ANNs) are robust tools for pattern recognition since they learn complex, nonlinear surfaces amongst different classes. Cascade-correlation is a combination of two ideas: the cascade architecture and the learning algorithm. In cascade architecture, the addition of hidden units is done only one at a time and do not change after addition. The new hidden units are created and installed in the learning algorithm. The proposed method uses the glaucoma assessment parameters CDR, RDR and ISNT rule values, and these are the input to the CCNN and the CCNN classifier [13] which in turn determines the normal and glaucoma subjects using the ground truth obtained from the ophthalmologist. Figure 6 shows the training of CCNN for glaucoma detection.

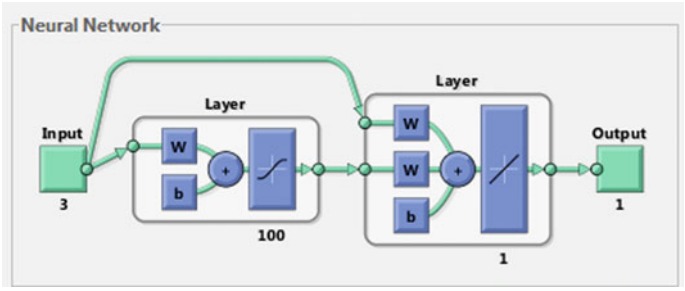


Fig. 6 CCNN training for glaucoma detection

Table 1 Parameter calculation for the images

Image	Clinical CDR	Obtained CDR	ISNT rule analysis				RDR (I)	RDR (S)
			Inferior (I)	Superior (S)	Nasal (N)	Temporal (T)		
1	0.61	0.60	3636	3669	3873	3098	0.09	0.09
2	0.48	0.49	4378	3692	4019	3318	0.07	0.06
3	0.24	0.26	2159	2011	1837	1569	0.31	0.30
4	0.65	0.64	1227	1468	1329	1589	0.05	0.05
5	0.43	0.46	2805	2878	2698	2502	0.18	0.18
6	0.20	0.23	1512	1422	1153	1049	0.33	0.32
7	0.54	0.51	2066	2310	1845	2688	0.06	0.07
8	0.71	0.64	4197	5118	2379	1058	0.03	0.04
9	0.89	0.91	3130	3433	1923	2124	0.01	0.02
10	0.81	0.77	4722	5928	4832	5233	0.02	0.03

3.2 Result of Glaucoma Detection Parameter Analysis

The comparison result between obtained CDR and the clinical CDR received from the ophthalmologist shows the accuracy of the proposed method, and it is given in (6).

$$cdr_{err} = |cdr_{cl} - cdr_{obt}|. \tag{6}$$

where $cdr_{err}, cdr_{cl}, cdr_{obt}$ are the CDR error, clinical CDR and obtained CDR.

For all the images, the mean error in estimating cup-to-disc area ratio is 0.16, and it shows the inconsistency of CDR measure alone to assess glaucoma. In classification, the glaucomatous images are identified using the images with CDR value greater than 0.3, RDR value less than 0.2 at inferior and/or superior, and the same image violates the ISNT rule, otherwise the image is diagnosed as normal. All the parameters of ten sample subjects from tested images are shown in Table 1.

Table 2 Performance analysis with CDR, RDR and ISNT features

Method	Sensitivity	Specificity	Youden's Index	Accuracy
CDR	92.50	89.4737	0.8197	91.7722
CDR,RDR, ISNT	93.2773	89.7436	0.83021	92.4051

Table 3 Classification accuracies using CCNN (%)

Classifier	Sensitivity (%)	Specificity (%)	Youden's Index	Accuracy (%)
NN	91.7355	89.1897	0.8093	91.1392
CCNN	93.2773	89.7436	0.83021	92.4051

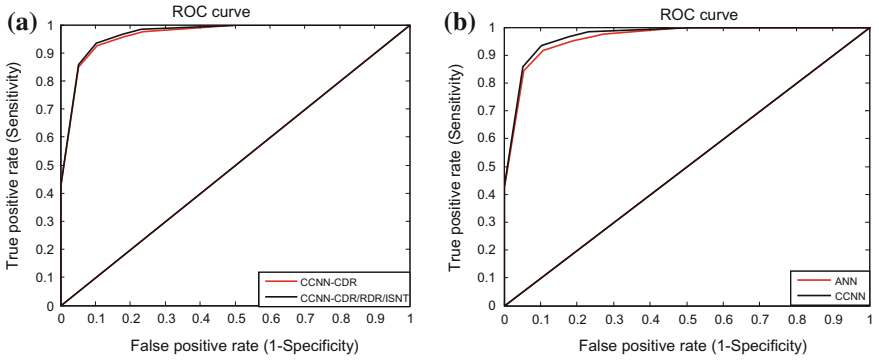


Fig. 7 ROC curve of **a** performance analysis **b** classification accuracies

3.3 Classification Performance

The performance of the proposed glaucoma detection system is evaluated using ROC curve. ROC curve is a plot of the true positive rate (TPR) versus false positive rate (FPR) [23]. The fraction of glaucomatous patients and that are diagnosed as positive is denoted using TPR. FPR is the fraction of normal subjects and that are diagnosed as positive. The glaucoma detection system was tested on 158 real images using the CCNN classifier. Thus the sensitivity, specificity, Youden's index and accuracy using the CDR, RDR and ISNT rule analysis are used to evaluate the performance of the proposed system. The performance of the features in the proposed method is given in Table 2.

The proposed classification method is compared using neural network (NN). Table 3 shows the classification accuracies of CCNN.

ROC curves in Fig. 7a, b show the performance analysis of the features and classification accuracies of CCNN classifier.

Table 4 Glaucoma stage identification

Range of RDR	Disc Damage stage	Number of images in the stage	Stage of Glaucoma
$0.29 < R_I$ or $R_S \leq 0.2$	1	33	Early
$0.19 < R_I$ or $R_S \leq 0.1$	2	66	Moderate
R_I or $R_S < 0.1$	3	12	Advanced

3.4 Glaucoma Stage Identification

In the proposed method, the severity stages 1, 2 and 3 denote the early stage, moderate stage and advanced stage, and these are categorized using the ratios R_I and R_S, the RDR of I and S quadrants. From the classification result of 158 images, it is concluded that out of 128 glaucoma images, 111 images are classified as glaucomatous, and the result is shown in Table 4.

Out of 111 glaucomatous images, 33 images are at the early stage, 66 images are at the moderate stage and 12 images are at the advanced stage. The obtained results approximate the results received from an expert ophthalmologist. The results show that the proposed method facilitates the automatic detection of the early stage of glaucoma, and it improves TNR and FNR.

4 Conclusion

The proposed approach initially does the automatic optic disc localization and ROI extraction by applying the circular Hough transform on the canny edge image of the intensity component of the HSI model of the input image. The OD and OC are segmented using the LMeP feature maps by applying the modified fuzzy C-means clustering technique. The glaucoma parameters CDR, RDR and ISNT values are computed using the segmented OD and OC. The CCNN classifier differentiates the normal and glaucoma subjects. RDR analysis identifies the severity level of the glaucoma subject as early/moderate/advanced. Since no user interaction is required, the results achieved are more accurate and objective compared to manual interpretations, and it facilitates glaucoma detection at an early stage. The optimized neural network can be used in future to get the results more accurately.

References

1. WHO: Glaucoma bulletin (2011). http://www.who.int/bulletin/volumes/82/11/feature1104/en/
2. Inoue, N., Yanishima, K., Magatani, K., Kurihara, T.A.K.T.: Development of a simple diagnostic method for the glaucoma using ocular Fundus pictures. In: 27th IEEE International Conference on Engineering in Medicine and Biology, pp. 3355 –3358 (2005)

3. Marin, D., Gegundez-Arias, M.E., Suero, A., Bravo, J.M.: Obtaining optic disc center and pixel region by automatic thresholding methods on morphologically processed fundus images. Comput. Methods Programs Biomed. **118**(2), 173–185 (2015)
4. Dai, B., Wu, X., Bu, W.: Optic disc segmentation based on variational model with multiple energies. Pattern Recognit. (2016)https://doi.org/10.1016/j.patcog.2016.11.017
5. Ren, F., Li, W., Yang, J., Geng, H., Zhao, D.: Automatic optic disc localization and segmentation in retinal images by a line operator and level sets. Technol. Health Care **24**(2), S767–S776 (2016)
6. Zilly, J., Buhmann, J.M., Mahapatra, D.: Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation (2016). https://doi.org/10.1016/j.compmedimag.2016.07.012
7. Mookiah, M.R.K., Rajendra Acharya, U., Lim, C.M., Petznick, A., Suri, J.S.: Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. Knowl.-Based Syst. **33**, 73–82 (2012)
8. Chen, X., Xu, Y., Yan, S., Wong, D.W.K., Wong, T.Y., Liu, J.: Automatic feature learning for glaucoma detection based on deep learning. In: MICCAI(2015), LNCS, vol.9351, pp. 669–677. Springer (2015)
9. Noronhaa, K.P., Rajendra Acharya, U., Prabhakar Nayak, K., Martis, R.J., Bhandary, S.V.: Automated classification of glaucoma stages using higher order cumulant features. Biomed. Signal Process. Control. **10**, 174–183 (2014)
10. Issac, A., Partha Sarathi, M., Kishore Dutta, M.: An adaptive threshold based image processing technique for improved glaucoma detection and classification. J. Comput. Methods Progr. Biomed. **122**(2), 229–244 (2015)
11. Maheshwari, S., Pachori, R.B., Rajendra Acharya, U.: Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images. IEEE JBHI (2016). https://doi.org/10.1109/jbhi.2016.2544961
12. Devasia, T., Jacob, P., Thomas, T.: Fuzzy clustering based glaucoma detection using the CDR. Signal Image Process.: Int J. (SIPIJ) **6**(3) (2015)
13. Fahlman, S.E., Lebiere, C.: The cascade-correlation learning architecture. NIPS, pp. 524–532 (1989)
14. Devasia, T., Jacob, P., Thomas, T.: Automatic optic disc localization in color retinal fundus images. Adv. Comput. Sci. Technol. **11**(1), 1–13 (2018)
15. Varma, R., Spaeth, G.L., Steinmann, W.C., Katz, L.J.: Agreement between clinicians and an image analyzer in estimating cup-to-disc ratios. Arch Oph. **107**, 526–529 (1989)
16. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing. Prentice Hall Publications (2008)
17. Murala, S., Jonathan Wu, Q.M.: Local mesh patterns versus local binary patterns: biomedical image indexing and retrieval. IEEE JBHI **18**(3), 929–938 (2014)
18. Gomathi, M., Thangaraj, P.: A parameter based modified fuzzy possibilistic C-Means clustering algorithm for lung image segmentation. global. J. Comput. Sci. Technol. **10**(4), 85–91 (2010)
19. Devasia, T., Jacob, P., Thomas, T.: Automatic optic disc boundary extraction from color fundus images. Int. J. Adv. Comput. Sci. Appl. **5**(7), 117–124 (2014)
20. Krieglstein, G., Weinreb, R.: Glaucoma Progress III (Essentials In Ophthalmology). Springer, Berlin Heidelberg (2009)
21. Spaeth, G.L., Henderer, J., Liu, C., Kesen, M., Altangerel, U., Bayer, A., Katz, L.J., Myers, J., Rhee, D., Steinmann, W.: The disc damage likelihood scale: reproducibility of a new method

of estimating the amount of optic nerve damage caused by glaucoma. Trans. Am. Ophthalmol. Soc. **100**, 181–186 (2002)
22. Harizman, N., Oliveira, C., Chiang, A., et al.: The ISNT rule and differentiation of normal from glaucomatous eyes. Arch. Ophthalmol. **124**, 1579–1583 (2006)
23. Medeiros, F.A., Sample, P.A., Zangwilt, L.M., Leibmann, J.M., Girkin, C.A., Weinreb, R.N.: A statistical approach to the evaluation of covariate effects on the receiver operating characteristic curves of diagnostic tests in glaucoma. investigative ophthalmology and visual. Science **47**(6), 2520–2527 (2006)

Performance Evaluation of MRI Pancreas Image Classification Using Artificial Neural Network (ANN)

B. Aruna Devi and M. Pallikonda Rajasekaran

Abstract Magnetic resonance imaging (MRI) has been used in the diagnosis and detection of pancreas tumor. The disadvantage of MRI is long time-consuming in the manual conclusion by a radiologist. Automated classifiers can update the diagnosis activity, in terms of both accuracy and time necessity. This paper is a trial to use artificial neural network (ANN) and least squares support vector machine (LSSVM) to automatically classify 168 human pancreas MR images under two categories, either normal or abnormal pancreas, and the features were extracted by gray-level cooccurrence matrix (GLCM). ANN plays a vital role, specifically in the application of pancreas tumor detection. Only a few reviews are feasible that lead to the improvement of these algorithms to enhance the diagnosis with respect to specificity and sensitivity. LSSVM is a pattern recognition algorithm which learns to assign labels to objects. Classification accuracy compares with the two methods, ANN and LSSVM. ANN provides the best classification accuracy of 96% compared to LSSVM.

1 Introduction

Medical image processing [1] is a powerful analysis used in all stages of cancer assessment. It supports in diagnosing the cancer cells from starting stage to the various steps of its growing stage. Medical imaging involves acquisition of images from the patient's body. The computer and image-processing techniques can provide detailed information about the tumor area. The main concept of CAD [2] is to assist the radiologist to understand the medical images by adopting trustful-computer system to produce "second opinions". CAD can support to diagnose the accuracy of

B. Aruna Devi (✉) · M. Pallikonda Rajasekaran (✉)
Department of Electronics and Communication Engineering, Kalasalingam University, Krishnankoil, India
e-mail: b.arunadevi@klu.ac.in

M. Pallikonda Rajasekaran
e-mail: m.p.raja@klu.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_65

radiologists, reduce the workload, moderate cancer misplaced in behalf of fatigue, over looked or data overburden, and progress inter- and intra-reader variability. The final decision was made by radiologists. Various medical image analyses based on machine learning have been introduced for various applications such as MRI [3], CT, ECG, ultrasound imaging, and microscopy. The integration of various image-processing techniques such as image preprocessing, feature extraction, feature selection, image classification, and image segmentation is essential in CAD system. MRI is an advanced approach that has to be proven in the study of medical image analysis.

MRI provides elaborate information about the organs of our body. The levels of details such as superior contrast and resolution are provided by MRI and it is extraordinary when compared to other imaging modalities. MR images can present biological variations as normal or abnormal information.

The pancreas is one of the important parts of our body. If it is not functioning properly, it will create diabetes, weight loss, nausea, and vomiting. If we are not taking care of pancreas, it will lead to pancreas tumor, pancreatic cancer, acute pancreatitis, adenocarcinoma, and chronic pancreatitis. But if the tumor is found at an earlier stage, it can be cured and patients' survival rate can be increased.

In this research work, MR pancreatic image is analyzed by comparing two methods of classification, namely ANN [4] and LSSVM [5]. ANN is used to aid doctors in solving medical problems. They have been applied in radiology, urology, laboratory medicine, and cardiology. ANN has the preferred feature technique for a large feature dataset task. It is the best tool in recognizing and classifying the images. Radiologists achieve accuracy in recognizing pancreatic tumor between 70 and 88%.

The system using ANN and LSSVM proposed in accuracy achieved for diagnosis for pancreatic tumor. Diagnostic and neural analysis of pancreatic tumor (DANAOPT) showed the results comparable to the results of radiologists. It was also found that the images are hard to recognize by DANAOPT differed from these causing problems to radiologists. Cooperation between humans and computers could therefore lower the probability of mistakes. Results obtained are also dependent on the size and quality of database used.

Least squares support vector machine (LSSVM)

The LSSVM is one of the most efficient machine learning algorithms that is used for image classification. It is used in wider areas along with medicine. LSSVM is a supervised classifier and accurate-learning technique. It is derived from the statistical learning developed by Vapnick in 1982. Diagnosis error obtained by the above-mentioned system was smaller than in systems which uses other known methods.

1.1 Literature Survey

Pavana et al. [6] proposed texture feature extraction by GLCM parameters and Haralick features. Image classification has done by SVM and has the accuracy

91.67%. Only 40 images were used in this paper, and accuracy should be improved. Jingjing et al. [7] described image classification algorithm defects can be reduced based on feature extraction and classification rate. Varsha et al. [8] provided the detection of pancreatic tumor using minimum distance classifier, which gives an average segmentation accuracy of 65.26%. This accuracy level should be improved. Murat et al. [9] performed EUS image classification using ANN that gives an accuracy of 91.7%. In this paper, research is done with EUS images. It is difficult to accurately differentiate diseases on EUS images. Jayaprakash et al. [10] described the feature extraction and content-based image retrieval(CBIR). This paper used 50 images and five textural properties, and this can improve the efficiency of the retrieval system and create a large data bank. Kanika et al. [11] presented the comparison of classification techniques for satellite images. In this paper, KNN provides good accuracy, but it has high time complexity. Harshavardhan et al. [12] described GLCM method of grayscale analysis. Mohanaiah et al. [13] presented GLCM method to extract texture features from the motion estimation of images. In this paper, features of four only computed from FPGA. It is used for real-time applications. Manavalan et al. [14] compared feature extraction using GLCM method, combining GLCM with other methods.

This paper gives more accuracy of computations. Chang et al. [15, 16] described the SVM is a method in sonography for the detection of breast cancer. Savkare et al. [17] classified the blood images using an SVM classifier and identified both normal and infected cells. In this method, there is no comparison of classification. Luiza Antonie [18] shows an SVM classifier which was used with statistical features for automated segmentation and classification of brain MR images. Fritz et al. [19] analyzed GLCM large no of texture features were useful to train the images, proved smallest classification error. Pawan et al. [20] proposed the image classification using SVM classifier. This paper lacks in comparative analysis of classification. Gering et al. [21] proposed the EM algorithms in the diagnosis of anomalisms. These algorithms are described to distinguish tumors from the adjoining brain tissues by training entirely on normal brain images in order to observe anomaly from uniformity. This needs long computations. The observation-based approaches describe segmentation and classification results, but these approaches stand in need to complete the training. In medical image investigation, the decision of tissue type and classification of tissue are achieved by utilizing texture. MR image texture is proven to be helpful to resolve the tumor type (Schad et al.) [22].

To resolve the texture classification dispute, many techniques have been established in past years such as multichannel methods, multi-resolution analysis, level set, Gabor filters, and wavelet transform (Dunn et al. Chang et al.) [23, 24]. Gabor filters are low due to their shortage of orthogonality that reaction in redundant features at disparate scales. While wavelet transform is reporting textures at the acceptable scale, by changing the spatial resolution. There is a challenging task in recruiting the features to discriminate between classes. The assessment of feasible feature subsets is mostly a hurtful task due to multiple computational attempts required. Zhou et al. [25] presented three types of features such as spatial, spectral, and temporal. Danielle et al. [26] compared three spatial feature extraction

methods—GLCM, SST, and TS. GLCM and SST give the improved classification accuracy. Conners et al. [27] suggested that concurrence matrix does not discriminate the visual pairs.

Based on this survey, we conclude that the image should be discriminated by second-order statistical texture feature GLCM and classified by a suitable classifier ANN and LSSVM). This will result that the image is normal or abnormal. Classification accuracy, sensitivity, specificity, and computational time are compared among the two methods,and ANN provides the better classification accuracy.

2 Materials and Methods

Input dataset:

The input dataset consists of MR pancreas images (168) which is collected from the KGS Scan center, Madurai, which are real human pancreas images. The number of MRI pancreas images 168 of which are 100 are normal and 68 are abnormal images. The pancreas images are in the plane of axial and have dimension 512 * 512. Figure 1 presents the sample normal images, and Fig. 2 presents the sample

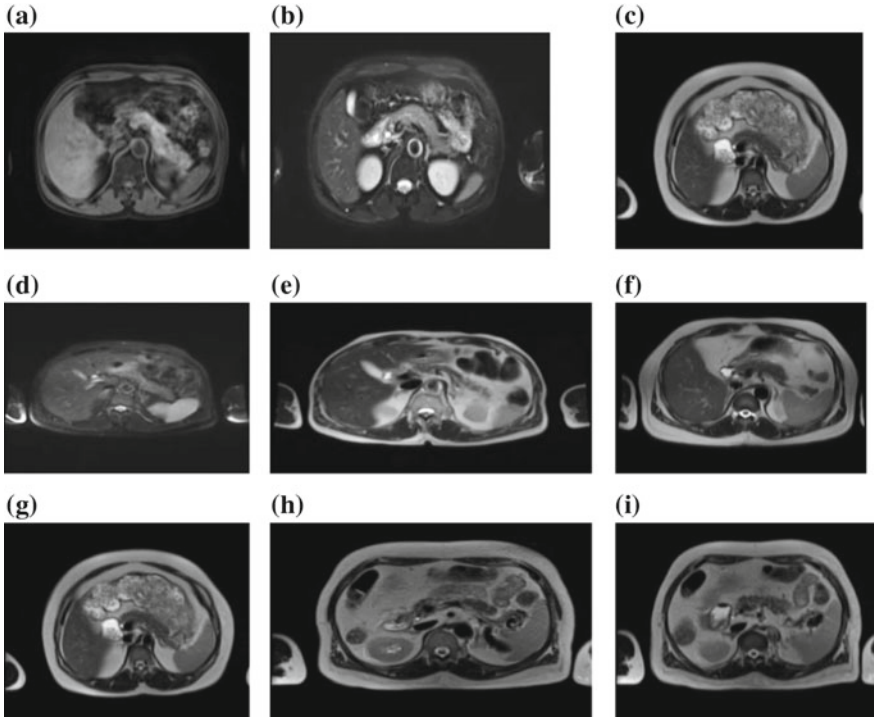


Fig. 1 Sample normal images

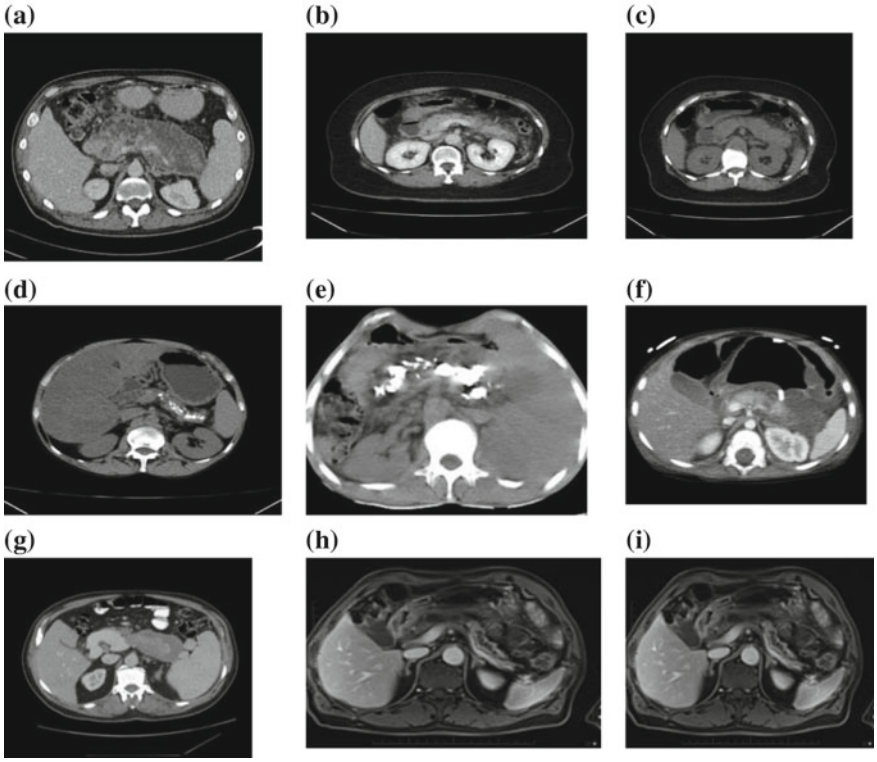


Fig. 2 Sample abnormal images

abnormal pancreas images. In our method, first of all, we extract features from the image using GLCM method. Ten feature dataset values are given to the classification stage based on ANN and LSSVM methods, we can conclude which method is best to identify the image is normal or abnormal. Figure 3 presents the methodology of classification techniques.

3 Feature Extraction

GLCM has an effective-tool basis for the use in texture. It is a stage of transforming images into a vector for further classification. We use Haralick (1973) feature extraction, namely the gray-level cooccurrence method (GLCM). This well-known statistical approach for extracting second-order texture information is placed on the evaluation of the discrete second-order probability function C $(I, J/\Delta x, \Delta y)$ (Haralick 1973). It helps to explain the fine points around the total image content. In this paper, the image is discriminated into ten feature vectors. The ten feature vectors are energy, entropy, correlation, contrast, homogeneity, cluster prominence,

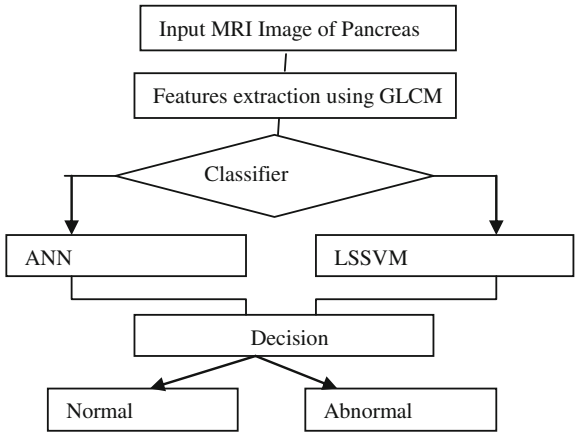


Fig. 3 Methodology of classification techniques

cluster shade, sum of variance, sum of average, and information measure of correlation. These features are extracted by using GLCM method. The feature dataset is a combination of ten feature vector values. GLCM method gives smallest classification error when using large feature dataset. This dataset based on ten features which gives more classification accuracy. GLCM not only depends on single space inter-sampling but also provides more inter-sampling distances between pixels. Energy: $\sum_i \sum_j g_{ij^2}$, energy belongs to the pixels repetition rate. Contrast: $\sum_i \sum_j (i-j)^2 g_{ij}$, contrast is the local intensity variation. Entropy: $\sum_i \sum_j g_{ij} \log_2 g_{ij}$, entropy relates to the inhomogeneous images and homogeneous images. Inhomogeneous images have low first-order entropy, while a homogeneous image has a high entropy. Correlation: $\sum_i \sum_j (ij) g_{ij} - \mu_x \mu_y / \sigma_x \sigma_y$, correlation is a gray-level linear dependence between pixel and neighborhood. Sum of variance: $\sum_{i=2}^{2n} (i-sa)^2 g_{x+y}(i)$, sum of variance refers to high weights of the elements from the average value. Cluster shade and cluster prominence are used for a sum histogram of images. Homogeneity: $\sum_i \sum_j (1/1+(i-j)^2) g_{ij}$, homogeneity gives a difference in histogram of images. Sum of average, information measure of correlation are $\sum_{i=2}^{2n} i g_{x+y}(i)$, (IMC1) = HXY − HXY1/ max{ HX, HY}, HXY = $-\sum_i \sum_j g_{ij} \log_2 g_{ij}$ where HX and HY are entropies of g_x *and* g_y, HXY1 = $-\sum_i \sum_j g_{ij} \log_2 \{g_x(i) g_y(j)\}$, respectively. These values are arranged in an excel file. These parameters are calculated using GLCM method and create GLCM feature dataset values. This feature dataset value is extracted from MR images and analyzed by the classification techniques.

4 Image Classification Techniques

4.1 ANN Classification Technique

An ANN is a biologically inspired nervous system. The ANN has a large number of processing elements termed as neurons, which is interconnected in the network. Neural networks are trained by knowing feature dataset. Once the feature dataset is trained, it solves the prediction values of unknown data. ANN has multiple inputs, outputs, and hidden layers. In this research work, input layer is assigned to be 11 neurons, which is a feature dataset. Output layer is a single neuron which informs that whether the image is normal or abnormal. Hidden layer is assigned to ten feature datasets. Classification accuracy for this system is 96%.

4.2 LSSVM Classification Technique

LSSVM is an effective base for binary classification. It has an ability of generating fast classifying functions followed by training. MSE is mathematically for variance analysis. For example, MSE is used to calculate the error variance. LSSVM is the structural risk minimization concept from the statistical learning technique. Its kernel is to control the practical risk and classification ability in order to maximize the margin between the classes and minimize the true costs. Ten feature subsets extracted from GLCM method are the inputs to the LSSVM algorithm. In our method, normal and abnormal pancreases are the two classes. Each subject is represented by a vector in all images. RBF kernel function is proven to be easeful in LSSVM classification. We use RBF kernel function, and it provides the classification accuracy 91%.

5 Training and Testing

Feature dataset values are analyzed by the classification stage such as ANN and LSSVM. The entire feature data set is divided into two parts—one for training and other for testing. ANN is a neural network training tool in MATLAB. The input layer comprised of 11 neurons, hidden layer compred of 10 neurons, and the output layer having single neuron which describes normal or abnormal image. During training, ANN is trained by the input feature vectors and are discriminated from the given pancreas MRI images. 10-fold cross validation method is applied to train the neural network. The advantage of applying 10-fold cross-validation is that it lowers the error rates and gives a good error estimate MSE. After the completion of training phase, the unknown pancreas MRI normal and abnormal testing images are given to the ANN classifier. The ANN classifier classifies the given images into

Table 1 Result of classification methods

Classification techniques	Classification accuracy	Sensitivity	Specificity	Training time	Testing time
ANN	96.67	97.3	95.65	2.0748	1.4664
LSSVM	91.67	89.66	94.74	0.468	1.4352

normal or abnormal based on the extracted features. The testing of an ANN is done by simulating the ANN with testing set, and then calculate the MSE. Out of 168 images, 30% of images were utilized for training and 70% of images were utilized for testing. Accuracy, sensitivity, specificity, training time, and testing time were tabulated in Table 1. ANN provides 96% classification accuracy.

In LSSVM method, we should specify kernel functions to be used. Here, the radial basis function [30] is used. The ten feature subsets are used as inputs of LSSVM classification. The approach is recycled to fix its best parameters is a grid chase applying a cross-validation. Cross-validation method with ten folders is recycled to hunt the best parameters with an interval of values that perform a great accuracy. Among all 168 images, we have selected 30% for training and 70% for testing. The training set was utilized to train the LSSVM, and the testing set was utilized to assess the performance of the LSSVM. Results are shown in Table 1. LSSVM provides 91% classification accuracy.

Next, the performance assessment methods are used to check out the two techniques ANN and LSSVM. We survey the execution of the two methods in terms of sensitivity, specificity, and accuracy. The three terms are defined in Eqs. (2)–(4):

$$\text{Sensitivity}=TP/(TP + FN)\ 100\% \quad (2)$$
$$\text{Specificity} = TN/(TN + FN)\ 100\% \quad (3)$$
$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN)\ 100\% \quad (4)$$

Where

True positives (TPs) = accurately classified positive cases,
True negatives (TNs) = accurately classified negative cases,
False positives (FPs) = inaccurately classified negative cases,
False negatives (FNs) = inaccurately classified positive cases.

Sensitivity is the ratio of accurately classified positives; represent superior performance of the technique in predicting positives. Specificity measures how perform the system can estimate the negatives. Accuracy measures the overall exactness of the classifier in estimating both positives and negatives.

6 Result and Conclusions

In this research work, LSSVM and ANN are trained by a GLCM feature extraction dataset and classified the image as normal or abnormal. ANN provides the best results compared to LSSVM. ANN had MSE value $0.009398 * 10^{-07}$ at epoch 28.

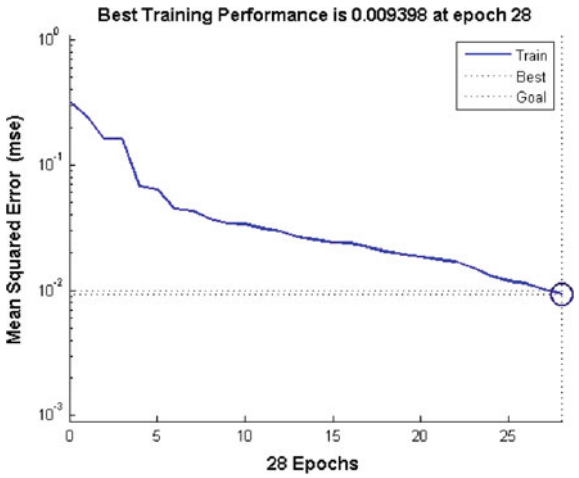


Fig. 4 Best training performance

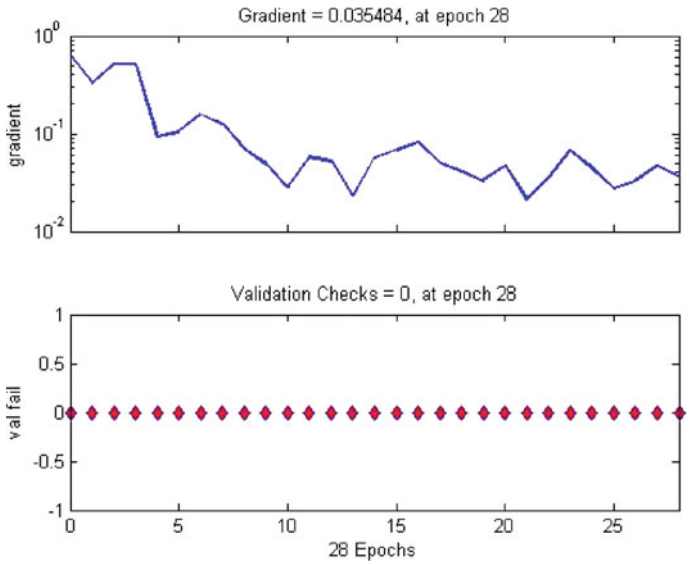


Fig. 5 Represents the gradient

Classification accuracy of the network is 96%. Sensitivity and specificity are 97% and 95%, respectively.

In LSSVM method, the classification accuracy value is 91%, and sensitivity and specificity are 89% and 94%, respectively. The computation time for ANN is 2.07 s, and for LSSVM, it is 0.468 s.

ANN also predicts the image as normal or abnormal which is useful to pancreas tumor-affected patients and is easy to identify tumor by the radiologist.

Table 1 describes the result of two classification methods. Figure 4 shows the best training performance. Figure 5 represents the gradient.

This paper does not require any ethical approval. Authors will be solely responsible if any issues arise in the future with regard to this.

Acknowledgements The authors thank the management of Kalasalingam University for providing financial assistance under the scheme of University Research Fellowship (URF). Also, we thank the Department of Electronics and Communication Engineering, Kalasalingam University, Tamil Nadu, India, for permitting to use the computational facilities available in Center for Research in Signal Processing and VLSI Design which was set up with the support of the Department of Science and Technology (DST), New Delhi, under FIST Program in 2013 (Reference No: SR/FST/ETI-336/2013 dated November 2013). Also, we thank KGS Scan center, Madurai for providing the pancreas images without which this work would not have been possible. This database is not publicly available, and we thank Dr. Srinivasan, KGS Scan center, Madurai, for permitting to use the MRI images which are more useful in this research work.

References

1. Rafael, C.G., Woods, R.E.: Digital Image Processing, vol. 1, 2nd edn. Prentice Hall (2002)
2. Pratt, W. Digital Image Processing, vol. 3. Wiley (2001)
3. Armstrong, T.S., Cohen, M.Z., Weinbrg, J., Gilbert, M.R.: Imaging techniques in neuro oncology. Semin. Oncol. Nurs. **20**(4), 231–239 (2004)
4. Reddick, W.E., et al.: Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. IEEE Trans. Med. Imaging **16**(6), 911–918 (1997)
5. Shubhangi, D.C., et al.: Support vector machine (SVM) classifier for brain tumor detection. In: International Conference on Advances in Computing, Communication and Control, pp. 444–448 (2009)
6. Pavana, S., et al.: Diabetes Mellitus Detection Based on Facial Texture Feature Using the GLCM. IRJET **04**(07) (2017)
7. Jingjing et al.: Application of Visual Saliency and Feature Extraction Algorithm Applied in Large-Scale Image Classification (2016)
8. Turkar, V., et al.: Pancreatic tumor detection using image processing. Proc. Comput. Sci. **49** (2015), 11–16 (2015)
9. Ozkan, M., etal.: Age-based computer-aided diagnosis approach for pancreatic cancer on endoscopic ultrasound images. EURO-EUS Sci. J. **5**(2) (2015)
10. Jaya Prakash, K. et al.: (2012)"Development of Pancreatic CT-Scan image dataset and Retrieval Process for Diagnosis" Journal of Scientific and industrial research
11. Kanika, et al.: A comparative study of supervised image classification algorithms for satellite images. Int. J. Electr. Electron. Data Commun. **1**(10) (2013). ISSN: 2320-2084
12. Harshavardhan et al.: GLCM architecture for image extraction. Int. J. Adv. Res. Electron. Commun. Eng. (IJARECE) **3**(1) (2014)
13. Mohanaiah et al.: Image texture feature extraction using GLCM. Int. J. Sci. Res. Publ. **3**(5) (2013). 1 ISSN 2250-3153
14. Manavalan et al.: Comparative analysis of feature extraction methods for the classification of prostate cancer from TRUS medical images. IJCSI Int. J. Comput. Sci. Issues **9**(1), 2 (2013). ISSN (Online): 1694-0814
15. Chang, R.F., et al.: Improvement in breast tumor discrimination by support vector machines for diagnosis of breast tumors on US images. Acad. Radiol. **10**(2), 189–197 (2003)

16. Chang, R.F., et al.: Support vector and speckle-emphasis texture analysis. Ultrasound Med. Biol. **29**(5), p. 679–686 (2003)
17. Savkare et al.: A comparative study of supervised image classification algorithms for satellite images. Int. J. Electr. Electron. Data Commun. **1**(10) (2012). ISSN: 2320-2084
18. Antonie, L.: Automated segmentation and classification of Brain Magnetic Resonance Imaging. C615 Project (2008). http://en.scientificcommons.org/42455248
19. Fritz et al.: Statistical Texture Measures Computed from Gray Level Coocurrence Matrices (2008)
20. Pawan et al.: Rough Set Based 1-v-1 and 1-v-r Approaches to Support Vector Machine Multi-classification, vol. 177, Issue 18, pp. 3782–3798. Elsevier (2014)
21. Gering, D.T.: Recognizing deviations from normalcy for brain tumor segmentation". Lect. Notes Comput. Sci. **2488**, 388–395 (2002)
22. Schad, L.R.: MR tissue characterization of intracranial tumors by means of texture analysis. Magn. Resonan. Imaging **11**(6), 889–896 (1993)
23. Dunn, C.: Optimal Gabor filters for texture segmentation, IEEE Trans. Image Process **4**(7), 947–964 (1995)
24. Chang, T.: Texture analysis and classification with tree structured wavelet transform. IEEE Trans. Image Process. **2**(4), 429–441 (1993)
25. Zhou et al.: Geographical Understanding and Analyses of Remotely Sensed Imagery (1999)
26. Danielle et al.: A comparison of spatial feature extraction algorithms for land-use classification with spot hrv data (1992)
27. Conners et al.: A theoretical comparison of texture algorithms",IEEE Trans. on Pattern Analysis and Machine Intell. PAMI-2, 204–222 (1980)

User Behavioral Characteristics Identification from Mobile Call Logs

Sravya Paritala, Rishitha Bobba, Haritha Akkineni, V. S. Lakshmi Papineni and Lakshmi Gogulamudi

Abstract There is a huge penetration of mobile phones and is forecasted to progress in its growth. By 2019, there will be five billion users of mobile phone (Statista - The statistics portal for market data, market research and market studies, (2017), [1]). According, Statista – A portal for statistics. This tremendous usage of mobile phones created lot of call log data. This work concentrates on one of the biggest common and woeful issue which relates to the subjective prosperity of the humanity which is the stress. Clustering and Correlation algorithms were applied on call log data. Various behavioral characteristics like mental overload, disturbed sleep, somatic complaints, psychological distress etc and abnormal behaviors were inferred which could enable the decision makers to draw out valuable insights. Finding the association between mobile phone use and behavioral characteristics of users is the main intention of this investigation. This investigate can be of vital value to professionals, analysts, medical experts and therapists who study the patient's cases.

S. Paritala · R. Bobba · H. Akkineni (✉) · V. S. L. Papineni · L. Gogulamudi
Department of IT, PVP Siddhartha Institute of Technology, Vijayawada, India
e-mail: akkinenih@gmail.com

S. Paritala
e-mail: paritalasravya2@gmail.com

R. Bobba
e-mail: bobbarishitha20@gmail.com

V. S. L. Papineni
e-mail: papinenivsl@gmail.com

L. Gogulamudi
e-mail: lakshmineeraja@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_66

1 Introduction

Mobile phones and smart phones have changed the way of communication. The most commonly used communication tool today is a mobile phone. They have become a mainstay in the lives of most people. Mobile phone provides a means of a wireless form of a communication which has impacted emotional and mental conditions of people in many areas of the world. The collection of call log data of a phone provides a huge insight of the behavior of a person. This insight can be done through data analytics. This mobile analytics gives the ability to turn information into insights, which truly influence the person's behavior, interests, stress levels, mental overload, disturbed sleep, somatic complaints, psychological distress etc and abnormal behaviors etc. – as well as a deep understanding of how those individuals behave at a particular time. Moreover, these insights determine how much influence does mobile phone has on their behavior in terms of stress.

Stress has become a major concern, as it is turning into the underlying driver for many diseases. Because of it, the life expectancies and personal contentment would certainly be affected. Stress is a hazard factor for colossal number of diseases likewise on the grounds that stress is firmly connected to unhealthy behavior. [2, 3]. Stress is a prominent condition in the present day life and has demonstrated that measure of aggregate stress takes a great part in an expansive range of physical, mental and behavioral conditions such as tension, social isolation, depression, sleep, cognitive impairments, and immunological disorders, neurodegenerative diseases and other therapeutic conditions [4], Thus, measuring stress in day by day life circumstances has turned into a vital challenge [5]. An early forecast of stress symptoms can undoubtedly anticipate circumstances that are hazardous for human life. [6].

A disputable report has recommended that, there is a danger of brain cancer up to 40 percent if we use a mobile phone for just thirty minutes per day. As indicated by World Health Organization research, people who used the device more than ten years were more likely to be diagnosed with cerebrum tumors.

In this paper, the main concentration is on call log dataset and applying clustering and correlation algorithm. On analyzing, the call logs by applying various clustering algorithms, we can identify the people who are more stressed and have any somatic complaints so that we can take precautionary measures. So finally, a methodology was proposed to identify behavioral attributes like stress, mental overload, disturbed sleep, somatic complaints, psychological distress etc and abnormal behaviors. As exact and proficient evaluation of unhealthy stress exposure could enormously help to carry out timely and appropriate support measures.

2 Related Work

Varied range of works has been done in this area. The current investigations have concentrated on understanding aspects of human mental conditions. These days various types of information can be extracted from the telephones, the analysis and data mining of call details is as yet a developing subject. This information is more valuable in the telecom business, where they can acquire helpful data about their clients. The mental conditions of a particular person can be predicted by analyzing his/her calls data. In [7] author recommends user subscription plan by analyzing mobile call log.

The information in mobile devices found at crime sites might provide investigators a few snippets of valuable pieces of information. The field of digital forensics which investigates any digital data at the crime scene or related to the crime now encompasses mobile phone data as well. A review of various cell phone advancements, equipment, and examination instruments is given in [8]. Has become out of the standard routine with regards to computer legal sciences.

Digital information that is verifiably created in mobiles has been broke down to infer relationship amongst areas and travel requests for various sorts of social events in 9. A comparative study has been made on mobile phone usage data of an upsetting (exam period) and an ordinary week is looked at in [10] for seven subjects. In [11] wrist sensor data and smart phone data has been analyzed to foresee high and low levels of worry, as measured by the full PSS poll.

3 Investigate on Basic Call Log Data

The fundamental data investigation depicted accumulates straightforward statistics on a single data set, the mobile call information. For instance, if we examine the past voice call data comprising the basic data such as contact numbers of dialed, incoming, and missed calls, and record of calls over a year, timestamp, and durations of incoming and dialed calls. The data about the most called/received numbers could be utilized beneficially to naturally give feelings of anxiety among the users. For instance, if the user is talking over the mobile phone for a longer duration, he/she may have high stress level and difference in the behavioral features.

The study group consisted of students in a CUG (Close User Group). Individual call log data was collected from their own phone call history given by the service provider. 80,000 records have been gathered for analysis. The sample of the data is shown in the given figure. Sampling has been done for analysis purpose (Fig. 1).

ExampleSet (49 examples, 0 special attributes, 7 regular attributes) Filter (49 / 49 examples): all

Row No.	user	other	direction	duration	timestamp
1	7680945483	9494114249	Incoming	211	Wed Mar 15 19:17:44 +0100 2017
2	7396666120	9666723917	Outgoing	31	Mon Feb 11 07:18:23 +0000 2017
3	7396666120	9542115015	Incoming	45	Mon Feb 11 07:45:42 +0000 2017
4	7396666120	9494954295	Outgoing	10	Mon Feb 11 08:04:42 +0000 2017
5	7396666120	9494954295	Outgoing	0	Mon Feb 11 08:05:31 +0000 2017
6	7396666120	9494954295	Incoming	0	Mon Feb 11 08:06:18 +0000 2017
7	7396666120	7382222259	Outgoing	0	Mon Feb 11 08:06:31 +0000 2017
8	7396666120	9441151313	Incoming	124	Thu Feb 09 19:35:37 +0100 2017
9	7382222259	8977392604	Outgoing	474	Thu Feb 09 18:43:44 +0100 2017
10	7382222259	8790873576	Missed	0	Thu Feb 09 19:51:30 +0100 2017
11	7382222259	8977392604	Outgoing	0	Thu Feb 09 20:57:55 +0100 2017
12	7382222259	8977392604	Outgoing	605	Fri Feb 10 20:17:00 +0100 2017
13	7382222259	8977392604	Outgoing	1	Sat Feb 11 16:44:56 +0100 2017
14	7382222259	8790873576	Outgoing	59	Sat Feb 11 17:00:59 +0100 2017
15	7382222259	8977392604	Outgoing	1201	Mon Feb 13 14:53:26 +0100 2017
16	7382222259	7396666120	Outgoing	2	Tue Feb 14 07:19:17 +0100 2017

Fig. 1 Call log data

4 Methodology

Our motive to survey whether there are relationship between cell phone utilize and psychological wellbeing effects in a CUG of students. The stress levels can be inferred from data stored on people's mobile phones, using their contact lists, call logs. We experimented on the data using the clustering algorithm, k-means and Pearson Correlation Coefficient. By using k-means algorithm we have identified the group of users who are attending the calls for a long duration and whose mobile phone activity is high. We designed a survey form questioning the issues concerning health, psychosocial factors and mobile phone use. The number of mobile calls made or received, the number of texts made or received, possibility to be disturbed at all hours, somatic complaints, anxiety, staying up late to use the mobile phone and insomnia, psychological distress and unhealthy life style, reach ability of the receiver were the topics which were queried upon. The survey forms were manually distributed among the student community. Out of the approximately 80 students, a total of 49 responded by the collection date. Based on the frequency distribution, the feedback received was divided into low, medium and high categories. Data was collected over a three month period from the month of June 2017 (Fig. 2).

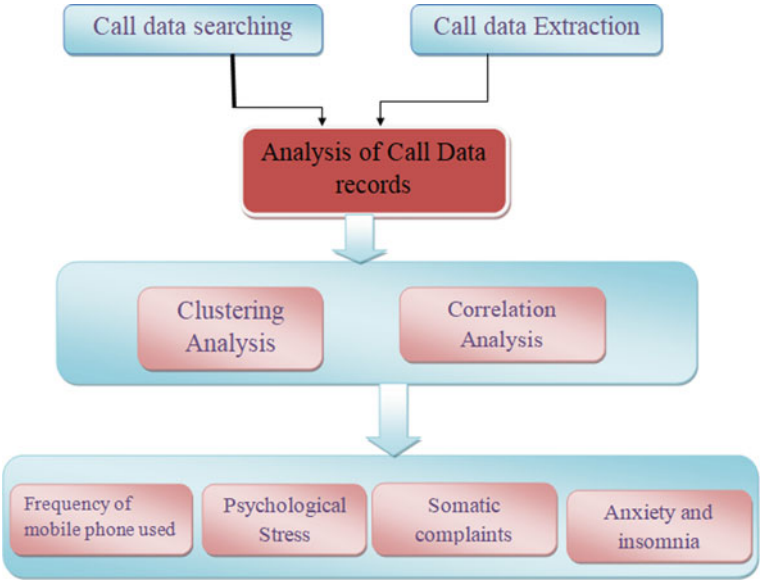


Fig. 2 Methodology

5 Proposed Work

To identify person's behavioral characteristics and stress levels based on the mobile call logs is the main intention of the work. The k-means clustering mechanism divided the mobile call logs into 5 groups based on the mixed Euclidian distance to find out the similarity between the calling patterns. The frequency of mobile phone usage based on the duration of the call was also identified. This gives us the users who were mostly using mobiles. A statistical software package SAS was used for the analysis purpose.

Pearson correlation analysis was used to examine association between the mobile phone exposure variables and the mobile phone usage. The simple statistics like mean and standard deviation, the minimum and maximum values were calculated.

Correlation analysis is used to measure the degree of relationship between the variables under consideration. Correlation coefficient is the measure of correlation. The degree of relationship is expressed by coefficient which range from correlation ($-1 \leq r \geq +1$). The correlation analysis enables us to have an idea about the degree & direction of the relationship between the two variables under study. If the correlation is positive, it means that when one variable increases, the other tends to increase. If the correlation is negative, it means that when one variable increases, the other tends to decrease. If r is getting close to 1 we can infer that there is a

strong relationship between variables. If it is getting close to 0 then there is weak relationship [12].

So the clustering algorithm [13] is used to find out the group of users whose mobile activity is more and the data inferred from the survey finds out the association between the qualitative aspects of mobile phone use with stress and the frequency of mobile phone use, mental overload, and change in behavioral characteristics of a person. The numbers with highest mobile activity was taken from the mobile call log data. The same person has answered that his/her mobile activity as high which was derived from survey form.

From the transitive property $((p \leftrightarrow q) \& (q \leftrightarrow r)) \rightarrow (p \leftrightarrow r)$. We came to a conclusion that the mobile numbers from the call log data and the survey forms whose mobile activity is intensive tend to suffer from distress, somatic disorders etc.

6 Experimental Results and Discussion

The call log data set contains 83000 records. The record set has attributes such as user, other, direction, duration and timestamp. Out of these data sets, we used sampling and applied clustering algorithms on 50 data sets for better understanding. All the attributes are taken into consideration for clustering. The same dataset is clustered by algorithm such as k-means. In this clustering technique, the call log records get clustered based on distances. The different parameters used here are k = 5, maximum optimization = 100 which specifies the maximum number of iterations for one run of k-means.

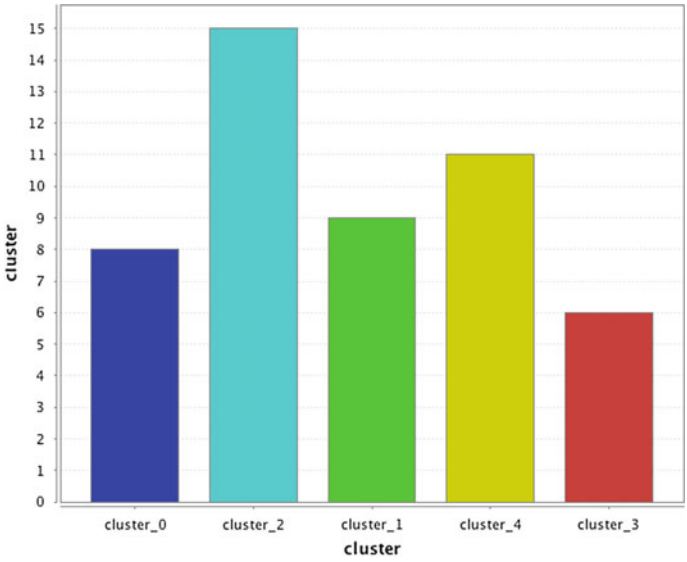


Fig. 3 User grouping into different cluster

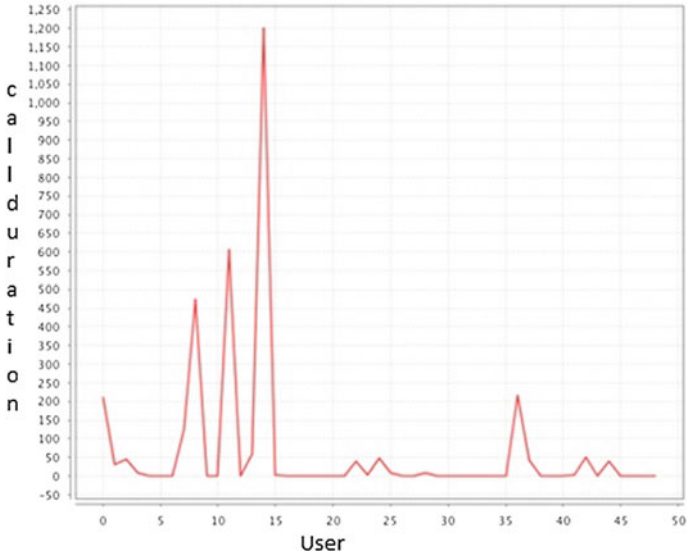


Fig. 4 Duration of call logs

The Fig. 3 represents users in different clusters that were clustered based on distances. Since k = 5 there are 5 clusters. Here the clustered data visualization is done using Rapid miner studio tool. The users with similar calling patterns were grouped into cluster which is depicted in the graph.

The Fig. 4. represents the statistics based on duration of calls. From the above graph, we can observe the users call duration where we can identify the user who uses his/her phone for more time.

Pearson Correlation Coefficients
Number of Observations

	somatic_complaints	possibility_to_be_disturbed_at_a	anxiety_and_insomnia
frequency_of_mobile_phone_used frequency of mobile phone used	0.26812 50	0.23856 50	0.47703 49

Pearson Correlation Coefficients
Number of Observations

	pschyological_distress_and_unhea	how_often_the_receiver_is_availa
frequency_of_mobile_phone_used frequency of mobile phone used	0.7931 50	-0.36109 50

Pearson Correlation Coefficients
Number of Observations

	text_messages_made_or_receive_ea	staying_up_late_to_use_the_mobil
frequency_of_mobile_phone_used frequency of mobile phone used	-0.31888 50	-0.03791 50

Fig. 5 Association using Pearson correlation coefficients

In Fig. 5. Frequency of mobile phone usage is positively correlated with somatic complaints, possibility of being disturbed at night, anxiety and insomnia, psychological distress and unhealthy life style. There is a weak correlation between values Frequency of mobile phone usage and somatic complaints as well as possibility of being disturbed at night as the value are 0.26812, 0.23856 respectively which are closely tending to zero. There is a neutral association between mobile phone usage and anxiety and insomnia, whereas mobile phone is strongly associated with psychological distress and unhealthy life style as the value is 0.7931 and it tends to 1. The accessibility of the receiver, text messages made, staying up late to use the mobile are inversely proportional to the usage of mobile phone. There is a weak association between the variables as almost all parameters are towards zero.

7 Conclusion

The main goal of this paper is to identify the users whose mobile phone usage is more. The more they use the more they are stressed. We solved it by applying k-means and Pearson correlation analysis. The results convincingly suggest that this analysis is sufficient to identify the person with more stress. There are likely relationship between cell phone factors and behavioral results among the students. High recurrence of cell phone use at was a hazard factor for announcing sleep disturbances and symptoms of somatic complaints at 6 months follow up. The hazard for revealing mental health symptoms at check-up was huge among those who had reported that they perceived the psychological distress and unhealthy life style to be stressful. This stress recognition strategy in view of mobile phone data can take leverage of broad utilization; it can be connected in a several real world situations. As such, the aim of this study was to see if there was an association between mobile phone use and behavioral characteristics of users in a CUG. This investigate can be of vital value to professionals, analysts, medical experts and therapists who study the patient's cases.

References

1. Statista - The Statistics Portal for Market Data, Market Research and Market Studies. https://www.statista.com/ (2017)
2. Adam, T.C., Epel, E.S.: Stress, eating and the reward system. Physiol. Behav. **91**(4), 449–458 (2007)
3. Schwabe, L., Dickinson, A., Wolf, O.T.: Stress, habits, and drug addiction: a psychoneuroendocrinological Perspective. Exp. Clin. Psychopharmacol. **19**(1), 53 (2011)
4. Cohen, S., Kessler, R.C., Gordon, L.U.: Measuring Stress: A Guide for Health and Social Scientists. Oxford University Press, Oxford (1997)
5. Plarre, K., Raij, A., Hossain, S., Ali, A., Nakajima, M., Al'absi, M., Ertin, E., Kamarck, T., Kumar, S., Scott, M., Siewiorek, D., Smailagic, A., Wittmers, L.: Continuous inference of

psychological stress from sensory measurements collected in the natural environment. In: 10th International Conference on Information Processing in Sensor Networks (IPSN), pp. 97–108 (2011)
6. Healey, J., Picard, R.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Trans. Intell. Transp. Syst. **6**(2), 156–166 (2005)
7. Adinarayana, S., Yogeswara Rao, K.: User plan recommendation using mobile call log analysis. Int. J. Innov. Res. Comput. Commun. Eng. **4**(9) (2016)
8. Punja, S.G., Mislan, R.P.: Mobile device analysis. small scale digital device Forensics. Journal **2**(1), (2008)
9. Calabrese, F., et.al.: Analyzing cell-phone mobility and social events. Workshop on the Analysis of Mobile Phone Networks. pp. 59–61. MIT, Cambridge (2010).
10. Bauer, G., Lukowicz, P.: Can smartphones detect stress-related changes in the behavior of individuals. In: International Conference on Pervasive Computing and Communications Workshops, pp. 423–426. IEEE (2012)
11. Sano, A., Picard, R.W.: Stress recognition using wearable sensors and mobile phones. In: Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 671–676. IEEE (2013)
12. Talukdar, J., Sanjib, K. K.: Analysis on retrospective cardiac disorder using statistical analysis and data mining techniques. Int. J. Appl. Eng. Res. 6778–6787 (2017)
13. Akkineni, H., Papineni, V. S. L., Burra, V.B.: Hybrid method for framing abstractive summaries of tweets. Int. J. Intell. Eng. Syst. (2017)

Local Contrast Regularized Contrast Limited Adaptive Histogram Equalization Using Tree Seed Algorithm—An Aid for Mammogram Images Enhancement

V. Muneeswaran and M. Pallikonda Rajasekaran

Abstract Despite the availability of huge volume technical advancement and numerous diagnostic tools, identification and spotting out of breast cancer at an early stage is still a challenging task. In this paper, a tree seed optimization (TSA) algorithm for optimizing the enhancement parameter of Local Contrast (LC) regularized Contrast Limited Adaptive Histogram Equalization (CLAHE) is presented. The TSA method of tuning the enhancement factor of LC-CLAHE achieves very good enhancement of mammogram images compared to other contrast enhancement methods available in the literature. Images from MIAS database is used for testing the proposed method. Contrast Improvement Factor (CIF) is used as an effective quality testing parameter. Experimental results explicate that the proposed method preserves the local information present in the image and also provides optimized contrast enhancement.

1 Introduction

Statistics of Oncology in India portrays that the breast cancer increases the mortality rate of women in a large percentage. Early recognition of breast tumor is critical and mammography is the essential imaging system for the location of it. Probability of successful treatment relies on the earlier identification of breast cancer. The essential objective of mammography screening is to identify little, non-discernable diseases in its beginning time. Because of poor perceivability, low complexity, and uproarious nature of mammograms, it is hard to interpret the neurotic changes of the breast [1]. Screening mammography is usually preferred on the grounds as it incites the malignant breast tumor and its subtle signs such as microcalcification and mass

V. Muneeswaran (✉) · M. Pallikonda Rajasekaran
Kalasalingam Academy of Research and Education, Krishnankoil 626126, Tamil Nadu, India
e-mail: munees.klu@gmail.com

M. Pallikonda Rajasekaran
e-mail: m.p.raja@klu.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_67

lesions. Due to the low contrast of mammographic image that occurs during the acquisition process, the subtle signs of breast cancer could not be easily visualized by the radiologist. The distinguishing level of contrast between the affected tissue and the normal tissue lies out of the range from human perception and radiologist expertise. Image enhancement is necessary for all computer-aided diagnosis [2, 3]. The two phases in which the process of image enhancement can be introduced are

- During the acquisition process
- Postprocessing stage of input image

The latter is usually preferred as the former incurs due to increasing acquisition time [2]. Consequently, improving the visual quality of the image turns out to be essential while screening mammograms.

2 Literature Review

Researchers across the globe have contributed a lot for medical image processing [4, 5]. A few among them in the context of mammogram image processing was described in this section. Harshul Srivatava and his research team proposed a novel quality evaluation index for mammogram image. The quality evaluation index is based on tangent algebra-based nonlinear operators [6]. A contributive approach to brightness preservation and contrast enhancement was proposed by Shanmuga Vadivu et al., Enhancement of image is obtained by dividing the histogram of the input image using mean as a dividend [7]. Langarizadeh et al. used the three basic processes such as Histogram Equalization, median filtering, and histogram searching, which were simultaneously used to bring out improvement in digital mammogram images [8]. Fuzzy techniques in defining the clip limit of Contrast Limited Adaptive Histogram Equalization were introduced by Jenifer et al., in 2016. This method overrides the manual setting of crisp clip limit [9]. In 2013, H.S. Jagannath et al. in their research article described the morphological enhancement of calcifications in digital mammogram images [10]. Effective removal of superfluous information of the tissue that surrounds the lesion was proposed by Cuiping Ding et al. in 2016. It improves the contrast based on nonlinear local enhancement [11]. One of the nonlinear filters called truncated Volterra filter to enhance the contrast and visualization of mass in digital mammograms was proposed by Vikrant Bhateja et al. in 2013 [12]. The shortcomings in the contrast enhancement of mammogram images can be successfully removed by incorporating optimization techniques.

3 Materials and Methods

3.1 Database Used

In order to benchmark a proposed algorithm, it is always a paradigm that it must use a standard test database (data set) in order to directly compare the results obtained. Since most of the mammographic databases hosted are not publicly available, the author used the most easily and most commonly used databases called the Mammographic Image Analysis Society (MIAS) database [13].

3.2 Contrast Limited Adaptive Histogram Equalization

CLAHE is one of type of Adaptive Histogram Equalization. It partitions the primary image into a few non-covering subimages. Histogram of subimages are cut with a constraint that measure the improvement of each and every pixel beyond it. The details in the image are shown up decisively in respect to the background. At the same time, the background of the image is improved similarly as the frontal area which prompts the high-quality output images [9].

4 Tree Seed Optimization

Tree seed optimization (TSA) algorithm manages three primary processes, such as arbitrary tree location, looking seeds, and determination of best arrangement. Arbitrary tree location refers to locating a tree inside the inquiry space, looking with seeds expounds the area look operation by a seed in view of the data from control parameter Search Tendency (ST) [14]. Choice of best arrangement is restrained in light of the target capacity of the issue. Finally, substitution process happens by the substitution of trees with seeds. The successive steps in TSA calculation are given as [15, 16]:

1. Exemplifying the quantity of population (N), control parameter(ST), termination condition X, and arrangement step x.
2. **Random trees location**
   With reference to the value of upper and lower bounds for a given inquiry space, create introductory area of trees arbitrarily utilizing equation (1)

$$T = L + rand(H - L) \tag{1}$$

   The irregular trees were additionally checked for fitness utilizing the objective function. A sample best answer for minimization process could be given as Eq. (2)

$$T_{best} = min\, f(T_i); i = 1, 2, 3, N \tag{2}$$

3. **Searching seeds**
   Searching seeds of a tree is defined utilizing a control parameter ST in the range [0 1]. In light of the rand value chosen in figuring of initial tree values, the seed for the trees can be computed utilizing one of the two rules as shown in Eqs. (3) and (4)

$$S_{((\forall rand < ST))} = T + (T_{best} - T_r) \tag{3}$$

$$S_{((\forall rand > ST))} = T + (T - T_r) \tag{4}$$

4. **Comparison of seeds with tree**
   The seed values are compared with the trees for better performance.
5. Test for termination If $x < X$, go to Step 3. Continue the seed searching procedure as the problem gets optimal solution.

5 Local Contrast-based CLAHE Tuned by Tree Seed Optimization

Histogram Equalization consistently creates over upgradation of the input image which prompts the loss of neighborhood pixels in the points of interest. AHE contrasts from the normal Histogram Equalization in the regard that HE produces just a single histogram while versatile strategy figures a few histogram, comparing to a particular segment of the image and makes use of it to reorganize the intensity of image. The transformation function can be articulated as given in Eq. (5)

$$T = \frac{E(Obtained from TSA).M}{\sigma} \tag{5}$$

where 'E' is the enhancement parameter which has to be identified by tree seed optimization algorithm TSA. A defines the global mean of the image, and σ defines the standard deviation obtained from user-defined mask (Local Standard Deviation). Using the transformation equation in Eq. (5), the output image can be obtained from Eq. (6)

$$B = T * (A - m) + m \tag{6}$$

where 'm' is the local mean obtained by the following equation (Eq. (7))

$$m(i, j) = \frac{1}{x * x} \sum_{i=0}^{x-1} \sum_{j=0}^{x-1} f(i, j) \tag{7}$$

The local standard deviation σ is given by Eq. (8)

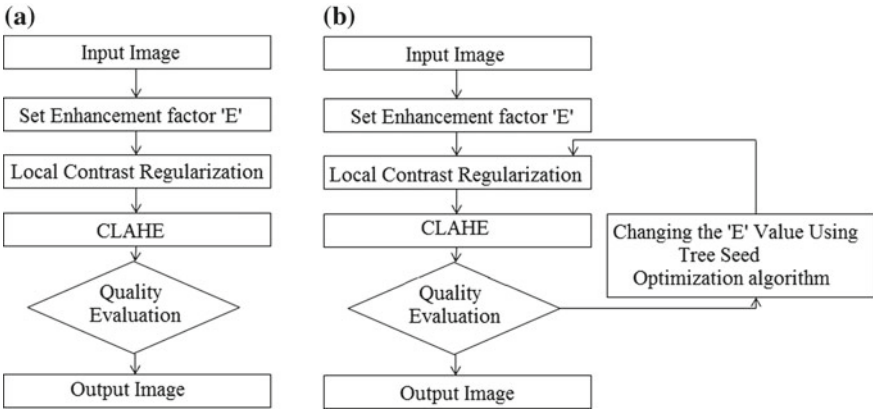


Fig. 1 **a** Conventional LC-CLAHE algorithm; **b** Tree seed optimization tuned LC-CLAHE

$$\sigma = \sqrt{\frac{1}{x * x} \sum_{i=0}^{x-1} \sum_{j=0}^{x-1} (f(i, j) - m(i, j))^2} \tag{8}$$

The resultant image $m(i, j)$ is delivered to CLAHE. CLAHE—a variation of AHE—lessens the noise enhancement. f(i, j) is the input image. Utilizing CLAHE likewise we have discovered that it is additionally not all that appropriate to images with fine subtle elements. In histogram-modified HM CLAHE, the author has proposed general adjustment of histogram alongside CLAHE. But in mammogram images, neighborhood subtle elements are more critical than universal points of interest for the location of malignant cells. So LC-CLAHE will utilize a nearby differentiation improvement to feature the fine subtle elements present and an upgrade value to control the limit of improvement alongside standard CLAHE [17]. So LC with CLAHE produces an ideal balance upgradation with all neighborhood data of mammogram image which may not be acquired utilizing CLAHE. Figure 1a depicts the steps associated with the proposed technique. As an initial step, the input image and the parameter 'E' are fed as input to Local Contrast Regularization. In LCR, the image will be converted in such a way to deliver the better points of interest covered up and that resultant image is given to CLAHE which will additionally augment the image with quality check. Figure 1 shows the incorporation of tree seed optimization algorithm in LC-CLAHE. The optimization capability of TSA was used in finding the optimal value of enhancement factor 'E'. Evaluation parameter used as fitness function and image quality indicator in this work is Contrast Improvement Index. The value of Contrast Improvement Index can be found using the following expression

$$CII = \frac{C_{Processed}}{C_{Origin}} \tag{9}$$

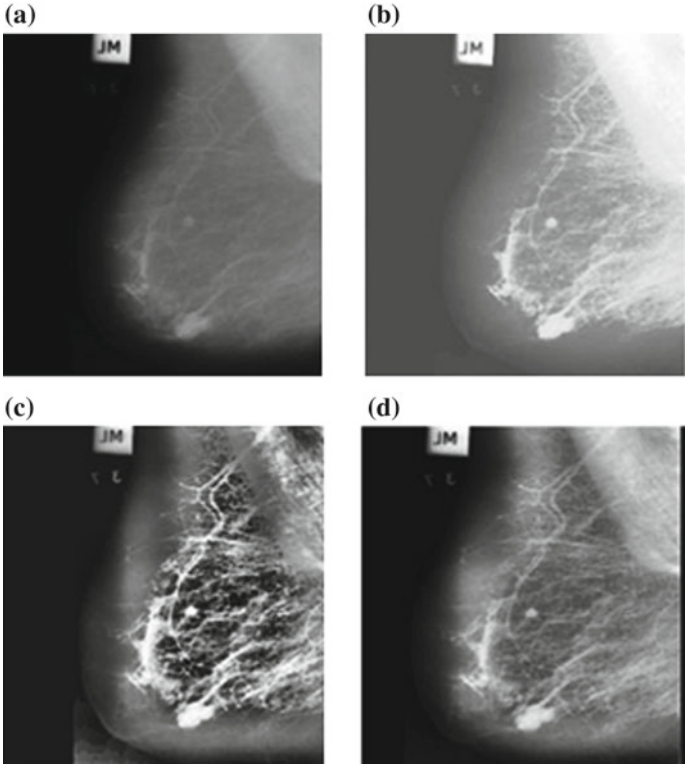


Fig. 2 **a** Input image mdb005; output image obtained using **b** CLAHE **c** conventional LC-CLAHE algorithm; **d** tree seed optimization tuned LC-CLAHE

6 Experimental Results

From the results, it is clear that LCM-CLAHE provides an optimal contrast enhancement for image. Figure 2a represents the fatty mammogram image mdb005. Various enhanced mammographic images obtained from AHE, CLAHE, conventional LC-CLAHE, tree seed optimization tuned LC-CLAHE of input image mdb005 are shown in Fig. 2b–d respectively. The Contrast Improvement Index rendered by various methods is shown in Table 1. From Table 1, it is obvious that the proposed TSA-based CLAHE offers highest values of Contrast Improvement Index.

Figure 3a represents the fatty glandular mammogram image mdb219, whereas the Fig. 3(b–d) represents various enhanced mammographic images obtained from AHE, CLAHE , conventional LC-CLAHE, tree seed optimization tuned LC-CLAHE, respectively.

Table 1 Comparison of Contrast Improvement Index for mammogram images

Images	Contrast Improvement Index			
	AHE	CLAHE	LC-CLAHE	TSA-based CLAHE
mdb005	0.82	2.52	3.92	4.50
mdb219	0.70	2.62	3.18	3.46
mdb211	0.68	1.82	1.97	2.53
mdb212	0.64	1.91	2.05	3.10
mdb315	0.47	2.01	2.54	5.50
mdb322	0.51	1.87	2.65	4.87

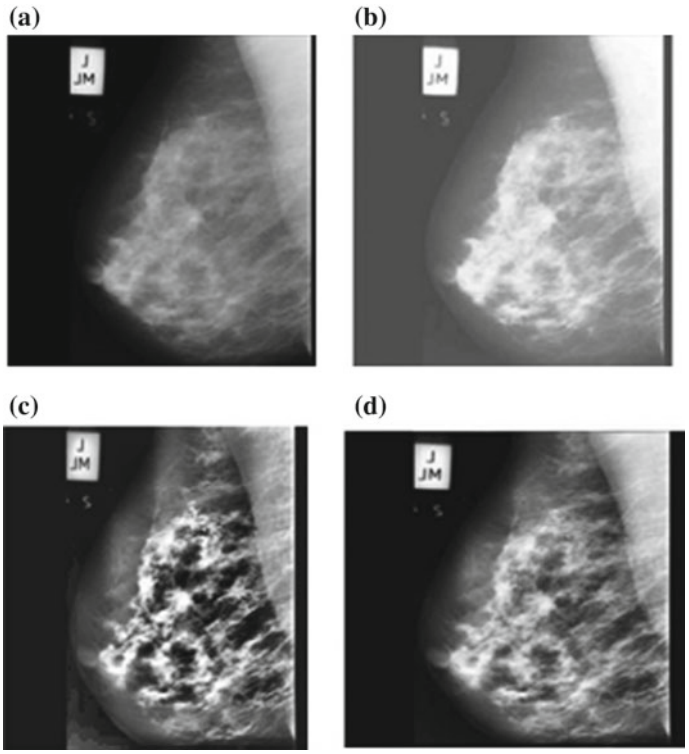


Fig. 3 **a** Input image mdb219; output image obtained using **b** CLAHE **c** conventional LC-CLAHE algorithm; **d** tree seed optimization tuned LC-CLAHE

7 Conclusion and Future Work

A novel approach for mammogram image improvement utilizing tree seed optimization tuned LC-CLAHE that provides neighborhood differentiation has been

introduced in this paper. The proposed strategy gives a superior differentiation and detailed protection of information in mammogram image. The mammogram images improved utilizing proposed technique are fulfilling both subjective and target assessment. The results' aftereffects of LCM-CLAHE demonstrate that this strategy is more powerful without trading off complexity and additionally unique data. A wide range of mammogram images like thick glandular can be enhanced successfully utilizing this technique. Future work is to improve the proposed technique for different sorts of medical images.

Acknowledgements The authors thank the Department of ECE, Kalasalingam University, for permitting to use the computational facilities available in Centre for Research in Signal Processing and VLSI Design which was setup with the support of the Department of Science and Technology (DST), New Delhi, under FIST Program in 2013 (Reference No: SR/FST/ETI-336/2013 dated November 2013).

References

1. Akila, K., Jayashree, L.S., Vasuki, A.: Mammographic image enhancement using indirect contrast enhancement techniques a comparative study. Procedia Comput. Sci. **47**, 255–261 (2015)
2. Muneeswaran, V., Pallikonda Rajasekaran, M.: Analysis of particle swarm optimization based 2D FIR filter for reduction of additive and multiplicative noise in images. In: Arumugam, S., Bagga, J., Beineke, L., Panda, B. (eds) Theoretical Computer Science and Discrete Mathematics. ICTCSDM 2016. Lecture Notes in Computer Science, vol. 10398. Springer, Cham (2017)
3. Muneeswaran V., Pallikonda Rajasekaran, M.: Beltrami-regularized denoising filter based on tree seed optimization algorithm: an ultrasound image application. In: Satapathy, S., Joshi, A. (eds) Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 1. Smart Innovation, Systems and Technologies, vol. 83. Springer, Cham (2018)
4. Muneeswaran V., Pallikonda Rajasekaran, M.: Gallbladder shape estimation using tree-seed optimization tuned radial basis function network for assessment of acute cholecystitis. In: Bhateja, V. et al. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol. 695. Springer, Cham (2018)
5. Muneeswaran, V., Pallikonda Rajasekaran, M.: Automatic Segmentation of gallbladder using bio-inspired algorithm based on a spider web construction model. J. Supercomput. (2018). https://doi.org/10.1007/s11227-017-2230-4
6. Srivatava, H., Mishra, A.K., Bhateja, V.: Non-linear quality evaluation index for mammograms. In: 2013 Students Conference on Engineering and Systems (SCES), pp. 1–5. Allahabad (2013). https://doi.org/10.1109/SCES.2013.6547534
7. Shanmugavadivu, P., Balasubramanian, K., Muruganandam, A.: Particle swarm optimized bi-histogram equalization for contrast enhancement and brightness preservation of images. Vis. Comput. **30**(4), 387399 (2014). https://doi.org/10.1007/s00371-013-0863-8
8. Langarizadeh, M., Mahmud, R., Ramli, A.R., Napis, S., Beikzadeh, M.R., Rahman, W.E.Z.W.A.: Improvement of digital mammogram images using histogram equalization, histogram stretching and median filter. J. Med. Eng. Tech. **35**(2), (2011)
9. Jenifer, S., Parasuraman, S., Kadirvelu, A.: Contrast enhancement and brightness preserving of digital mammograms using fuzzy clipped contrast-limited adaptive histogram equalization algorithm. Appl. Soft Comput. J. **42**, 167–177 (2016). https://doi.org/10.1016/j.asoc.2016.01.039

10. Jagannath, H.S., Kumar, V., Virmani, J.: Morphological enhancement of microcalcifications in digital mammograms. J. Inst. Eng. India Ser. B **93**(3), 163172 (2012). https://doi.org/10.1007/s40031-012-0020-1
11. Ding, C., Dong, M., Zhang, H., Ma, Y., Yan, Y., Zwiggelaar, R.: Nonlinear local transformation based mammographic image enhancement. In: Tingberg, A., Lng, K., Timberg, P. (eds.) Breast Imaging. IWDM 2016. Lecture Notes in Computer Science, vol. 9699. Springer, Cham (2016)
12. Bhateja, V., Urooj, S., Pandey, A., Ekuakille, A.L., Misra, M.: Improvement of masses detection in digital mammograms employing non-linear filtering. In: International Mutli-Conference on Automation Computing Communication Control Compressed Sensing (iMac4s), pp. 406–408. Kottayam, (2013). https://doi.org/10.1109/iMac4s.2013.6526445
13. Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., et al. (2015). Mammographic Image Analysis Society (MIAS) Database v1.21 [Dataset]. https://www.repository.cam.ac.uk/handle/1810/250394
14. Kiran, M.S.: TSA: tree-seed algorithm for continuous optimization. Expert Syst. Appl. **42**(19), 6686–6698 (2015)
15. Kiran, M.S.: An implementation of tree seed algorithm (TSA) for constrained optimization. In: Lavangnananda, K., Phon-Amnuaisuk, S., Engchuan, W., Chan, J.H. (eds.) IES 2015. PALO, vol. 5, pp. 189–197. Springer, Cham (2016)
16. Muneeswaran, V., Pallikonda Rajasekaran, M.: Performance evaluation of radial basis function networks based on tree seed algorithm. In: Proceeding of the 2016 International Conference of Circuit Power and Computing Technologies, pp.1–4. IEEE Explore (2016)
17. Mohan, S., Mahesh, T.R.: Particle swarm optimization based contrast limited enhancement for mammogram images. In: 2013 7th International Conference on Intelligent Systems and Control (ISCO), pp. 384–388. Coimbatore, India (2013). https://doi.org/10.1109/ISCO.2013.6481185

Design of Multi-layer Perceptron for the Diagnosis of Diabetes Mellitus Using Keras in Deep Learning

N. Muni Kumar and R. Manjula

Abstract Diabetes mellitus is one of the deadly diseases growing at a rapid rate in the developing countries. Diabetes has affected over 246 million people worldwide. According to the World Health Organization (WHO) report, by 2025 this number is expected to rise to over 380 million. If the diabetes mellitus is left untreated, it will result in blindness, amputation, kidney failures, heart attack, and stroke. So, the timely diagnosis of diabetes is a great boon to the people around the world. This study has taken the diabetes dataset and applied the deep learning method for the diagnosis of diabetes using Keras library. From our results, it is evident that, the increase in number of nodes in hidden layer of the Multi-layer Perceptron, resulted substantial increase in diagnosing accuracy.

1 Introduction

Diabetes mellitus [1] is one of the deadly diseases affecting over 400 million people worldwide and expected to rise to 640 million by 2040. It is a chronic disease that normally occurs when the pancreas is no longer able to produce insulin and the other scenario where the body cannot utilize the produced insulin. The first scenario in which pancreas does not produce insulin is called Type-I diabetes, and the second scenario is in which the pancreas does not produce enough insulin is called Type-II diabetes. According to The National Diabetes Statistics Report 2017, Estimates of Diabetes and its burden in the United States, it is estimated that 30.3 million people of all ages of the US population had diabetes in 2015, of which 7.2 million people were not aware that they had diabetes. The percentage of adults with diabetes is increasing with age. Diabetes was the seventh leading cause of death in the USA in 2015. The total direct and indirect estimated cost of diagnosed diabetes

N. Muni Kumar (✉) · R. Manjula
SCOPE, VIT University, Vellore, Tamil Nadu, India
e-mail: muni.kumarn2014@vit.ac.in

R. Manjula
e-mail: rmanjula@vit.ac.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_68

in the USA in 2012 was $245 billion, and average medical expenditure for people with diagnosed diabetes were about $13,700 per year. The average medical expenditures among people with diagnosed diabetes were about 2.3 times higher than expenditures for people without diabetes.

An estimated 33.9% of US adults aged 18 years or older had prediabetes in 2015, nearly half of adults aged 65 years or older had prediabetes. Also, the survey presented that, in 2011−2014, comparatively more men (36.6%) than women (29.3%) had prediabetes. The number of people with diabetes has nearly quadrupled since 1980. Prevalence is increasing worldwide, particularly in low- and middle-income countries. Diabetes of all types can lead to complications in many parts of the body and increase the risk of premature deaths.

2 Related Works

Emrane et al. [2] developed an expert clinical decision support system to predict diabetes mellitus. In this system, they used decision trees and k-nearest neighbor (KNN) algorithms and experimental results proved that C4.5 classification model provides better accuracy for diabetes disease prediction. The experiment was conducted on the PIMA Indians diabetes dataset. Anjali and Varun [3] attempted to develop a diabetes prediction method using support vector machine (SVM) on different datasets with more than 1 lakh instances. The experiment was carried on WEKA environment and achieved 72% accuracy for diabetes prediction. Aparimita et al. [4] proposed the prediction and classification of diabetes mellitus using artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS).

Veena and Anjali [5] carried out a machine learning approach for the prediction and diagnosis of diabetes mellitus. The results proved that the ada boost algorithm with decision stump as base classifier performs better compared to that of support vector machine, naïve Bayes, and decision trees. Ayush and Divya [6] designed a diabetes prediction system based on the personal daily lifestyle activities. This paper used classification and regression trees (CART) algorithm for the prediction of diabetes and obtained 75% accuracy in prediction. Durairaj and Kalaiselvi [7] proposed a backpropagation algorithm for the prediction of diabetes on the PIMA Indians diabetes dataset. The classification accuracy of backpropagation network was proved to be better than other methods.

Sapna and Pravin Kumar [8] proposed the diagnosis of diabetes disease from clinical big data using neural network. They compared quasi-Newton backpropagation and resilient backpropagation methods and proved resilient backpropagation performed better. Zhilbert et al. [9] developed an intelligent system for diabetes prediction. They focused on the joint implementation of the support vector machine (SVM) and naïve Bayes statistical modeling and proved the joint implementation performs better compared to the individual implementations.

Olaniyi and Adnan [10] proposed a backpropagation network to predict the diabetes mellitus given the various input attributes to the network. Sigmoid transfer function was used in the hidden layers and the output layers. They proved artificial

neural network trained with backpropagation gave 82% accuracy. Linli [11] proposed a weight-adjusted voting approach for the diagnosis of diabetes. In this approach, the author combined three classifiers: support vector machines, artificial neural network, and naïve Bayes, to diagnose diabetes. Here, the PIMA Indian diabetes dataset was used and achieved an accuracy of 80.0%.

Jetri and Suharjito [12] implemented a method of neural network called extreme learning machine (ELM) and proved the accuracy of ELM is better than the accuracy of backpropagation in the diagnosis of diabetes mellitus.

Deepthi Jain and Divakar Singh [13] presented a feedforward neural network approach for the estimation of diabetes risk factors based on the body characteristics. They used neural network toolbox of MATLAB on the dataset with 50,788 records provided by the SAS data mining shootout competition.

Muhammad et al. [14] presented a study on the prediction of diabetes using different supervised learning algorithms on the diabetes patients between 25 and 78 years old, and their prediction accuracies were evaluated.

3 Methodology

3.1 Artificial Neural Network (ANN)

Artificial neural networks (ANNs) [15] is one of the soft computing techniques that simulate the behavior of human brain and has become popular in its applications in the fields of data mining, fault detection, image processing, pattern recognition, weather forecasting, job scheduling, medical diagnosis, etc. These networks are capable of learning through examples (train data), remember their past experiences, and perform the parallel processing. The learning activities in these types of networks are possible because of the neurons that could receive and process the information as similar to that of human brain.

A very basic neural network with two layers i.e., input layer and processing or output layer, is depicted in Fig. 1. The input layer receives input from the sensors and passes the input to the next layer for processing. The processing layer consists of summation and activation functions where the activation function is checked with the threshold value and generates the output signal of the neural network.

3.2 Multi-Layer Perceptron (MLP)

A multi-layer perceptron is a feedforward network with at least three layers: input layer, hidden layer, and output layer, as depicted in Fig. 2. In MLP, except the input layer, other layers use nonlinear activation functions. MLP uses a supervised learning called backpropagation for training. The MLP consists of minimum of three layers with nonlinear activation functions, which makes this a deep neural

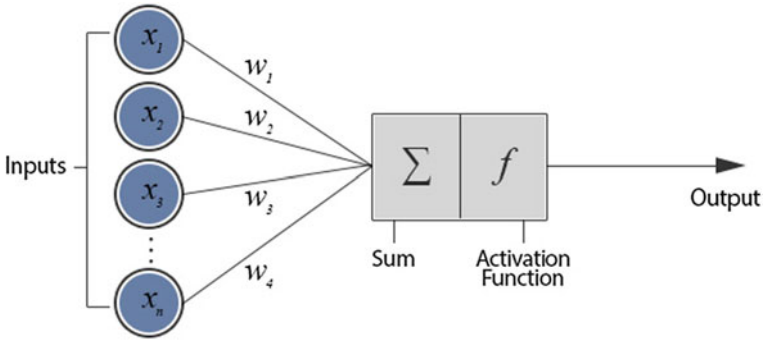


Fig. 1 A simple neural network with two layers. This shows a figure consisting of simple neural network. The input layer with n-neurons, activation functions, and an output layer

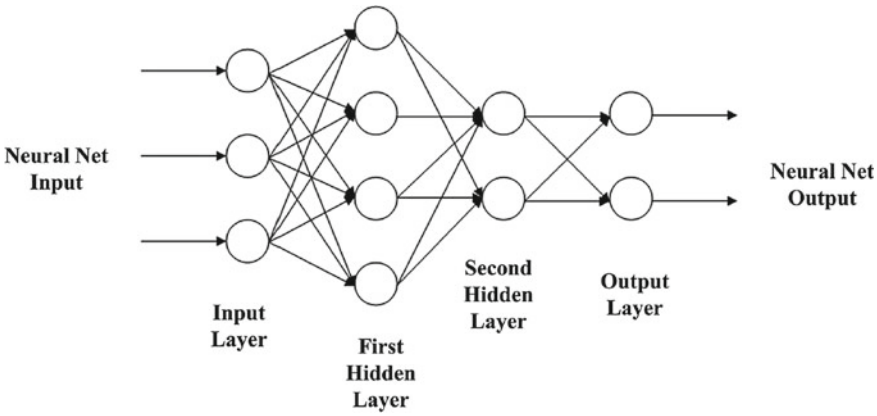


Fig. 2 A multi-layer perceptron. This shows a figure with one input layer, two hidden layers, and an output layer

network (DNN). Every node in the MLP contains certain weights w_{ij}, which are connected to the nodes in the following layers. MLP learns in its learning phase with the change of weights after the processing of every node. Based on the output of the unit compared to the expected result, the error is calculated and then the weights get updated which is carried out through backpropagation method.

3.3 Proposed Model

In our proposed model, we installed the Theano and Tensorflow libraries as pre-requisite for using the Keras library. Keras allows for easier development of neural networks that can run on both CPUs and GPUs for faster computations and parallel

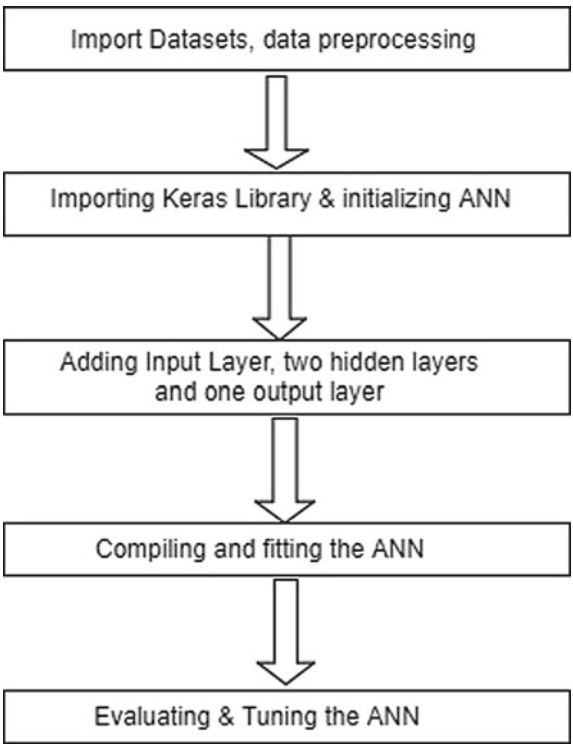


Fig. 3 Proposed model for the design of MLP. This shows a figure with the proposed model starting with the dataset imports, import libraries, model, compilation, and evaluation of ANN

processing seamlessly. The major task in the design of deep neural networks is organizing the layers. Using Keras, it is easier to build any complex neural network with minimal number of steps.

Keras allows for easier development of neural networks that can run both on CPUs and GPUs for faster computations and parallel processing seamlessly. Using Keras it is easier to build any complex neural network with minimal number of steps as model is the core data structure in Keras which is used for organizing layers.

The proposed model shown in Fig. 3 lists all the steps for the design and construction of a multi-layer perceptron (MLP) network. Initial steps to begin the model is the installation of required packages and carrying out the data preprocessing steps like encoding the categorical data, independent variables and splitting the data into training set and test set. After the completion of preprocessing, the ANN must be initialized and the required number of layers should be added to build the model. While building layers, suitable activation function should be supplied along with number of nodes in each layer. Among the available activation

functions, ReLU activation function is preferred for the hidden layers and sigmoid for the output layer. Our proposed model also uses ReLU and sigmoid activation functions.

3.4 Pseudocode

The pseudocode for the proposed design of multi-layer perceptron (MLP) network for the diagnosis of diabetes using PIMA Indian diabetes dataset is as follows.

1. Import the libraries for data preprocessing and the PIMA Indian diabetes dataset downloaded from the UCI repository.
2. Preprocess the input data like encoding categorical data, independent variables, splitting the dataset into training set and test set, and perform feature scaling if required.
3. Import the Theano, Tensorflow, Keras library, and required Python packages for setting up the Python environment.
4. Initialize the ANN.
5. Add the input layer and the first hidden layer.
6. Add the extra hidden layers, apply the activation functions.
7. Add the output layer along with activation function.
8. Compile the ANN.
9. Fit the ANN to the training set.
10. Predict the test results.
11. Evaluate the ANN.
12. Improve and tune the ANN.

4 Experiment

The proposed work is implemented using the Python environment on Windows 10 operating system with Intel Core i5 processor and 8 GB of RAM. The experiment requires some initial prerequisites for setting up the environment which will be discussed as environmental setup in the following section. For the experiment, we have chosen PIMA Indian diabetes dataset from UCI data repository. The experiment was conducted with the Python source code for the above-given pseudocode.

4.1 Environmental Setup

Our proposed model requires any operating system with Python 2.7 installed. In Python environment, the libraries can be installed with the following commands.

```
$ pip theano install        # installation of theano
$ pip install tensorflow    # installation of tensorflow
$ pip install keras         # installation of keras
```

Before using the above commands, the pip must be installed on Python system. In our experimental setup, we have installed Anaconda environment for the execution of Python programs in Spider editor. Anaconda is an open-source distribution of Python language for large-scale data processing, predictive analytics, and scientific computing, which simplifies package management and deployment.

4.2 Dataset

The dataset used in this experiment was downloaded from the University of California, Irvine (UCI) repository of machine learning databases. The name of the dataset is PIMA Indians diabetes dataset, which consists of 768 instances of 9 attributes belonging to the Indian people living in Arizona, USA.

5 Results and Discussion

The classification accuracies obtained by this research work on the PIMA Indian diabetes dataset for different multi-layer perceptrons (MLPs) were presented in the following table. The experimental results are tabulated for three perceptrons as follows:

The notation 8-12-8-1 is used for the multi-layer perceptron indicating 8 nodes in the input layer, 12 and 8 nodes in the hidden layers, and 1 node in the output layer. Similarly, three more MLP networks are designed with the increase in nodes in the hidden layers as 8-32-32-1, 8-64-64-1, and 8-128-128-1. We have used the ReLU activation function in the input layer and the hidden layers, and sigmoid activation function in the output layer. The following table lists the results obtained by executing each network for 10 runs with 150 epochs in each model.

The results are collected in ten runs for four different MLP networks for 150 epochs each network and compared in Table 1. A graph was drawn with the results obtained in Table 1 as shown in Fig. 4. From the above results, it is evident that, the increase in number of nodes in the hidden layer resulted in substantial increase in diagnosing accuracy.

Table 1 Comparison of the performances of four different MLP networks

	Performance of four different multi-layer perceptrons			
S. No.	8-12-8-1 (%)	8-32-32-1 (%)	8-64-64-1 (%)	8-128-128-1 (%)
1	78.91	82.68	83.33	86.46
2	79.56	84.11	82.81	86.33
3	77.99	82.16	83.72	87.11
4	78.78	82.81	84.77	87.89
5	79.43	83.20	84.51	86.72
6	79.69	82.68	85.29	84.77
7	78.26	83.98	84.64	85.29
8	79.17	83.72	85.68	88.15
9	78.26	82.16	84.38	87.24
10	80.47	83.20	83.46	86.42
Average	79.05	83.02	84.26	86.67

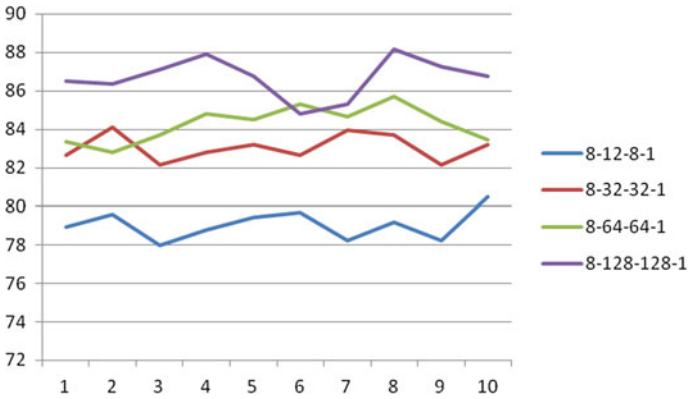


Fig. 4 Graph with classification accuracies of four different MLP networks. This figure shows the comparisons of the MLP networks drawn in separate color for each MLP

6 Conclusion

This paper presented a new methodology for the design and development of a multi-layer perceptron network using Keras. We have used PIMA Indian diabetes dataset for the diagnosis of diabetes mellitus. It was observed that the designed model achieved greater accuracy by designing deep neural networks with the increase in the number of hidden layers and nodes per layer. This work can be further extended to obtain an accuracy of more than 90% with deep neural networks so that diabetes can be accurately diagnosed during the early stages.

Acknowledgements I would like to take this opportunity to express my profound gratitude and deep regard to my guide, Prof. R. Manjula, SCOPE, VIT University, for her exemplary guidance, valuable feedback, and constant encouragement in completing this paper. Her valuable suggestions were of immense help in getting this work done. Also, I would like to extend my sincere gratitude to my parents, my wife, and my lovable daughter for their constant support and encouragement in completing this paper.

References

1. International Diabetes Federation. https://www.idf.org/about-diabetes/what-is-diabetes (2017)
2. Hashi, E.K., Zaman, M.S.U., Hasan, M.R.: An expert clinical decision support system to predict disease using classification techniques. In: International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 16–18. Bangladesh (2017)
3. Negi, A., Jaiswal, V.: A first attempt to develop a diabetes prediction method based on different global datasets. In: Fourth IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC), pp. 237–240 (2016)
4. Swain, A., Mohanty, S.N., Das, A.C.: Comparative risk analysis on prediction of diabetes mellitus using machine learning approach. In: IEEE International Conference on Electrical, Electronics and Optimization Techniques (ICEEOT), pp. 3312–3317 (2016)
5. Vijayan, V.V., Anjali, C.: Prediction and diagnosis of diabetes mellitus – A machine learning appraoch. In: IEEE Recent Advances in Intelligent Computational Systems (RAICS), Trivandrum (2015)
6. Anand, A., Shakti, D.: Prediction of diabetes based on personal lifestyle indicators. In: IEEE 1st International Conference on Next Generation Computing Technologies (NGCT), pp. 673–676. Dehradun (2015)
7. Durairaj, M., Kalaiselvi, G.: Prediction of diabetes using back propagation algorithm. Int. J. Emerg. Technol. Innov. Eng. **1**(8), 21–25 (2015)
8. Sapna, S., Pravin, Kumar M.: Diagnosis of disease from clinical big data using neural network. Indian J. Sci. Technol. **8**(24), 1–7 (2015)
9. Tafa, Z., Pervetical, N., Karahoda, B.: An intelligent system for diabetes prediction. In: Proceedings of the 4th Mediterranean Conference on Embedded Computing, pp. 378–382 (2015)
10. Olaniyi, E.O., Adnam, K.: Onset diabetes diagnosis using artificial neural network. Int. J. Sci. Eng. Res. 754–759 (2014)
11. Li, L.: Diagnosis of diabetes using a weight-adjusted voting approach. In: IEEE 14th International Conference on Bioinformatics and Bioengineering, pp. 320–324 (2014)
12. Pangaribuan, J.J., Suharjito.: Diagnosis of diabetes mellitus using extreme learning machine. In: IEEE International Conference on Information Technology Systems and Innovation (ICITSI), pp. 33–38 (2014)
13. Jain, D., Singh, D.: A neural network based approach for the diabetes risk estimation. Int. J. Comput. Appl. 1–4 (2013)
14. Sapon, M.A., Ismail, K., Zainudin, S.: Prediction of diabetes by using artificial neural network. In: International Conference on Circuits, Systems and Simulations (IPCSIT), pp. 299–303. Singapore (2011)
15. Tripathy, B.K., Anuradha, J.: Soft Computing – Advances and Applications: Cengage Learning (2015)

A QoS-Enhanced Smart Packet Scheduler for Multi-core Processors in Intelligent Routers Using Machine Learning

Suman Paul and Malay Kumar Pandit

Abstract In this paper, we have developed a QoS-enhanced smart scheduler for multi-core processor in intelligent IP routers using machine (deep) learning. The proposed packet scheduler is stochastic in nature, it can process real-time traffic packets, and it is reconfigurable one. Each core maintains a guaranteed steady-state core load distribution ratio with an aim of keeping the processor utilization close to 100%. The scheduler has the advantages of having each core processing the incoming traffic in a utilization-driven deadline-aware mode, and the load imbalance is minimized dynamically at run-time in an intelligent way using our proposed machine (deep) learning algorithm in order to retain a steady-state load distribution among the cores. This results in minimization of computational overhead to each core, higher throughput, lower value of mean waiting time, and PLR.

1 Introduction

Quality of service (QoS) plays a significant role for performance issues in communication networks as well as network connecting devices, like core and edge routers, switches. A simulation framework for the first time has been for implementing our new idea of scheduler for multi-core intelligent IP routers. This work is an extension and further analysis of our ongoing research [1]. In this paper, we present a detailed analysis of a QoS-enhanced stochastic real-time packet scheduler in a multi-core processor in a router considering multimedia IP traffic for load balancing among the cores. In this case, we have taken three classes—VoIP, IPTV, and HTTP. The proposed multi-core scheduler has following advantages. Each core

S. Paul · M. K. Pandit
Department of ECE, Haldia Institute of Technology, Haldia, West Bengal, India
e-mail: mkpandit.seci@gmail.com

S. Paul (✉)
School of Engineering and Technology, West Bengal University of Technology (Maulana Abul Kalam Azad University of Technology West Bengal), Kolkata, India
e-mail: paulsuman999@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_69

is aware of class-dependent deadlines of real time traffic, VoIP, IPTV, and best effort traffic HTTP. The scheduler is utilization-driven, maintains a guaranteed steady-state process utilization ratio in the order 0.80:0.16:0.04 for above-mentioned multimedia traffic flows, and enforces the processor utilization for each core to meet close to 100%. The load imbalance is minimized in an intelligent way at run-time using our newly proposed machine (deep) learning algorithm in order to retain a steady-state core load distribution ratio as time passes. This results the benefit of parallel processing which outcomes higher throughput, highest system utilization, lower value of mean waiting time, and packet loss rate (PLR).

Various network QoS parameters are throughput, fair bandwidth allocation, packet loss rate (PLR), latencies, jitter, etc. In order to optimize the system performance in a multi-core processor in a router, the incoming packets are processed concurrently by parallel computation in each core. Moreover, an important issue is that the load in each core of the system should be balanced and as well as highly utilized. An efficient load balancing scheme dynamically allocates packets to cores, monitors load of each core, and decides at run-time to balance the load from the overloaded to underutilized cores.

2 Literature Review

Allocation of task, scheduling the different types of traffic, both real-time and best effort Internet traffic and the load distribution, and maintaining a steady state of balanced load among the cores in a multi-core processor within a router are emerging fields of research and study. Among various distinguished studies, the authors in [2, 3] have proposed various methods for multi-core systems, but these are not applicable for real-time tasks such as deadline-sensitive multimedia IP traffic. In [4, 5], the authors have proposed schedulers for real-time multi-core system. These schedulers are not utilization-driven, and it has no advantage of reconfigurability. In [6], the authors have proposed a real-time multi-core scheduler considering static and dynamic power as well as load balancing issues. But the scheduler is not a reconfigurable one.

3 Proposed Multi-core Scheduling Framework in IP Router with Load Balancing

The working flow model of the proposed scheduler with load balancing scheme in a multi-core router based on our proposed machine (deep) learning algorithm is illustrated in Fig. 1. The model is defined by two transition probability matrices (TPM), denoted by Eqs. (1) and (2). The proposed model is a finite-state machine (FSM) of a $m \times n$ matrix, and the working principle is based on Markov chain. The

elements in the C_{TPM} matrix in Eq. (1) denote a core's probability of transition of load from the mth to nth core. The core load distribution (ratio) vector is defined by matrix (2). The PLR (error) probability matrix of the cores E_c is stated in (3). As stated in QUEST scheduler [1], each core is represented by an FSM stated in (4). The elements in the FSM of each core are estimated using *machine learning Metropolis—Hastings algorithm* in order to attain a target steady-state process utilization ratio for three multimedia IP traffic—VoIP, IPTV, and HTTP. This is in the order of 0.8:0.16:0.04. An initial error probability matrix (in Eq. 5) denotes the root-mean-squared error of cache miss (L1 and L2) and deadline miss errors considering practical values for each core.

$$C_{TPM} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & .. & c_{1n} \\ c_{21} & c_{22} & c_{23} & .. & c_{2n} \\ : & : & : & : & : \\ c_{m1} & c_{m2} & c_{m3} & .. & c_{mn} \end{pmatrix} \tag{1}$$

$$\mathrm{C} = (\mathrm{C}_1 \quad \mathrm{C}_2 \quad .... \quad \mathrm{C}_\mathrm{N}) \tag{2}$$

$$E_C = \begin{pmatrix} 1-e_1 & 1-e_2 & 1-e_3 & 1-e_4 \\ e_1 & e_2 & e_3 & e_4 \end{pmatrix} \tag{3}$$

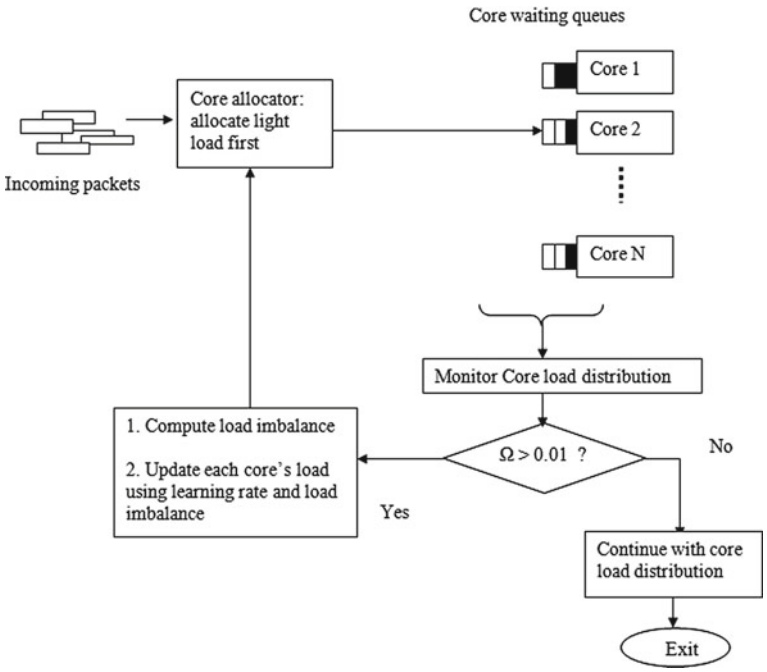


Fig. 1 Proposed scheduling framework for multi-core router with dynamic load balancing using deep learning

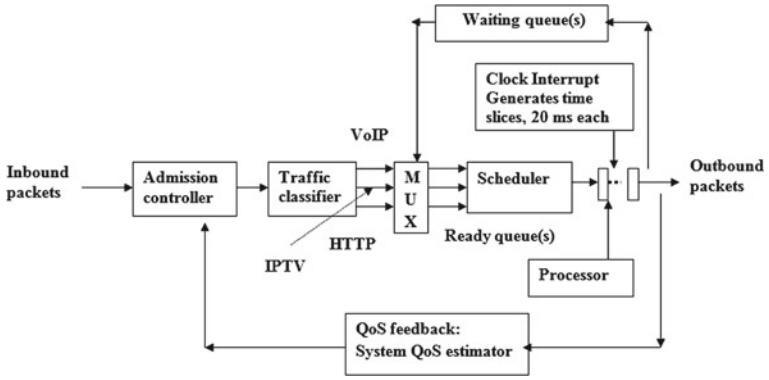


Fig. 2 Working principle of each core running M/BP/1/./QUEST scheduling framework

$$T = \begin{pmatrix} 0.90 & 0.08 & 0.02 \\ 0.39 & 0.56 & 0.05 \\ 0.42 & 0.18 & 0.40 \end{pmatrix} \tag{4}$$

$$E = \begin{pmatrix} 0.98 & 0.9 & 0.8 \\ 0.02 & 0.1 & 0.2 \end{pmatrix} \tag{5}$$

The simplified diagram of service model of each core is defined as M/BP/1/./QUEST [1] which is depicted in Fig. 2. As shown in Fig. 1, a run-time core allocation map table keeps records of packets to an allocated core and monitors the core load distribution ratio. Incoming packets are allocated to a core based on principle of run-time '*light loaded first*'. A load imbalance detector monitors length of waiting queue in each core applying the principle of '*random early detection*' [7] which is used in most of the modern routers. The load imbalance among the cores is updated by applying our proposed algorithm in Sect. 4.2.

4 The Multi-core Scheduling Algorithm with Core Load Balancing

4.1 Review of Our Proposed QUEST Scheduler for a Single Core

In order to understand the proposed scheduler for multi-core platform with load balancing, we *first review our earlier work for a single processor* [1] using an initial TPM in (4) and an error probability matrix in (5).

```
1.Generate random number R, 0.01≤R≤1;
2. Set: Time quantum T_Q: 20 ms;
3. Initialize: timer, t=0
4.      switch (initial_process) {
5.          CASE  initial_process:P1 //Pi denotes class specific process, P1, P2 and P3
                                       denotes VoIP, IPTV and HTTP, respectively//
6.                  if (0.01≤R≤ p11) then        //pij denotes probability of a process to be
                                       their own state or to make transition to a new
                                       state//
7.                      execute P1;
8.                  else if ((p11+0.01)≤R≤(p11+0.08)) then
9.                      execute P2;
10.                 else execute P3;
11.                 end if;
12.         CASE  initial_process:P2
13.                 if (0.01≤R≤p22) then
14.                     execute P2;
15.                 else if ((p22+0.01)≤R≤(p22+0.39)) then
16.                     execute P1;
17.                 else execute P3;
18.                 end if;
19.          CASE  initial_process:P3
20.                 if (0.01≤R≤0.4) then
21.                     execute P3;
22.                 else if ((p33+0.01)≤R≤(p33+0.42)) then
23.                     execute P1;
24.                 else execute P2;
25.                 end if;    }
26.    end for;
27.  Place Pi in expired queue;
```

4.2 Proposed Multi-core Scheduler with Load Balancing Using Machine (Deep) Learning

In this subsection, we discuss our newly proposed intelligent scheduler with load balancing for multi-core processor based on machine (deep) learning.

```
1. Define: N;                                          //N, no. of cores //
2. For Core Ci,  i=1to N                    // i stands for ith core //
3. Initialize: Elements cij of core allocation  TPM, CTPM ;
4. Initialize: Core load distribution ratio (vector)  including load
        imbalance pattern observed, L=[L1 : L2 .... LN];   // Li, individual core load//
5. Initialize: PLR probability matrix, Ec ;          //PLR: Packet loss rate//
6.      epoch=0;
7.        For i,j= 1 to N
8.           While Ωi >0.01                          // Ω is load imbalance //
9.              Compute  Ω, Ωi =L(k+1)- L(k);              // k states kth iteration //
10.             Compute ΔLij, ΔLij =η. Ωi. Lij ;   // Weight (load) update , η: Learning
                                                        rate //
11.              epoch=epoch+1;              //loop convergence//
12.           End;
13.       End;                                  // For loop for i,j ends //
14.   End;
```

5 Simulation Results

In this work, we apply simulation platform using MATLAB and DEVS suite [8], a discrete event simulation tool. For simplicity, we consider four identical cores. We take an initial TPM of individual core's unbiased probability of transition of load from one core to another stated in (6). A packet loss rate (error probability) matrix of the cores E_c is stated (7). Considering random arrival of processes for simulation, we take an initial *imbalanced run-time core load distribution* stated in (8). Initially, the $core_1$ and $core_2$ are underutilized, whereas other cores are heavily loaded. The following simulation parameters were used which are stated in Table 1.

$$C_{TPM} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix} \tag{6}$$

Table 1 Simulation parameter

Parameters	Conditions
Arrival rate:	200 packets/s
Link capacity:	10 Mbps
Load imbalance, Ω:	0.01
Service discipline in each core:	QUEST
Simulation time:	397 s

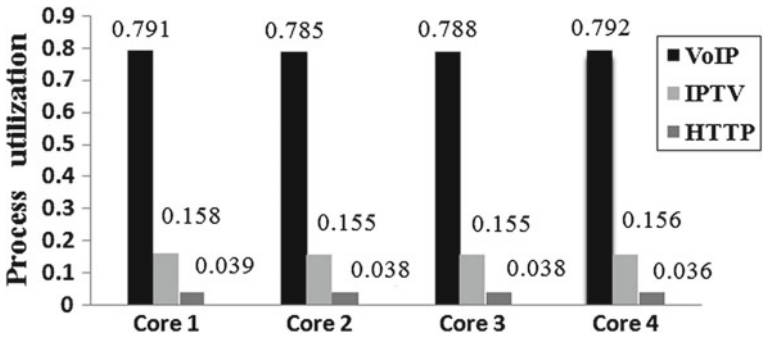


Fig. 3 Steady-state process utilization ratio of VoIP, IPTV, and HTTP

$$E_C = \begin{pmatrix} 0.099 & 0.098 & 0.097 & 0.096 \\ 0.001 & 0.002 & 0.003 & 0.004 \end{pmatrix} \tag{7}$$

$$C_L = (\,0.23 \quad 0.24 \quad 0.26 \quad 0.27\,) \tag{8}$$

Moreover, this is to be noted that learning rate parameter η has been taken in the order of ($0.05 \leq \eta \leq 0.25$) settling to a solution. We have observed that a value of 0.13 gives the best result. Accordingly, while running each core (represented by an FSM), a steady-state guaranteed process utilization ratio is achieved in the order of 0.80:0.16:0.04 for VoIP, IPTV, and HTTP, respectively. The process utilization ratio for each core is illustrated in Fig. 3.

6 Conclusion and Future Work

In this paper, we have proposed and performed the analysis of a QoS-enhanced stochastic real-time packet scheduler for multimedia IP traffic in a multi-core processor in a router based on machine (deep) learning. The scheduler is unique in that each core processes the incoming packets in a deadline-aware and utilization-driven mode. A machine (deep) learning algorithm has been proposed in order to minimize the load imbalance among the cores and to maintain a steady-state core load distribution. The performance of the scheduler can be further investigated by increasing the number of cores. An extension of analyzing the impact of scheduling jitter and the scheduler's response to noise are ongoing.

References

1. Paul, S., Pandit, M. K.: A QoS-enhanced intelligent stochastic real-time packet scheduler for multimedia IP traffic. Multimed. Tools Appl. 1–24 (2017). https://doi.org/10.1007/s11042-017-4912-6
2. Kavousianos, X., Chakrabarty, K., Jain, A., Parekhji, R.: Test schedule optimization for multicore SoCs: handling dynamic voltage scaling and multiple voltage islands. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **31**(11), 1754–1766 (2012)
3. Geng, X., Xu, G., Wang, D. Shi, Y.: A task scheduling algorithm based on multi-core processors. In: Proceedings of the International Conference on Mechatronic Science, Electric Engineering and Computer (MEC'11), pp. 942–945. Jilin, China (2011)
4. Han, J.-J., Wu, X., Zhu, D., Jin, H., Yang, L., Gaudiot, J.L.: Synchronization-aware energy management for VFI-based multicore real-time systems. IEEE Trans. Comput. **61**(12), 1682–11696 (2012)
5. Seo, E., Jeong, J., Park, S., Lee, J.: Energy efficient scheduling of real-time tasks on multicore processors. IEEE Trans. Parallel Distrib. Syst. **19**(11), 1540–1552 (2008)
6. Cho, K.-M., Tsai, C.-W., Chiu, Y.-S., Yang., C.-S.:A high performance load balancing strategy for real-time multicore systems. Sci. World J. Hindawi, **2014**, 1–14 (2014). (article ID: 101529)
7. QoS: Congestion Avoidance Configuration Guide, Cisco IOS XE Release 3S, pp. 1–6. https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/qos_conavd/configuration/xe-3s/qos-conavd-xe-3s-book/qos-conavd-oview.pdf
8. DEVS suite Discrete event system simulator suite, Arizona Center of Integrative Modeling and Simulation of Arizona State University.http://acims.asu.edu/software/devs-suite

Achieving Robustness of Mesh Watermarking Techniques Toward Mesh Geometry and Topology-Invariant Operations

Sagarika Borah and Bhogeswar Borah

Abstract Eminent usage of three-dimensional (3D) media on the Internet nowadays brings the concern related to their security. Watermarking is a branch of security mechanism that protects these 3D models from malicious users, and it also detects and localizes the attacks. While talking about attacks to 3D media, there are a number of mesh processing operations which do not alter the actual geometry or topology of the mesh. These operations are not considered as attack, and the mesh watermarking process should be able to resist such operations. These operations include *mesh translation*, *mesh scaling*, *mesh rotation*, and *vertex reordering*. This paper discusses different approaches to make the mesh watermarking operations invariant to the above routine operations. The approaches have also been implemented to 3D meshes.

1 Introduction

Daily escalation of Internet technology usage and higher dimensional media acquisition techniques has lead to raise in concern related to the security of these media. Malicious alteration of these 3D models gives rise to information loss, getting manipulated, and finally falsely interpreted by authorized persons. So to handle such situation watermarking techniques are used. Watermarking is the process of inserting some information to the host media. The information embedded to the host media may be related to not related to it. 3D surface models are modeled by various mathematical representation formats like parametric surfaces, implicit surfaces, point clouds, voxels, polygon mesh; among which polygon meshes are mostly used due to their simpler structure and are easier to work with [1]. A 3D

S. Borah (✉) · B. Borah
Department of Computer Science and Engineering, Tezpur University, Napaam, India
e-mail: sagarika08connect@gmail.com

B. Borah
e-mail: bgb@tezu.ernet.in

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1_70

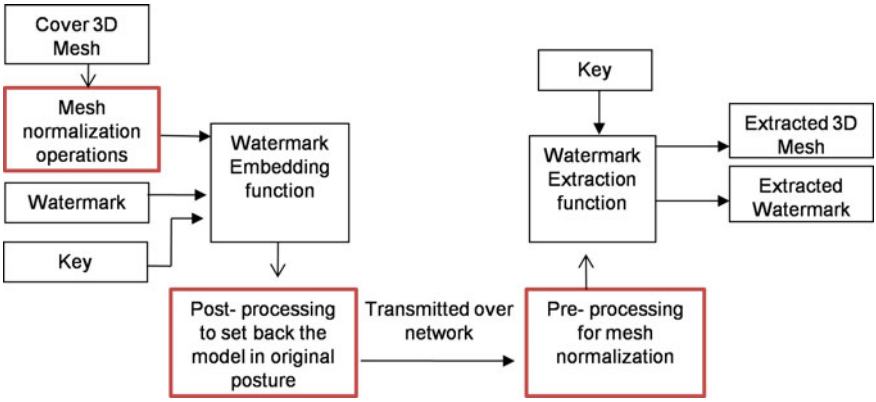


Fig. 1 A typical 3D mesh watermarking flow diagram

triangular mesh M defines a tuple $\{V, E, F\}$ of vertices $V = \{v_i | v_i \in \mathbb{R}^3, 1 \leq i \leq m\}$, edges $E = \{e_{ij} = (v_i, v_j) | v_i, v_j \in V, i \neq j\}$ and faces F, which are usually triangles $F = \{f_{ijk} = (v_i, v_j, v_k) | v_i, v_j, v_k \in V, i \neq j, j \neq k, k \neq i\}$ where m is the number of vertices. *Mesh geometry* specifies positional, and other geometric properties and *mesh topology* shows the incidence relationships between the mesh elements. The nature of watermark varies with the application domain. The various application domains are copyright protection, authentication, broadcast monitoring, and traitor tracing. The different watermarking scheme requirements based on the security applications are as follows. **Robust watermarking requirements** includes robustness of the watermark to any sort of malicious or non-malicious attacks. This class of watermarking also wants a watermark to be imperceptible to human eye. **Fragile watermarking requirements** include higher data capacity for use in diverse application and at the same time should be imperceptible to human vision. Apart from that a fragile watermark is fragile to any modification in the mesh. However, a subsection of fragile watermarking is called *semi-fragile watermarking* where a watermark needs to resist a group of content preserving mesh operations. Figure 1 shows a typical flow diagram of a mesh watermarking scheme and it also shows the locations where the normalization steps could be placed.

2 Affine Transformations on Mesh and Vertex Reordering

Affine transformations on meshes include mesh translation, mesh scaling, and mesh rotation (they are together also termed as RST operations). However, there are other higher versions of affine transformations like shearing, projective mapping but these are not considered to be non-malicious, and they impact on the mesh geometry as well.

Mesh translation: A mesh translation is simply a repositioning operation on the 3D mesh along a straight-line path from one coordinate location to another in the space. To translate a 3D mesh, a translation factor $\{f_x, f_y, f_z\}$ is added to the three mesh coordinate values as shown in Eq. 1.

$$\begin{aligned} x' &= x + f_x \\ y' &= y + f_y \\ z' &= z + f_z \end{aligned} \tag{1}$$

Mesh scaling: Scaling operation basically changes the size of the 3D model which includes expanding or squeezing with a scaling factor. Scaling can be of two types: uniform scaling and non-uniform scaling. However, non-uniform scaling is considered malicious and it modifies the mesh geometry. So to make the mesh robust against uniform scaling a normalization of the vertex information is carried out as shown in Eq. 3. Here the scaling factor α is computed in different ways. Some of the works on mesh watermarking [7] computed scaling factor as the distance of two furthest points on the model from Eq. 2.

$$\alpha = max\{(x_i)^2 + (y_i)^2 + (z_i)^2\} - min\{(x_i)^2 + (y_i)^2 + (z_i)^2\} i = 1, 2, \ldots N_v \tag{2}$$

$$\begin{aligned} x' &= \alpha * x \\ y' &= \alpha * y \\ z' &= \alpha * z \end{aligned} \tag{3}$$

Mesh rotation: Rotation operation in 3D space is complicated than in 2D images. During 3D mesh rotation, an axis of rotation needs to be defined. In 2D, the axis of rotation is always perpendicular to the xy plane, i.e., the z-axis, but in 3D the axis of rotation can have any spatial orientation as shown in Fig. 2. Rotation operation to a 3D model is performed by shifting the mesh elements positions in an angular path in a particular direction. This direction actually is determined from the axis of rotation and the reference plane(xy plane, yz plane, zx plane) at which the repositioning of mesh elements happens. Equations 4, 5, and 6 show the rotation operations on 3D mesh elements.

$$\begin{aligned} x' &= x * \cos q - y * \sin q \\ y' &= x * \sin q + y * \cos q \\ z' &= z \end{aligned} \tag{4}$$

$$\begin{aligned} y' &= y * \cos q - z * \sin q \\ z' &= y * \sin q + z * \cos q \\ x' &= x \end{aligned} \tag{5}$$

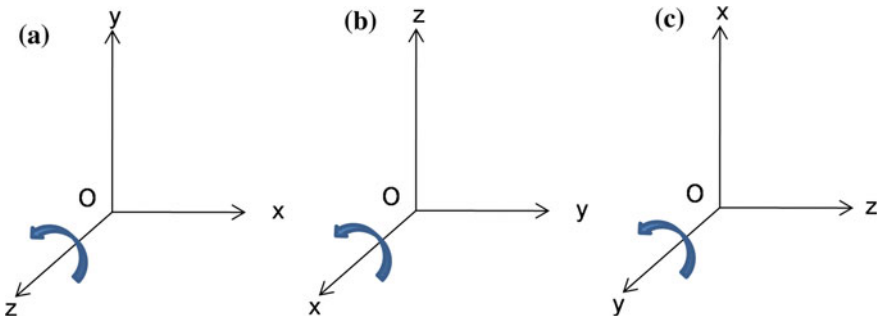


Fig. 2 Mesh rotation operation with respect to **a** z-axis, **b** x-axis, and **c** y-axis

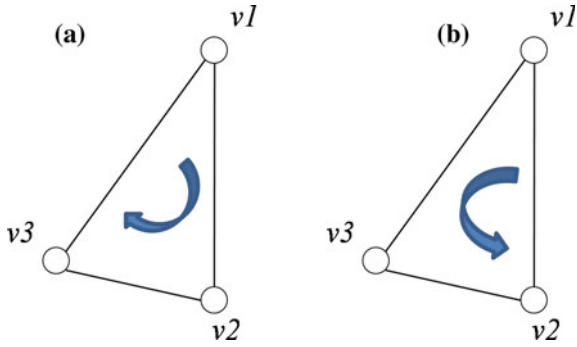


Fig. 3 Example: Vertex ordering, **a** ordering will read the vertices in the sequence v1,v2,v3, **b** ordering will read the vertices in the sequence v1,v3,v2

$$\begin{aligned} z' &= z * \cos q - x * \sin q \\ x' &= z * \sin q + x * \cos q \\ y' &= y \end{aligned} \tag{6}$$

2.1 Vertex Reordering

The order of the vertices in a mesh is determined by the order they were assigned when created. So every triangle in a triangular mesh, three vertices of the triangles have a particular orientation or ordering. If anyone reorders the vertices, the geometry or the topological information of the mesh does not changes. Rather it creates a sort of synchronization problem while reading the mesh elements. Sometimes the mesh processing operations follow a particular vertex sequence ordering. If that ordering changes, the mesh processing operation will not work later due to synchronization problem. Figure 3 shows the vertex ordering and how it can be reordered.

3 Robustness Against RST and Vertex Reordering Operation

To make the mesh robust against RST operations and vertex reordering, two different types of approaches are there: *selecting proper embedding primitives* and *performing certain preprocessing steps*.

3.1 Selection of Proper Embedding Primitives

One class of mesh watermarking algorithms has solved the problem of RST invariance by selecting proper embedding primitives which are invariant to such operations. The various mesh watermark embedding primitives can be divided into two groups, namely *geometrical embedding primitives* and *topological embedding primitives*. Geometrical embedding primitives comprise of various positional as well as distance information of the mesh elements, and topological attributes include mesh elements connectivity information. Table 1 shows different mesh embedding primitives, their invariance toward the RST attacks and list of mesh watermarking algorithms that have used the particular embedding primitive.

3.2 Mesh Preprocessing Steps

In this approach, a number of steps are carried out on the mesh before the actual watermarking process which is also mentioned as preprocessing steps. These steps are performed one after another to generate a mesh invariant to such modification. These are also known as mesh normalization operations.

Mesh translation invariance: Translation of the mesh Cartesian origin to the mesh centroid (center of mass) so that it becomes invariant to any translation operation.

$$O = \frac{1}{area_t} \sum_{i=1}^{m} area_i * \frac{A_i + B_i + C_i}{3} \tag{7}$$

Here the $area_i$ is the area of the triangle formed by A_i, B_i, $and C_i$ and $area_t$ is the total area of the triangles. Now to compute the vertices of the translated mesh,

$$\begin{aligned} x_i' &= x_i - o_x \\ y_i' &= y_i - o_y \\ z_i' &= z_i - o_z \end{aligned} \tag{8}$$

Table 1 Table showing the list embedding primitives and their invariance toward RST

Mesh embedding primitives	Used by methods	Invariance toward
Geometrical embedding primitives		
Vertex coordinates	–	Nothing
Length of an edge	–	TR
Length of mesh centroid to face centroid	Ref. [8]	TR
Length of mesh centroid to vertex	Ref. [8]	TR
Distance between one ring vertex neighbor and centroid to vertex	Ref. [9]	TR
Length of face centroid to vertex	Ref. [3]	
Ratio of length of two segments of line	Ref. [5]	RST
Ratio of face edges	–	RST
Angles of a face	–	TR
Vertex norms	–	RST
First-order moments	–	RST
Second-order moments	–	RST
Ratio between two angles	–	RST
Area of a face	–	TR
Ratio of areas of two faces	–	RST
Volume of polyhedron	–	TR
Ratio of volumes of polyhedrons		RST
Topological embedding primitives		
Vertex connectivity	[4]	RST
Edge connectivity	–	RST

Mesh uniform scaling invariance: Uniform scaling or normalizing the mesh vertex coordinate values to a range is carried out by finding the distance between the two furthest vertices. The rest of the vertices are divided by the scaling factor α as shown in Eqs. 9 and 10 with a rotation matrix whose. Here NV is the total number of vertices.

$$\alpha = max\{(x_i')^2 + (y_i')^2 + (z_i')^2\} - min\{(x_i')^2 + (y_i')^2 + (z_i')^2\} i = 1, 2, \ldots NV \tag{9}$$

$$x_i'' = \frac{x_i'}{\alpha}, y_i'' = \frac{y_i'}{\alpha}, z_i'' = \frac{z_i'}{\alpha} \tag{10}$$

Mesh rotation invariance: To make the mesh robust to rotation operation or rather to make it pose invariant, one of the solution is to compute models orthogonal axis. *Principal component analysis* (PCA) has been used to compute the three orthogonal axes [7]. This process aligns the 3D model's orientation based on PCA eigenvectors. In the first step, the covariance matrix for the model is computed as shown in Eq. 11. The eigenvectors are computed, sorted, and scaled to unit Euclidean length. The original 3D model is rotated with a rotation matrix whose rows are scaled eigenvectors

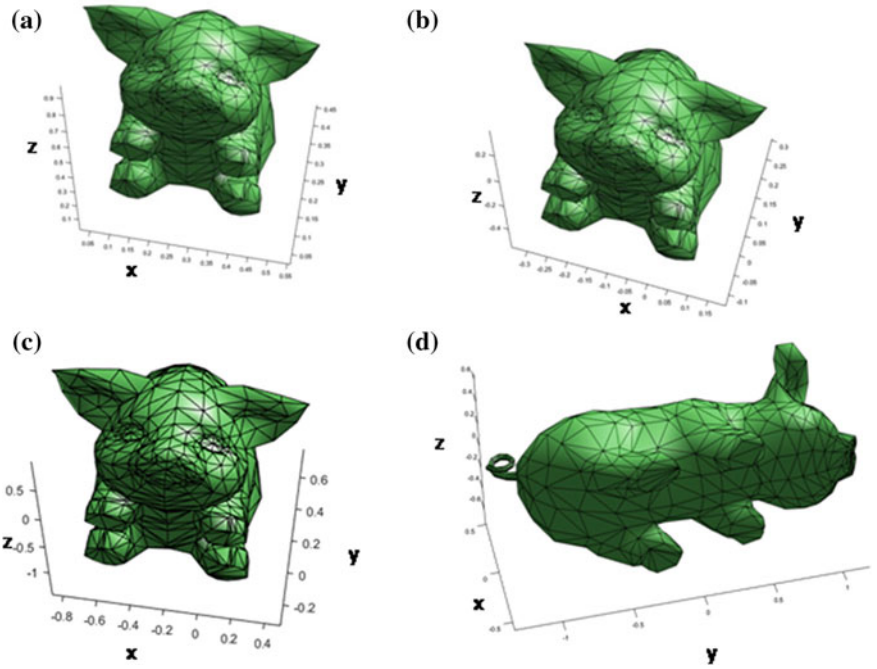


Fig. 4 Example: **a** original 3D model, **b** 3D model invariant to translation, **c** 3D model uniform scaling invariant, and **d** 3D model rotation invariant

as shown in equation. However PCA alignment may sometimes come out to be imprecise and can produce poor alignments [2].

$$I = \begin{bmatrix} \sum_{i=1}^{N_v} x_i^2 & \sum_{i=1}^{N_v} x_i y_i & \sum_{i=1}^{N_v} x_i z_i \\ \sum_{i=1}^{N_v} y_i x_i & \sum_{i=1}^{N_v} y_i^2 & \sum_{i=1}^{N_v} y_i z_i \\ \sum_{i=1}^{N_v} z_i x_i & \sum_{i=1}^{N_v} z_i y_i & \sum_{i=1}^{N_v} z_i^2 \end{bmatrix} \tag{11}$$

$$\begin{bmatrix} x_i' \\ y_i' \\ z_i' \end{bmatrix} = R * \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \tag{12}$$

Minimum enclosing volume box can also be computed for the 3D mesh using maximum normal distribution (MND) approach [6]. In this method, the three orthogonal axes are computed that fit better *human visual perception* perception. The maximum normal distribution is considered as one of the principal axis in this method. There is another operation which is a form of rotation operation, *mesh reflection*. In 3D reflection, the reflection takes place about a plane, whereas 2D reflection it used takes place about an axis. This involves 180° rotation of the mesh over a plane of rotation.

Mesh vertex reordering invariance: Vertex ordering is used normally while traversing the mesh vertices or mesh elements. The traversal process should not be dependent on the actual vertex ordering so that the traversal becomes invariant to vertex reordering operation.

- Distribution of mesh vertex norms is invariant to vertex reordering.
- To embed watermark bits, a particular mesh sequence is required. Vertex sequence computation should be independent of the mesh original vertex ordering. So it should be dependent on the mesh geometry information.

The preprocessing operations were applied to 3D meshes to make the watermarking process affine transformation invariant. An example can be seen in Fig. 4.

4 Conclusion

In case of mesh watermarking techniques, a robust watermarking scheme or a semi-fragile watermarking scheme should be able to resist a set of four simple mesh processing operations. Either selecting appropriate mesh embedding primitives or performing proper mesh normalization operation the robustness can be achieved. Mesh normalization is an important step which brings the mesh into a canonical state which is suitable to process with. In this paper, the various procedures to make the meshes invariant to the normal mesh operations are explained and implemented.

References

1. Botsch, M., Kobbelt, L., Pauly, M., Alliez, P., Lévy, B.: Polygon Mesh Processing. CRC press, Boca Raton (2010)
2. Chaouch, M., Verroust-Blondet, A.: A novel method for alignment of 3D models. In: 2008 IEEE International Conference on Shape Modeling and Applications, SMI 2008, pp. 187–195. IEEE, New York (2008)
3. Chou, C.M., Tseng, D.C.: Affine-transformation-invariant public fragile watermarking for 3D model authentication. IEEE Comput. Graph. Appl. **29**(2), 72–79 (2009)
4. Dittmann, J., Benedens, O.: Invertible authentication for 3D meshes. In: Electronic Imaging 2003. International Society for Optics and Photonics, pp. 653–664 (2003)
5. Ji, H., Yang, X., Zhang, C., Gao, X.: A new reversible watermarking of 3D models based on ratio expansion. In: 2010 3rd International Congress on Image and Signal Processing (CISP), vol. 8, pp. 3899–3903. IEEE, New York (2010)
6. Jiantao, P., Yi, L., Guyu, X., Hongbin, Z., Weibin, L., Uehara, Y.: 3d model retrieval based on 2D slice similarity measurements. In: Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT 2004, pp. 95–101. IEEE, New York (2004)
7. Molaei, A.M., Ebrahimnezhad, H., Sedaaghi, M.H.: A blind fragile watermarking method for 3d models based on geometric properties of triangles. 3D Res. **4**(4), 1–9 (2013)
8. Wu, H.T., Cheung, Y.M.: A fragile watermarking scheme for 3D meshes. In: Proceedings of the 7th Workshop on Multimedia and Security, pp. 117–124. ACM (2005)
9. Wu, H.T., Dugelay, J.L.: Reversible watermarking of 3D mesh models by prediction-error expansion. In: 2008 IEEE 10th Workshop on Multimedia Signal Processing, pp. 797–802. IEEE, New York (2008)

Author Index

A
Abdul Moiz Qyser, Ahmed, 9
Abiramy, N. V., 287
Adate, Amit, 211
Agarwal, Rakshanda, 515
Agrawal, Rashmi, 479
Akkineni, Haritha, 683
Amarendra, Ch., 559
Anand, D., 135
Annie Micheal, A., 219
Aruna Devi, B., 671
Ayyasamy, S., 183

B
Baby, Cyril Joe, 305
Balamuralikrishna, Thati, 163
Balamurugan, S., 153
Balasubramanian, Prashanth, 85
Banerjee, Sreeparna, 1
Baskar, V. Vijaya, 625
Benisha, M., 335, 367
Bhagya Sri, H. V., 267
Bharathi, C. R., 649
Bhatia, Varun, 325
Bhoi, Bandan Kumar, 237
Bobba, Rishitha, 683
Borah, Bhogeswar, 721
Borah, Sagarika, 721

C
Chamania, Dhruv, 211
Chandran, C. P., 505
Chandrasekhar Reddy, P., 447
Chandusha, K., 35
Chapaneri, Radhika, 325, 345
Chintalapudi, S. Rao, 35
Chrysolite, Evangelin, 107

D
Daimiwal, Nivedita, 625, 637
DasGupta, Suparna, 571
Das, Suman Kumar, 571
Devasia, Thresiamma, 659
Dey, Nilanjan, 193
Dhanalakshmi, D., 581
Dileep Kumar, K., 267

F
Fernandes, Steven Lawrence, 193

G
Ganesh, A. Balaji, 203
Geetha, P., 219
Ghose, Dillip K., 53
Gogulamudi, Lakshmi, 683
Govardhan, A., 19
Gupta, Rajeev Kumar, 437

H
Harinadha Reddy, K., 559
Harshali, Patil, 229
Hemalatha, M., 315
Hussain, Mohammed Ali, 163

J
Jacob, K. Poulose, 659
Jayanthi, VE., 121
John, Binu, 85
Joshi, N., 413
Joshi, Satej, 325

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 104,
https://doi.org/10.1007/978-981-13-1921-1

K
Kakollu, Vanitha, 447
Kar, Chinmoy, 1
Kashyap, Ramgopal, 437
Kishore Babu, S., 393
Kotipalli, Pushpa, 459
Krishna Prasad, M. H. M., 35
Krishna Prasad, P. E. S. N., 525
Kumar, M. Pradeep, 603

L
Lenin, K., 45

M
Maduraiveeran, C., 603
Manic, K. Suresh, 193
Manjula, R., 703
Martin, Betty, 625, 637
Mary Neebha, T., 107
Mathew, Jimmy, 75, 85
Menaha, R., 121
Misra, Neeraj Kumar, 237
Mohan, B. A., 259
Mohan Raju, M. A. S., 459
Muneeswaran, V., 693
Muni Kumar, N., 703
Murali Krishna, D., 459
Murali Krishna, M., 267

N
Nagadeepa, N., 487
Nagaraju, C., 535
Narayanan, Swathi Jamjala, 305
Narayana, V. Lakshman, 649
Narendrababu Reddy, G, 357
Narsimha, G., 447
Nesasudha, M., 107

P
Pallikonda Rajasekaran, M., 671, 693
Pandit, Malay Kumar, 713
Panduranga Vital, T., 267
Papineni, V. S. Lakshmi, 683
Paritala, Sravya, 683
Parveen Sultana, H., 515
Paul, Suman, 713
Pawan Sai, R., 65
Pemula, Rambabu, 535
Peram, Subba Rao, 247
Perumal, Boominathan, 305
Phani Kumar, S., 357
Poornapushpakala, S., 637
Potnuru, Devendra, 145
Pradhan, Manoranjan, 237
Pranay, Sai Satya, 515
Prasad, B. V. V. S., 247
Praveen Kumar, E. R., 459
Praveen, S. Phani, 591
Priyanka, J. S. V. S. Hari, 175

R
Rachana, K., 515
Ragaleela, D., 495
Rajasekaran, M. Pallikonda, 603
Raja Sekar, M., 613
Rajinikanth, V., 193
Raju, Ch. Padmanabha, 581
Ramakrishna Murthy, M., 175
Rama, S., 603
Ranjeetha, R., 367
Rao, B. Tarakeswara, 247
Rao, J. Ranga, 403
Rao, K. Thirupathi, 591
Rao, M. R. Narasinga, 135
Rashmi, R., 381, 469
Raveendra Babu, P., 403
Razia, Shaik, 135
Reddy, Ch. Rami, 547
Reddy, K. Harinadha, 547
Reddy, K. Venkata Siva, 547
Roseline, A. Christina, 277
Rout, Rashmishree, 237
Rupa, Ch., 403

S
Saha, Soumyabrata, 571
Sahoo, Kalyani, 297
Sahoo, Ramesh K., 297
Sai Charan, P. V., 97
Sailaja, D., 175
Sakthivel, S., 603
Samantaray, Sandeep, 53
Sandhya, N., 613
Saraswathi, P., 487
Saraswathi, S., 153
Sarojadevi, H., 259
Saruladha, K., 315
Satapathy, Suresh Chandra, 193
Sathyabama, K., 315
Seema, Purohit, 229
Sethi, Srinivas, 297
Seth, Siddharth, 425
Sevugapandi, N., 505
Shah, S., 413
Shah, Seema, 345
Shriram, Revati, 625, 637

Shukla, Ruchi, 437
Shylu Sam, D. S., 277
Singh, Bhupesh, 85
Singh, Praveen Kumar, 525
Sinha, Arihant, 75
Sivanagaraju, S., 495
Smilarubavathy, G., 183
Soundrapandiyan, Rajkumar, 305
Sri, M. Navya, 175
Srinivas, Jagirdar, 9
Sucheta, P. S. V., 19
Sudha, S. V., 287
Sultana, Parveen, 211
Sundararajan, M., 637
Sundhararajan, M., 625
Sunil, M. P., 65
Swami Das, M., 19

T
Thandaiah Prabu, R., 335, 367
Thilagaraj, M., 603
Thomas, Tessamma, 659
Thulasi Bai, V., 335, 367
Tummala, Ayyarao S. L. V., 145

U
Uplane, M. D., 381, 469

V
Vankayalapati, Hima Deepthi, 135
Vasavi, S., 393
Venkata Subba Reddy, Kothapuli, 9
Venkatesan, R., 203
Verma, Rajesh Kumar, 525
Videla, Lakshmi Sarvani, 135
Vijaya Lakshmi, D., 19

Z
Zaveri, Mukesh A., 425