Lecture Notes in Computational Vision and Biomechanics 31

J. Dinesh Peter
Steven Lawrence Fernandes
Carlos Eduardo Thomaz
Serestina Viriri *Editors*

Computer Aided Intervention and Diagnostics in Clinical and Medical Images

Springer

Lecture Notes in Computational Vision and Biomechanics

Volume 31

The research related to the analysis of living structures (Biomechanics) has been a source of recent research in several distinct areas of science, for example, Mathematics, Mechanical Engineering, Physics, Informatics, Medicine and Sport. However, for its successful achievement, numerous research topics should be considered, such as image processing and analysis, geometric and numerical modelling, biomechanics, experimental analysis, mechanobiology and enhanced visualization, and their application to real cases must be developed and more investigation is needed. Additionally, enhanced hardware solutions and less invasive devices are demanded.

On the other hand, Image Analysis (Computational Vision) is used for the extraction of high level information from static images or dynamic image sequences. Examples of applications involving image analysis can be the study of motion of structures from image sequences, shape reconstruction from images, and medical diagnosis. As a multidisciplinary area, Computational Vision considers techniques and methods from other disciplines, such as Artificial Intelligence, Signal Processing, Mathematics, Physics and Informatics. Despite the many research projects in this area, more robust and efficient methods of Computational Imaging are still demanded in many application domains in Medicine, and their validation in real scenarios is matter of urgency.

These two important and predominant branches of Science are increasingly considered to be strongly connected and related. Hence, the main goal of the LNCV&B book series consists of the provision of a comprehensive forum for discussion on the current state-of-the-art in these fields by emphasizing their connection. The book series covers (but is not limited to):

- Applications of Computational Vision and Biomechanics
- Biometrics and Biomedical Pattern Analysis
- Cellular Imaging and Cellular Mechanics
- Clinical Biomechanics
- Computational Bioimaging and Visualization
- Computational Biology in Biomedical Imaging
- Development of Biomechanical Devices
- Device and Technique Development for Biomedical Imaging
- Digital Geometry Algorithms for Computational Vision and Visualization
- Experimental Biomechanics
- Gait & Posture Mechanics
- Multiscale Analysis in Biomechanics
- Neuromuscular Biomechanics
- Numerical Methods for Living Tissues
- Numerical Simulation
- Software Development on Computational Vision and Biomechanics
- Grid and High Performance Computing for Computational Vision and Biomechanics
- Image-based Geometric Modeling and Mesh Generation
- Image Processing and Analysis
- Image Processing and Visualization in Biofluids
- Image Understanding
- Material Models
- Mechanobiology
- Medical Image Analysis
- Molecular Mechanics
- Multi-Modal Image Systems
- Multiscale Biosensors in Biomedical Imaging
- Multiscale Devices and Biomems for Biomedical Imaging
- Musculoskeletal Biomechanics
- Sport Biomechanics
- Virtual Reality in Biomechanics
- Vision Systems

More information about this series at http://www.springer.com/series/8910

J. Dinesh Peter · Steven Lawrence Fernandes
Carlos Eduardo Thomaz · Serestina Viriri
Editors

Computer Aided Intervention and Diagnostics in Clinical and Medical Images

Springer

Editors
J. Dinesh Peter
Department of Computer
Science and Engineering
Karunya Institute of Technology
and Sciences
Coimbatore, Tamil Nadu, India

Steven Lawrence Fernandes
Department of Electrical
and Computer Engineering
University of Alabama at Birmingham
Birmingham, AL, USA

Carlos Eduardo Thomaz
Brazilian Computer Society
Aalcides Platiny Alves Batista
Porto Alegre, Rio Grande do Sul, Brazil

Serestina Viriri
School of Computer Science
University of KwaZulu-Natal
Durban, South Africa

ISSN 2212-9391 ISSN 2212-9413 (electronic)
Lecture Notes in Computational Vision and Biomechanics
ISBN 978-3-030-04060-4 ISBN 978-3-030-04061-1 (eBook)
https://doi.org/10.1007/978-3-030-04061-1

Library of Congress Control Number: 2018961232

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Brain Tissue Entropy Changes in Patients with Autism Spectrum Disorder 1
Sudhakar Tummala

Differential Coding-Based Medical Image Compression 11
P. Chitra and M. Mary Shanthi Rani

Harmonization of White and Gray Matter Features in Diffusion Microarchitecture for Cross-Sectional Studies 21
Prasanna Parvathaneni, Shunxing Bao, Allison Hainline, Yuankai Huo, Kurt G. Schilling, Hakmook Kang, Owen Williams, Neil D. Woodward, Susan M. Resnick, David H. Zald, Ilwoo Lyu and Bennett A. Landman

Deep Neural Architecture for Localization and Tracking of Surgical Tools in Cataract Surgery 31
Neha Banerjee, Rachana Sathish and Debdoot Sheet

Efficient Segmentation of Medical Images Using Dilated Residual Networks 39
Lokeswara Rao Bonta and N. Uday Kiran

Non-rigid Registration of Brain MR Images for Image Guided Neurosurgery Using Cloud Computing 49
D. Preetha Evangeline and P. Anandhakumar

A Hybrid Fusion of Multimodal Medical Images for the Enhancement of Visual Quality in Medical Diagnosis 61
S. Sandhya, M. Senthil Kumar and L. Karthikeyan

An Amplifying Image Approach: Non-iterative Multi Coverage Image Fusion 71
K. Elaiyaraja, M. Senthil Kumar and L. Karthikeyan

U-Net Based Segmentation and Multiple Feature Extraction of Dermascopic Images for Efficient Diagnosis of Melanoma 81
D. Roja Ramani and S. Siva Ranjani

Secured Transmission of Medical Images in Radiology Using AES Technique 103
Pavithra Prabhu and K. N. Manjunath

A Review on Haze Removal Techniques 113
K. P. Senthilkumar and P. Sivakumar

Secured Image Transmission in Medical Imaging Applications—A Survey 125
Pavithra Prabhu and K. N. Manjunath

Evolution of Methods for NGS Short Read Alignment and Analysis of the NGS Sequences for Medical Applications 135
J. A. M. Rexie and Kumudha Raimond

Caries Detection in Non-standardized Periapical Dental X-Rays 143
D. Osterloh and Serestina Viriri

Segmentation of Type II Diabetic Patient's Retinal Blood Vessel to Diagnose Diabetic Retinopathy 153
T. Jemima Jebaseeli, C. Anand Deva Durai and J. Dinesh Peter

A Novel Corner Elimination Method for the Compression of Wireless Capsule Endoscopic Videos 161
Caren Babu and D. Abraham Chandy

Prediction of Two Year Survival Among Patients of Non-small Cell Lung Cancer 169
Yash Dagli, Saumya Choksi and Sudipta Roy

Prediction of Chronic Kidney Diseases Using Deep Artificial Neural Network Technique 179
Himanshu Kriplani, Bhumi Patel and Sudipta Roy

Monitoring Acute Lymphoblastic Leukemia Therapy with Stacked Denoising Autoencoders 189
Jakob Scheithe, Roxane Licandro, Paolo Rota, Michael Reiter, Markus Diem and Martin Kampel

Modified Low-Power Built-in Self-test for Image Processing Application 199
P. Anitha, P. Ramanathan and P. T. Vanathi

A Hassle-Free Shopping Experience for the Visually Impaired: An Assistive Technology Application 207
Sherin Tresa Paul and Kumudha Raimond

Retina as a Biomarker of Stroke 219
R. S. Jeena, A. Sukeshkumar and K. Mahadevan

Distributed Representation of Healthcare Text Through Qualitative and Quantitative Analysis 227
J. R. Naveen, H. B. Barathi Ganesh, M. Anand Kumar and K. P. Soman

Detection of Lymph Nodes Using Centre of Mass and Moment Analysis 239
R. Akshai, S. Rohit Krishnan, G. Swetha and B. P. Venkatesh

Estimation of Elbow Joint Angle from Surface Electromyogram Signals Using ANFIS 247
P. Rajalakshmy, Elizabeth Jacob and T. Joclyn Sharon

Video Stabilization for High-Quality Medical Video Compression 255
D. Raveena Judie Dolly, D. J. Jagannath and J. Dinesh Peter

Significance of Global Vectors Representation in Protein Sequences Analysis 261
Anon George, H. B. Barathi Ganesh, M. Anand Kumar and K. P. Soman

Texture Analysis on Thyroid Ultrasound Images for the Classification of Hashimoto Thyroiditis 271
S. Kohila and G. Sankara Malliga

Cluster Based Paddy Leaf Disease Detection, Classification and Diagnosis in Crop Health Monitoring Unit 281
A. D. Nidhis, Chandrapati Naga Venkata Pardhu, K. Charishma Reddy and K. Deepa

Detection of Glaucoma Using Anterior Segment Optical Coherence Tomography Images 293
P. Priyanka, V. Norris Juliet and S. Shenbaga Devi

Brain Tissue Entropy Changes in Patients with Autism Spectrum Disorder

Sudhakar Tummala

Abstract Autism Spectrum Disorder (ASD) is accompanied by brain tissue changes in areas that control behavior, cognition, and motor functions, deficient in the disorder. The objective of this research was to evaluate brain structural changes in ASD patients compared to control subjects using voxel-by-voxel image entropy from T1-weighted imaging data of 115 ASD and 105 control subjects from autism brain imaging data exchange. For all subjects, entropy maps were calculated, normalized to a common space and smoothed. Then, the entropy maps were compared at each voxel between groups using analysis of covariance (covariates; age, gender). Increased entropy in ASD patients, indicating chronic injury, emerged in several vital regions including frontal temporal and parietal lobe regions, corpus callosum, cingulate cortices, and hippocampi. Entropy procedure showed significant effect size and demonstrated wide-spread changes in sites that control social behavior, cognitive, and motor activities, suggesting severe damage in those areas. The neuropathological mechanisms contributing to tissue injury remain unclear and possibly due to factors including genetic, atypical early brain growth during childhood.

Keywords Magnetic resonance imaging · Entropy · Autism spectrum disorder

1 Introduction

Autism spectrum disorder (ASD) is a diverse group of neurodevelopmental disorders represented by repetitive behaviors, abnormal social interaction, and diminished cognition [1, 2]. Subjects with long-standing ASD show regional and voxel-level tissue changes in several brain sites that regulate motor, social-cognitive, and mood, as evaluated with high-resolution T1-weighted-imaging-based voxel-based morphometry (VBM) and diffusion tensor imaging (DTI)-based diffusivity procedures from

S. Tummala (✉)
Department of Electronics and Communication Engineering,
SRM University-AP, Amaravati 522503, Andhra Pradesh, India
e-mail: sudhakar.t@srmap.edu.in

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_1

magnetic resonance imaging (MRI) [3–6]. VBM procedure has limited sensitivity from the inherent limited range of probability values, and is not suitable to detect subtle chronic/acute gray or white matter changes; thus, the procedure is unable to differentiate acute from chronic tissue pathology. Although, DTI based mean diffusivity, axial and radial diffusivity metrics, can differentiate such acute versus chronic differences, those images also have an inferior spatial resolution, and require specialized preprocessing.

Image texture is a measure that quantifies spatial patterns of intensities/gray level values and these spatial patterns may differ with respect to the nature and degree of tissue injury. For this purpose, T1-weighted images were employed, which are better suited to detect the changes in intensity patterns due to higher spatial resolution and better gray and white matter contrast.

Entropy, a texture feature, measures the extent of homogeneity or randomness in a given region, based on characteristics of the intensity histogram from high-resolution structural images. The entropy values are inversely proportional to the amount of water content in the tissue. In acute stages of the ASD disease, due to axonal and neural swelling, the amount of water content increases and entropy values decreases, whereas in chronic stages of the disease, the amount of water content decreases and entropy values will be increased. The entropy technique has been used to assess the neural changes in different conditions, including Alzheimer's [7], Parkinson's disease [8], characterization of intracranial tumors [9], and acute inflammation in MS lesions [10]. There were few region-based studies related to image texture for assessing brain tissue in ASD subjects [11, 12]. However, to my knowledge applying *voxel-level* entropy procedure on a larger population to study brain tissue changes in ASD patients is novel.

Here, the aim was to investigate regional brain entropy changes in ASD subjects using high-resolution T1-weighted images from Autism Brain Imaging Data Exchange (ABIDE I) database. The hypothesis is that voxel-level entropy values are higher in ASD subjects in various brain sites involved in cognitive, motor and neuropsychologic regulation compared to control subjects, indicative of chronic tissue changes in those regulatory sites.

2 Materials and Methods

2.1 Study Population

From ABIDE I database, 115 subjects with ASD and 105 age- and gender-comparable control subjects were selected. The control subjects were recruited through several participating sites across the USA. Subjects with ASD were included based on clinicians Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Test

Revision (DSM-IV-TR; American Psychiatric Association [13]) diagnosis of autistic disorder, Asperger's disorder, or pervasive developmental disorder—not otherwise specified. All data collection procedures were carried out with approval from the Institutional Review Boards at all corresponding participating sites.

2.2 Magnetic Resonance Imaging

Brain MRI images were downloaded from ABIDE I database (Magnetom Tim-Trio Syngo B17; Siemens, Erlangen, Germany). The high-resolution T1-weighted scans were acquired at 3.0 T field strength with magnetization prepared rapid acquisition gradient echo sequence in the sagittal plane (repetition time TR = 1230, 590, 2300 ms; echo time TE = 1.73, 2.73, 2.9, 3.6 ms; inversion-time = 624, 900 ms; flip-angle = 9°, 10°; in-plane resolution = 0.5 mm × 0.5 mm, 1 mm × 1 mm; slice thickness = 1.0, 1.1, 1.2 mm).

2.3 Cognition Assessment

Study subjects at most of the participating sites were assessed for cognitive skills and ability using one of the Intelligence Quotient (IQ) tests including Wechsler Abbreviated Scale of Intelligence (WASI), Wechsler Adult Intelligence Scale (WAIS), Wechsler Intelligence Scale for Children (WISC), Writing skills test (WST), Hamburg Wechsler Intelligence test for children (HAWIK), Differential Ability Scales (DAS) for estimation of full-scale IQ, verbal IQ, and performance IQ [14–16].

2.4 Data Processing and Analysis

Several software tools were used for image visualization, data preprocessing, and analyses that included the statistical parametric mapping package SPM12 (Wellcome trust center for neuroimaging, UK; http://www.fil.ion.ucl.ac.uk/spm), MRIcroN [17], and in-house MATLAB-based routines (The MathWorks Inc., Natick, MA). For all ASD and control subjects, the images were visually examined using MRIcroN to ensure that there were no motion artifacts and serious anatomical defects.

2.5 Entropy Maps Calculation, Normalization, and Smoothing

For each subject, the T1-weighted image was bias corrected using SPM12 to eliminate any local magnetic field inhomogeneities. Then, the entropy at a given voxel "v" from the bias-corrected high-resolution T1-weighted images was computed in MATLAB from the following mathematical equation:

$$E = -\sum_{i=1}^{N} p_i \log(p_i)$$

by defining a $3 \times 3 \times 3$ Volume of Interest (VOI) with "v" as center. Where, N is the number of gray values in the VOI, and p_i is the probability of occurrence of ith gray value in the VOI. Mathematically, the VOI can be represented as below:

$$\mathrm{VOI} = \mathrm{I}(\mathrm{x}-3:\mathrm{x}+3, \mathrm{y}-3:\mathrm{y}+3, \mathrm{z}-3:\mathrm{z}+3)$$

where "I" is the bias-corrected T1-weighted image, and x, y, z are spatial coordinates. There were no free parameters in the framework. The procedure was the same at all voxels in the bias-corrected T1-weighted image. Further, the warping parameters for x, y, z directions obtained from the bias-corrected T1-weighted images via modified unified segmentation method in SPM12 were applied to the corresponding entropy maps to get normalized entropy maps. The normalized entropy maps were smoothed using an isotropic Gaussian filter in SPM12 to improve the signal-to-noise ratio (kernel size, 10 mm).

2.6 Global Brain Mask

For each subject, the T1-weighted images were segmented into white and gray matter probability maps using SPM12. Then these probability maps were normalized into MNI space using the warping parameters obtained from the normalization step. Then, all the normalized white and gray matter probability maps were averaged to create global white and gray matter probability maps, respectively. The global brain mask was created by thresholding (gray matter > 0.3 and white matter > 0.3) and combining the global white and gray matter probability maps. The global brain mask was used to limit the voxel level analyses to gray and white matter regions.

2.7 Statistical Analyses

Chi-square test and independent-samples T-tests were employed to determine group differences in demographic data and cognitive scores. The smoothed entropy maps between the groups were compared at each voxel using analysis of covariance (ANCOVA, SPM12, covariates: age and gender, uncorrected, $p<0.001$; minimum cluster-size 10 voxels). Brain regions with significant entropy differences between groups were superimposed onto background normalized T1-weighted images for structure classification.

3 Results

3.1 Demographic Variables

No significant differences were found in age ($p=0.32$), gender ($p=0.83$), and handedness ($p=0.59$, Table 1). However, ASD subjects showed significantly lower full-scale IQ ($p=0.02$), verbal IQ ($p<0.0001$), performance IQ ($p<0.0001$) scores compared to control subjects (Table 1).

3.2 Brain Regions with Entropy Changes

Several brain areas in ASD patients showed increased entropy values, showing chronic tissue changes in those regions, compared to control subjects (Fig. 1, $p<0.001$, uncorrected). Brain sites in ASD patients that showed increased entropy values included the bilateral superior frontal white matter (Fig. 2a), bilateral parietal cortices, and bilateral genu of corpus callosum spreading to anterior cingulate

Table 1 Differences in clinical variables were assessed using the χ^2 test for gender, handedness, and Mann–Whitney U test for rest of the variables. ASD: Autism Spectrum Disorder, DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Test Revision, IQ: Intelligence Quotient, N: Number of subjects, SD: Standard Deviation

Variables	Control (mean ± SD) N = 105	ASD (mean ± SD) N = 115	P-value
Age (years)	28.9 ± 9.4	20 ± 9	0.32
Gender	95 male	105 male	0.83
DSM-IV-TR		1.53 ± 0.9 (N = 105)	NA
Full-scale IQ	89.8 ± 24.7 (N = 58)	79.6 ± 21.8 (N = 59)	0.02
Verbal IQ	111.8 ± 14.9 (N = 90)	98.6 ± 18.3 (N = 94)	<0.0001
Performance IQ	111.3 ± 14.4 (N = 90)	96.1 ± 19.6 (N = 86)	<0.0001

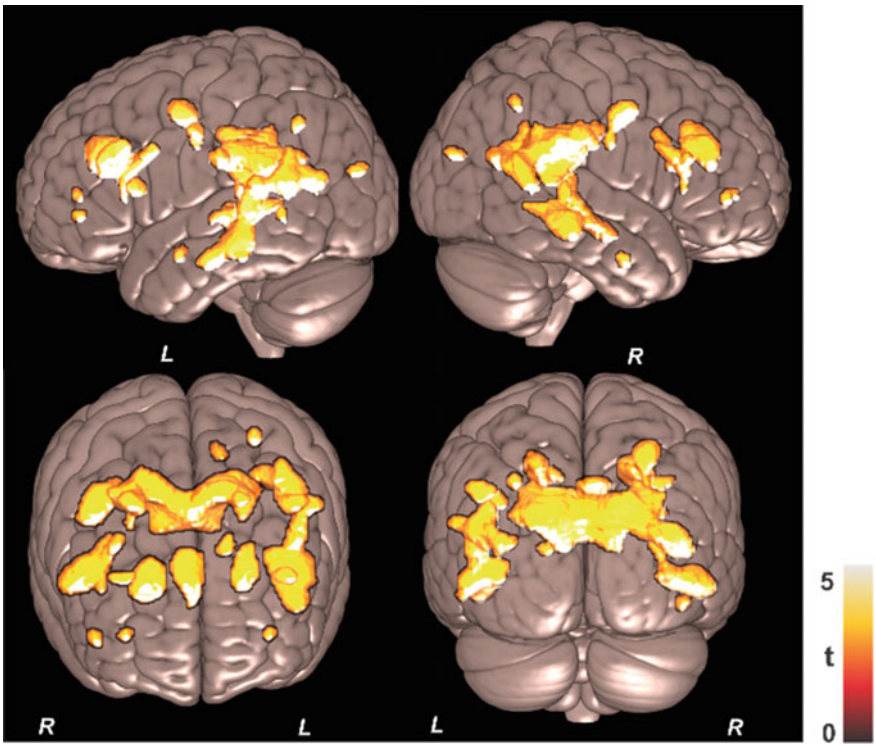


Fig. 1 Brain regions with statistically significant higher entropy values in ASD patients compared to controls were superimposed onto a 3D semitransparent whole brain for structure identification. Images were shown in both coronal and sagittal views. R: Right, L: Left. Color bar shows t-statistic values

cortices (Fig. 2b, c), bilateral splenium of corpus callosum spreading to posterior cingulate cortices (Fig. 2d, e), bilateral posterior hippocampus spreading to parietal white matter (Fig. 2g), bilateral mid-temporal white matter (Fig. 2f), left occipital cortex (Fig. 2b). No brain regions showed decreased entropy values in ASD patients compared to controls.

4 Discussion

4.1 Brain Tissue Changes in ASD

Previous studies speculated that several factors including accelerated/atypical brain growth start during early childhood, oligogenetic, and aberrant micro-organization of the cortex may trigger processes that lead to an axonal and neuronal loss in specific brain areas [18, 19]. White matter and cortical surface changes, gray matter

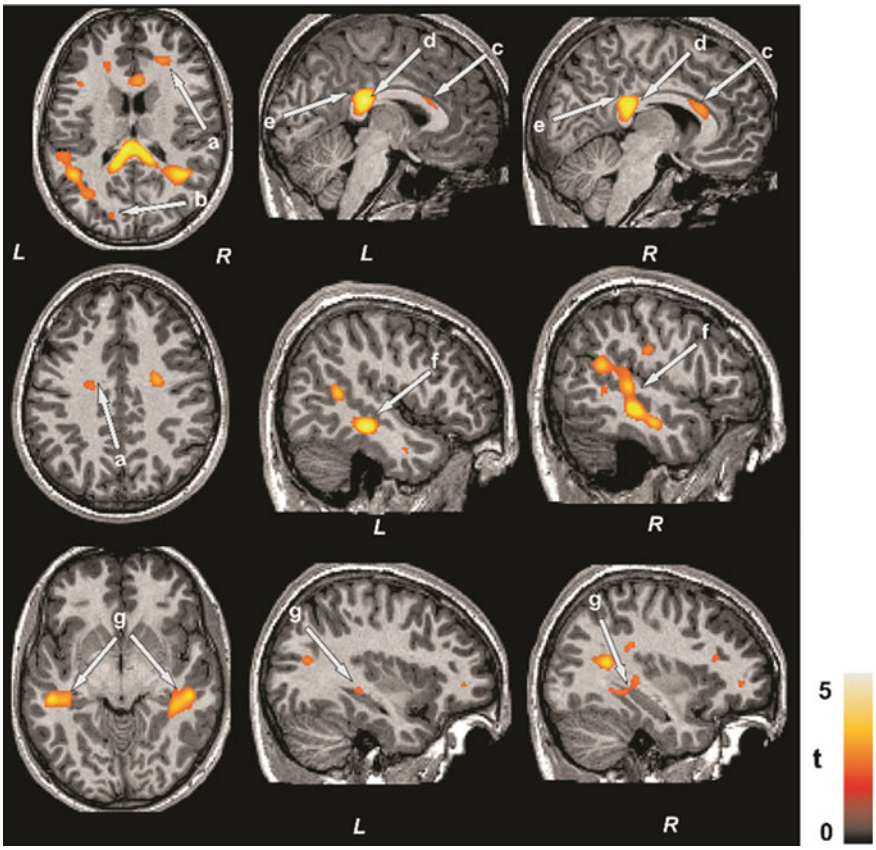


Fig. 2 Brain areas showing increased values in entropy in ASD over control subjects (uncorrected $p < 0.001$). These sites included bilateral superior frontal white matter (**a**), bilateral parietal cortices and left occipital cortex (**b**), bilateral genu of corpus callosum extending to anterior parietal cortices (**c**), bilateral splenium of corpus callosum extending to posterior parietal cortices (**d**, **e**), bilateral mid-temporal white matter (**f**), bilateral posterior hippocampi spreading to parietal cortices (**g**). Images are according to the neurological convention (L, left; R, right), and color bar indicates t-statistic values

volume alterations, functional connectivity changes, and metabolic deficits were examined by various MRI and spectroscopy procedures in ASD subjects [20–22]. Brain tissue chronic changes in ASD subjects, examined by DTI-based fractional anisotropy, mean and diffusivity-based procedures, were emerged in the corpus callosum, medial temporal cortex, ventrolateral prefrontal cortex, longitudinal fasciculus, uncinate fasciculus, and cingulum bundle sites [5, 23]. A recent longitudinal study in males spanning 9 years showed increased mean, axial, radial diffusion in posterior corpus callosum indicating chronic tissue changes in those areas [18]. However, now the tissue changes were examined using entropy procedure from T1-weighted images, which offer relatively superior spatial resolution and provides effective way of assessing extent and nature of tissue injury.

4.2 Entropy in Assessment of Brain Tissue Changes

Entropy was computed from intensities of high-resolution T1-weighted images, and it represents a potential marker of local tissue homogeneity. Tissue homogeneity decreases in chronic disease condition with tissue injury, and increases in acute condition due to excess water content due to neuronal swelling, and thus, entropy measures may be suitable to differentiate chronic from acute tissue pathology [24]. In this study, multiple brain regions were identified with increased entropy values, also indicating chronic disease condition. Moreover, various brain areas including the bilateral superior frontal white matter, bilateral parietal white matter, bilateral genu of corpus callosum spreading to anterior cingulate cortices, bilateral splenium of corpus callosum spreading to posterior cingulate cortices, bilateral posterior hippocampi, bilateral mid-temporal white matter showed wide-spread damage than previously identified brain tissue changes using other imaging-based procedures [4, 5, 18].

4.3 Entropy Changes: Neuropsychologic, Motor, and Cognitive Regulatory Sites

Brain tissue changes were apparent in sites important for neuropsychologic, motor, and cognitive control. These areas included bilateral superior frontal white matter, bilateral parietal white matter, bilateral parietal cortices, bilateral genu of corpus callosum spreading to anterior cingulate cortices, bilateral splenium of corpus callosum spreading to posterior cingulate cortices, bilateral posterior hippocampi, and bilateral mid-temporal white matter regions. The genu and splenium of corpus callosum responsible for information integration between hemispheres showed predominant wide-spread changes [25]. The hippocampi and its adjacent regions responsible for social cognition (social and emotional functioning) and memory integration and retrieval demonstrated chronic changes in ASD patients [26]. The posterior cingulate cortices and cingulum bundle, sites responsible for regulating neuropsychologic functions, including arousal state, balance and breadth of attention symptoms showed damage, and this damage may interrupt linkage of fiber crossings with other adjacent white matter tracts [27]. Along with neuropsychologic deficit, temporal and parietal cortices also responsible for regulating emotions and visual memory, language and object processing showed widespread damage [28, 29].

4.4 Limitations

There were few limitations in the study. The exact ASD disease duration was unknown due to several potential factors. These factors could include subjects living with the disease for several years but they were diagnosed and identified lately. Also,

the treatment duration for some of the ASD subjects may also have an impact on the findings here. However, our results show increased entropy values, indicating that the majority of ASD subjects showed chronic brain pathology. Also, significant differences emerged in full-scale IQ, verbal IQ, and performance IQ between groups should have influenced current study findings if used them as covariates.

4.5 Conclusions

Brain entropy values are significantly increased in patients with ASD over control subjects, suggesting predominantly chronic tissue changes in those sites. During chronic changes, tissue becomes more disorganized, randomness increases, and thus, entropy values are increased. Brain areas with increased entropy values in ASD subjects are localized in the critical motor, neuropsychologic, and cognitive control sites including the frontal and parietal white matter regions, genu of corpus callosum spreading to anterior cingulate cortices, splenium of corpus callosum spreading to posterior cingulate cortices, posterior hippocampi, mid-temporal white matter. The underlying neurobiological mechanisms contributing to brain tissue changes were not clear and probably due to genetic and accelerated abnormal brain growth during childhood in the disease. Therefore, the findings suggest that image entropy examination from T1-weighted imaging can demonstrate chronic/acute tissue changes in the autistic brain.

Acknowledgements I would like to thank Autism Brain Imaging Data Exchange for providing demographic as well as MRI data for this study.

References

1. Rapin I (1997) Autism. N Engl J Med 337:97–104
2. Geschwind DH, Levitt P (2007) Autism spectrum disorders: developmental disconnection syndromes. Curr Opin Neurobiol 17:103–111
3. Yang DY, Beam D, Pelphrey KA, Abdullahi S, Jou RJ (2016) Cortical morphological markers in children with autism: a structural magnetic resonance imaging study of thickness, area, volume, and gyrification. Mol Autism 7:11
4. Barnea-Goraly N, Kwon H, Menon V, Eliez S, Lotspeich L, Reiss AL (2004) White matter structure in autism: preliminary evidence from diffusion tensor imaging. Biol Psychiatry 55:323–326
5. Alexander AL, Lee JE, Lazar M, Boudos R, DuBray MB, Oakes TR, Miller JN, Lu J, Jeong EK, McMahon WM, Bigler ED, Lainhart JE (2007) Diffusion tensor imaging of the corpus callosum in Autism. Neuroimage 34:61–73
6. Ismail MM, Keynton RS, Mostapha MM, ElTanboly AH, Casanova MF, Gimel'farb GL, El-Baz A (2016) Studying autism spectrum disorder with structural and diffusion magnetic resonance imaging: a survey. Front Hum Neurosci 10:211
7. Maani R, Yang YH, Kalra S (2015) Voxel-based texture analysis of the brain. PLoS One 10:e0117759

8. Sikio M, Holli-Helenius KK, Harrison LC, Ryymin P, Ruottinen H, Saunamaki T, Eskola HJ, Elovaara I, Dastidar P (2015) MR image texture in Parkinson's disease: a longitudinal study. Acta Radiol 56:97–104
9. Kjaer L, Ring P, Thomsen C, Henriksen O (1995) Texture analysis in quantitative MR imaging. Tissue characterisation of normal brain and intracranial tumours at 1.5 T. Acta Radiol 36:127–135 (1995)
10. Michoux N, Guillet A, Rommel D, Mazzamuto G, Sindic C, Duprez T (2015) Texture analysis of T2-weighted MR images to assess acute inflammation in brain MS lesions. PLoS One 10:e0145497
11. Chaddad A, Desrosiers C, Toews M (2017) Multi-scale radiomic analysis of sub-cortical regions in MRI related to autism, gender and age. Sci Rep 7:45639
12. Radulescu E, Ganeshan B, Minati L, Beacher FD, Gray MA, Chatwin C, Young RC, Harrison NA, Critchley HD (2013) Gray matter textural heterogeneity as a potential in-vivo biomarker of fine structural abnormalities in Asperger syndrome. Pharmacogenom J 13:70–79
13. Diagnostic and statistical manual of mental disorders. American Psychiatric Association, Washington DC test revision (2000)
14. Wechsler D (1999) Wechsler abbreviated scale of intelligence psychological corporation. San Antonio, TX
15. Wechsler D (1997) WAIS-III: Wechsler adult intelligence scale. Psychological Corporation, San Antonio, TX
16. Wechsler D (2003) Wechsler intelligence scale for children San Antonio, 4th ed. TX, Psychological Corporation
17. Rorden C, Karnath HO, Bonilha L (2007) Improving lesion-symptom mapping. J Cogn Neurosci 19:1081–1088
18. Travers BG, Tromp do PM, Adluru N, Lange N, Destiche D, Ennis C, Nielsen JA, Froehlich AL, Prigge MB, Fletcher PT, Anderson JS, Zielinski BA, Bigler ED, Lainhart JE, Alexander AL (2015) Atypical development of white matter microstructure of the corpus callosum in males with autism: a longitudinal investigation. Mol Autism 6:15
19. Parellada M, Penzol MJ, Pina L, Moreno C, Gonzalez-Vioque E, Zalsman G, Arango C (2014) The neurobiology of autism spectrum disorders. Eur Psychiatr 29:11–19
20. Libero LE, Reid MA, White DM, Salibi N, Lahti AC, Kana RK (2016) Biochemistry of the cingulate cortex in autism: an MR spectroscopy study. Autism Res 9:643–657
21. Aoki Y, Kasai K, Yamasue H (2012) Age-related change in brain metabolite abnormalities in autism: a meta-analysis of proton magnetic resonance spectroscopy studies. Trans Psychiatr 2:e69
22. Friedman SD, Shaw DW, Artru AA, Dawson G, Petropoulos H, Dager SR (2006) Gray and white matter brain chemistry in young children with autism. Arch Gen Psychiatr 63:786–794
23. Libero LE, Burge WK, Deshpande HD, Pestilli F, Kana RK (2016) White matter diffusion of major fiber tracts implicated in autism spectrum disorder. Brain Connect 6:691–699
24. Fozouni N, Chopp M, Nejad-Davarani SP, Zhang ZG, Lehman NL, Gu S, Ueno Y, Lu M, Ding G, Li L, Hu J, Bagher-Ebadian H, Hearshen D, Jiang Q (2013) Characterizing brain structures and remodeling after TBI based on information content, diffusion entropy. PLoS One 8:e76343
25. Bloom JS, Hynd GW (2005) The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? Neuropsychol Rev 15:59–71
26. Geib BR, Stanley ML, Dennis NA, Woldorff MG, Cabeza R (2017) From hippocampus to whole-brain: The role of integrative processing in episodic memory retrieval. Human Brain Mapp 38:2242–2259
27. Leech R, Sharp DJ (2014) The role of the posterior cingulate cortex in cognition and disease. Brain 137:12–32
28. Pertzov Y, Miller TD, Gorgoraptis N, Caine D, Schott JM, Butler C, Husain M (2013) Binding deficits in memory following medial temporal lobe damage in patients with voltage-gated potassium channel complex antibody-associated limbic encephalitis. Brain 136:2474–2485
29. Sarkheil P, Goebel R, Schneider F, Mathiak K (2013) Emotion unfolded by motion: a role for parietal lobe in decoding dynamic facial expressions. Soc Cogn Affect Neurosci 8:950–957

Differential Coding-Based Medical Image Compression

P. Chitra and M. Mary Shanthi Rani

Abstract Modern trends of technology face the challenges of cost-effective massive data storage and transmission. Image compression is a master key for this issue. Basically, the process of image compression reduces redundant and irrelevant information from the original data resulting in reduced data file size. Vector quantization is a lossy image compression technique which helps to achieve higher compression with less computation complexity. The aim of the proposed work is to develop a novel medical image compression method that blends differential encoding and vector quantization (VQ). The basic idea is to transform the input image blocks into a set of difference vectors (difference between the each pixel intensity value and its respective mean). The difference vectors are normalized to preserve the sign and further quantized to generate the codebook. The algorithm is also investigated with other statistical moments like median and mode for finding the difference vectors. The experimental results with test medical images have demonstrated better performance of the proposed method when compared to similar methods.

1 Introduction

Image compression is the process of representing an image with reduced bit storage. It makes the reduction by eliminating the redundant data which gives a compact representation of the data. There are various and innumerable applications implementing image compression process to get efficient storage and transmission [1]. In particular, medical field has large set of valuable medical image data for sending and receiving through the internet. In addition to transmission time, the data should be recovered without losing their quality. Normally, image compression can be either lossy or lossless compression. The original data can be perfectly recovered using lossless image compression [2]. There is some popular lossless compression techniques commonly

P. Chitra · M. Mary Shanthi Rani (✉)
Department of Computer Science and Applications, The Gandhigram Rural Institute - Deemed to be University, Gandhigram 624302, Tamil Nadu, India
e-mail: m.maryshanthirani@ruraluniv.ac.in

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_2

used for better compression such as Huffman encoding, Run-length encoding, and Arithmetic Encoding, etc. In lossy compression, there is a minor loss in data by eliminating the irrelevant information from the original data. Vector Quantization, Block Truncation Coding, and Transform Coding are widely employed lossy compression techniques. Though transform-based compression methods achieve high quality, they have high computational complexity. On the other hand, VQ-based methods are simple and easy to implement. The quantization process reduces the dimension of the image size by applying clustering techniques. The number of clusters and the dimension of the cluster centroids determine the size of the quantized codebook. Compression ratio increases with the decrease in codebook size but with some trade-off in quality. A good quantization algorithm should achieve a balance between image quality and compression ratio. This proposed work delivers a near-lossless novel image compression method using differential coding and normalized vector quantization with good compression ratio.

2 Related Work

Vector quantization is a highly preferred technique to reduce the bit storage without much loss in image quality. Shanthi Rani et al. [3] described a method for image compression based on vector quantization using novel genetic algorithm for compressing medical images. Chiranjeevi et al. [4] proposed a vector quantization for image compression which uses Bat algorithm on the initial solution of LBG. Sumalatha and Subramanyam [5] proposed a hierarchical lossless image compression for telemedicine applications. The author has developed a method of adaptive multi-wavelet transform using multi-dimensional layered zero coding. Huiyan et al. [6] presented a medical image compression method based on vector quantization with variable block sizes in wavelet domain. The author has implemented a Local Fractal Dimension (LFD) which is used to analyze the local complexity of each wavelet coefficients and sub-band. Bhattacharyya et al. [7] proposed a Vector Quantization-based Image Compression method using Generalized Improved Fuzzy Clustering. The author has described the method as easy and fast with rapid convergence. Li et al. [8] developed an image compression method using transform-based vector quantization technique which has combined the features of transform coding and VQ. Abouali [9] proposed an object-based VQ for image compression. Chitra and Shanthi Rani [10] developed a novel compression approach based on Haar Wavelet and VQ. Phanprasit et al. [11] proposed a medical image compression method using vector quantization and system error compression for improving the bit rate and the image quality. Shanthi Rani et al. [12] presented a novel approach of vector quantization using modified particle swarm optimization algorithm which has generated an efficient codebook for compressing images. There are various existing methods [13–16] which employed VQ technique to create an efficient codebook and reduce the bit rate as well.

The organization of the paper is as follows: The detailed description of the proposed method is presented in Sect. 3. Results and comparative analysis of the proposed method is investigated in Sect. 4. Finally, the merits and future enhancement ideas are presented in Sect. 5.

3 Proposed Method

This section presents a novel approach of the proposed method using differential and vector quantization technique. Mainly, it has three phases in the encoding process. In the first phase, the input image is initialized and divided into non-overlapped blocks/vectors. Mean Differential Vector (MDV) has been created in the second phase for each vector by finding the difference of the mean value of each vector and each pixel from the original vector. Generally, the difference values are small in magnitude but have both positive and negative signs. As the sign will have more impact on the results, MDV is normalized to all positive values. Vector quantization is applied to the Normalized Mean Vector (NMV).

In the first phase, the input image is partitioned into non-overlapped nxn image blocks which are represented as a vector. Mean value is found out for each input vector and the MDV, i.e., difference between mean value and each pixel of the vector.

$$M_i = mean(v(i,:))$$

where "M_i" refers to the mean value of each input vector and "v_i" represents the ith input vector.

$$MDV(i,:) = M_i - v(i,:)$$

where MDV(i, :) denotes the ith mean differential vector which has both positive and negative value as it stores the difference. To preserve the sign, it is normalized to positive values by adding Maximum Absolute Difference (MAD) to each value in MDV as given in the following Equations.

$$MAD = abs(\max(MDV))$$

$$NMV(i,:) = MDV(i,:) + MAD$$

where NMV represents the Normalized Mean Vector. The normalized vector has been created for preserving the quality of image blocks. To improve the compression ratio, VQ is applied to NMV in the next phase. The codebook obtained in VQ process is further de-normalized by subtracting MAD from each code vector before generating the compressed stream so as to reduce the number of bits. Hence, the compressed stream will have de-normalized codebook, code indices, and compressed

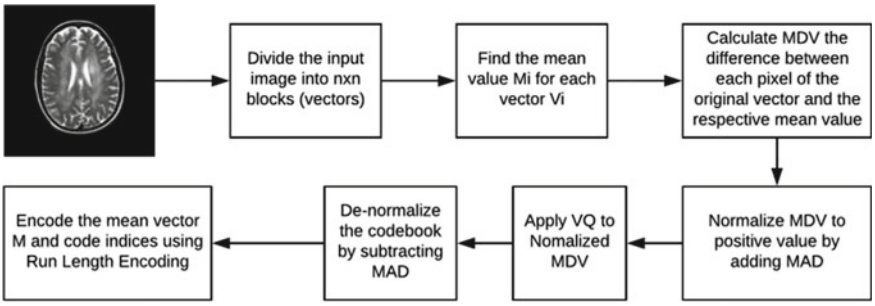


Fig. 1 Encoding process of the proposed method

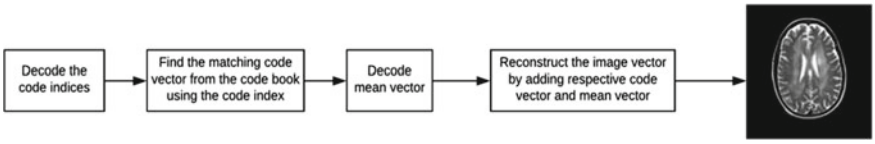


Fig. 2 Decoding process of the proposed method

mean vector. The mean vector and code indices are further compressed using run-length encoding.

The decoding process is the inverse of encoding process. The mean matrix and code indices are run-length decoded first and the matching code vector for each input vector is retrieved through its codebook index. The final image vector RI_i is reconstructed by adding the mean value to its respective reconstructed input vector R_i using the following Equation.

$$RI(i,:) = R(i,:) + M(i)$$

3.1 Proposed Flow Diagram

See Figs. 1 and 2.

4 Results and Discussions

The performance of the proposed method is briefly discussed in this section. The outcome of this proposed method shows better results than the other existing methods. The experimental results of the proposed method are analyzed for five medical test images such as MRI_Brain, MRI_Knee, MRI_Spine, Mammogram Image, and X-ray image. The efficiency of the proposed method is evaluated using standard

compression metrics like Peak Signal-to-Noise Ratio (PSNR), Compression Ratio (CR), Structured Similarity Metric (SSIM), and Bit Rate (BR) which have proved the superior performance of the proposed method in quality and compression aspects as well.

$$PSNR = 20 \log_{10} \frac{255}{\sqrt{MSE}}$$

$$Q = \frac{4\sigma_{xy}\mu_x\mu_y}{\left(\sigma_x^2\sigma_y^2\right)\left(\mu_x^2\mu_y^2\right)}$$

$$CR = \frac{Original\ Image\ Size}{Compressed\ Image\ Size}$$

$$Bpp = \frac{Original\ Image\ Size\ in\ Pixels}{Total\ number\ of\ bits\ in\ Compressed\ image}$$

Another big challenge of compression methods is the computation complexity. An effective compression method should strive to achieve good quality and compression ratio without much computation complexity.

The performance analysis of the proposed method in terms of three important metrics; quality, compression efficiency, and computation complexity is presented in Table 1. Experiments are also conducted to investigate the performance using other statistical moments median and mode other than mean and are listed in Table 1.

It is worth noting that the proposed method has achieved PSNR value in the considerable range (40–50) optimal for medical image diagnosis and SSIM closer to 1 exhibiting near-lossless quality achievement by the proposed method. Table 1 also reports good CR and BR which are vital parameters for assessing the compression efficiency of the proposed method. Moreover, the proposed method is also tested using various image block sizes. It has been observed that the block size $[2 \times 2]$ produces higher range of PSNR and CR as well. Table 2 investigates the comparative analysis between the proposed and existing methods for MR image.

It is obvious from Table 2 that the proposed method outperforms similar existing methods by achieving a PSNR value in the range (38–54).

Table 3 results show that the proposed method achieves higher PSNR range (38–40) over the existing CLC [11] and CLC + SEC [11] methods.

The results of the proposed method are visually represented for medical test images (Fig. 3).

Table 1 Performance of the proposed method with block size [2 × 2] and codebook size [128]

Image	Methods	Quality metrics		Compression efficiency		Computation time (In s)
		PSNR	SSIM	CR	BR	
MRI_Brain	Mean	49.61	0.99	6.37	1.25	9.34
	Median	48.76	0.98	6.43	1.24	25.54
	Mode	47.44	0.98	6.66	1.20	27.03
MRI_Knee	Mean	49.16	0.97	3.25	2.45	11.93
	Median	48.63	0.97	3.24	2.46	31.61
	Mode	47.11	0.96	3.23	2.47	32.40
MRI_Spine	Mean	59.24	0.99	3.02	2.64	15.03
	Median	59.13	0.99	3.00	2.66	53.44
	Mode	57.51	0.99	3.01	2.65	23.92
Mammogram	Mean	53.17	1.02	4.82	1.65	5.19
	Median	52.81	0.99	4.85	1.64	9.44
	Mode	51.27	0.98	5.10	1.56	11.59
X-ray	Mean	58.63	0.99	2.74	2.91	7.64
	Median	57.96	0.99	2.75	2.90	12.71
	Mode	56.53	0.99	2.75	2.90	13.96

Table 2 Comparative analysis of the existing and proposed method using medical images

Images 512 × 512 pixels	Algorithms	Bit rate (bpp)			
		0.19		0.25	
		PSNR	MSE	PSNR	MSE
MR cardiac	CLC [11]	26.78	136.48	28.05	101.88
	CLC + SEC [11]	35.10	20.10	35.27	19.32
	Proposed method	**54.53**	**0.21**	**54.87**	**0.12**
MRI skull	CLC [11]	28.19	98.65	28.82	85.33
	CLC + SEC [11]	35.80	17.10	36.08	16.04
	Proposed method	**38.14**	**10.68**	**38.87**	**9.23**
Ultrasound liver cyst	CLC [11]	25.01	205.15	26.37	150.00
	CLC + SEC [11]	30.54	57.42	31.07	50.83
	Proposed method	**38.85**	**9.70**	**39.08**	**7.32**

Table 3 Comparative analysis of the existing [11] and proposed method using nonmedical images

Images 512 × 512 pixels	Algorithms	Bit rate (bpp)			
		0.19		0.25	
		PSNR	MSE	PSNR	MSE
Airplane	CLC [11]	23.65	280.60	25.17	197.73
	CLC + SEC [11]	28.15	99.56	29.20	78.17
	Proposed method	38.01	10.28	39.02	8.13
Girl	CLC [11]	26.51	145.24	27.66	111.45
	CLC + SEC [11]	31.94	41.60	31.95	41.50
	Proposed method	39.31	7.61	40.34	6.01
Lena	CLC [11]	25.47	184.54	26.84	134.61
	CLC + SEC [11]	31.17	49.67	31.35	47.65
	Proposed method	38.37	9.44	39.35	7.54
Pepper	CLC [11]	22.89	334.26	25.54	181.59
	CLC + SEC [11]	28.84	84.93	29.61	71.14
	Proposed method	38.52	9.13	39.23	7.76

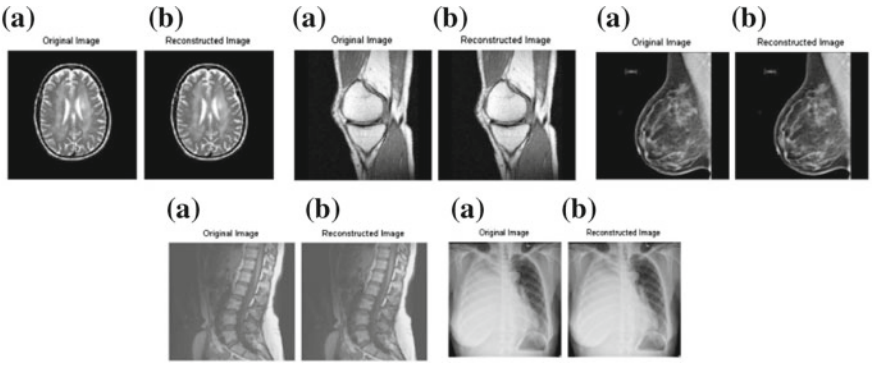


Fig. 3 **a** Original image. **b** Reconstructed image using the proposed method

5 Conclusion

The proposed method presents a novel approach of differential coding with normalized vector quantization technique. The proposed method achieves higher compression without compromising the quality of an original image. Furthermore, the proposed method has minimized the computation complexity. Hence, the proposed method is a good choice for medical image storage and transmission.

References

1. Somasundaram K, Shanthi Rani MM (2011) Novel K-means algorithm for compressing images. Int J Comput Appl 18(8):9–13
2. Somasundaram K, Shanthi Rani MM (2011) Mode based K-means algorithm with residual vector quantization for compressing images. In: International conference on "control, computation and information systems". Springer CCIS 140, pp 105–112
3. Shanthi Rani MM, Chitra P, Vijayalakshmi R Image compression based on vector quantization using novel genetic algorithm for compressing medical images. Int J Comput Eng Appl XII(I):104–114
4. Chiranjeevi K, Jena U (2016) Fast vector quantization using a Bat algorithm for image compression. Eng Sci Technol Int J 19:769–781
5. Sumalatha R, Subramanyam MV (2015) Hierarchical lossless image compression for telemedicine applications. In: Eleventh international multi-conference on information processing-2015 (IMCIP-2015). Procedia Comput Sci 54:838–848
6. Huiyan J et al (2012) Medical image compression based on vector quantization with variable block sizes in wavelet domain. Hindawi Publishing Corporation. Comput Intell Neurosci 2012:3, Article ID 541890
7. Bhattacharyya et al (2014) Vector quantization based image compression using generalized improved fuzzy clustering. In: 2014 international conference on control, instrumentation, energy & communication (CIEC):662–666
8. Robert Y, Li et al (2002) Image compression using transformed vector quantization. Image Vis Comput Elsevier 20:37–45
9. Abouali AH (2015) Object-based VQ for image compression, Elsevier. Ain Shams Eng J 6:211–216
10. Chitra P, Shanthi Rani MM (2018) Modified haar wavelet based method for compressing medical images. Int J Eng Techniq (IJET) 4(1)
11. Phanprasit Tanasak et al (2015) Medical image compression using vector quantization and system error compression. IEE J Trans Electr Electron Eng 10:554–566
12. Shanthi Rani MM, Chitra P, Mahalakshmi K (2017) A novel approach of vector quantization using modified particle swarm optimization algorithm for generating efficient codebook. Int J Advanc Res Comput Sci 8(9)
13. Shanthi Rani MM, Chitra P (2009) Adaptive classified pattern matching vector quantization approach for compressing images. In: 2009 international conference on image processing, computer vision & pattern recognition proceedings, Las Vegas, USA, pp 532–538
14. Shanthi Rani MM (2014) Residual vector quantization based iris image compression. Int J Comput Intell Stud 3(4):329–334, Inderscience Publishers

15. Shanthi Rani MM, Chitra P (2016) Novel hybrid method of haar-wavelet and residual vector quantization for compressing medical images. In: 2016 IEEE conference on advances in computer applications (ICACA), vol 1, pp 321–326
16. Shanthi Rani MM, Chitra P (2018) A hybrid medical image coding method based on haar wavelet transform and particle swarm optimization technique. Int J Pure Appl Math 118(8):3056–3067

Harmonization of White and Gray Matter Features in Diffusion Microarchitecture for Cross-Sectional Studies

Prasanna Parvathaneni, Shunxing Bao, Allison Hainline, Yuankai Huo, Kurt G. Schilling, Hakmook Kang, Owen Williams, Neil D. Woodward, Susan M. Resnick, David H. Zald, Ilwoo Lyu and Bennett A. Landman

Abstract Understanding of the specific processes involved in the development of brain microarchitecture and how these are altered by genetic, cognitive, or environmental factors is a key to more effective and efficient interventions. With the increasing number of publicly available neuroimaging databases, there is an opportunity to combine large-scale imaging studies to increase the power of statistical analyses to test common biological hypotheses. However, cross-study, cross-sectional analyses are confounded by inter-scanner variability that can cause both spatially and anatomically dependent signal aberrations. In particular, scanner-related differences in the diffusion-weighted magnetic resonance imaging (DW-MRI) signal are substantially different in tissue types like cortical/subcortical gray matter and white matter. Recent studies have shown effective harmonization using the ComBat technique (adopted from genomics) to address inter-site variability in white matter using diffusion tensor imaging (DTI) microstructure indices like fractional anisotropy (FA) or mean

P. Parvathaneni (✉) · Y. Huo · B. A. Landman
Electrical Engineering, Vanderbilt University, Nashville, TN, USA
e-mail: Prasanna.Parvathaneni@vanderbilt.edu

S. Bao · I. Lyu · B. A. Landman
Computer Science, Vanderbilt University, Nashville, TN, USA

K. G. Schilling · B. A. Landman
Vanderbilt University Institute of Imaging Science, Vanderbilt University, Nashville, TN, USA

A. Hainline · H. Kang
Biostatistics, Vanderbilt University, Nashville, TN, USA

N. D. Woodward · B. A. Landman
Department of Psychiatry and Behavioral Sciences, Vanderbilt University School of Medicine, Nashville, TN, USA

D. H. Zald
Department of Psychology and Psychiatry, Nashville, TN, USA

O. Williams · S. M. Resnick
National Institutes of Health, Bethesda, USA

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_3

diffusivity (MD). In this study, we propose (1) to apply the correction at voxel level using tract-based spatial statistics (TBSS) in FA, (2) to correct variability across scanners with different gradient strengths in DTI, and (3) to apply the ComBat technique to advanced DW-MRI models, i.e., neurite orientation dispersion and density imaging (NODDI), to correct for variability of orientation dispersion index (ODI) in gray matter using gray matter-based spatial statistics tool (GSBSS). We show that the biological variability with age is retained or improved while correcting for variability across scanners.

Keywords Harmonization · NODDI · Brain microstructure
Gray matter surface-based analysis

1 Introduction

Understanding the development of white matter microstructure is essential to understanding neurodevelopment. With the increasing number of publicly available neuroimaging databases, including the Adolescent Brain Cognitive Development (ABCD), the UK BioBank and Connectome studies, there is an opportunity to combine large-scale imaging resources to increase the power of statistical analyses to test the common biological hypothesis. Community-wide efforts are underway to address standardization of acquisitions and analyses for imaging biomarkers as described in imaging biomarker roadmap for cancer studies [1]. Using diffusion tensor imaging (DTI), ENIGMA's disease working groups have begun to analyze data across cohorts for differences in a range of DTI measures and discovering factors that consistently affect brain structure and function [2]. However, the community of clinicians, engineers, and physicists is not yet ready to agree on a single best practice approach to advanced DW-MRI. Therefore, it is imperative to understand when (and how) diverse protocols can be analyzed to enable comparison and optimization across protocols based on practical study design and imaging constraints.

Application of advanced DW-MRI methods in clinical research studies has been hindered by a lack of consensus on best protocols. DW-MRI suffers from between-scanner variation that hinders comparisons of images across imaging sites, scanners and over time, a model that can preserve biological variability and remove the unwanted variation introduced by the site is needed. Harmonization of data across scan protocols and site differences is an important preliminary step to conduct group analysis involving cross-sectional data acquired across different regions or on different scan protocols (Fig. 1). Using fractional anisotropy (FA) and mean diffusivity (MD) maps, Fortin et al. compared several harmonization approaches and found that ComBat performed best at inter-site variability correction [3]. ComBat has also been applied to harmonization of cortical thickness measures across scanners and sites [4] and multicenter radiomic studies with positron emission tomography [5]. With the introduction of advanced diffusion MRI models, e.g., the neurite orientation dispersion density imaging (NODDI), it is possible to analyze microstructural changes within gray matter (GM).

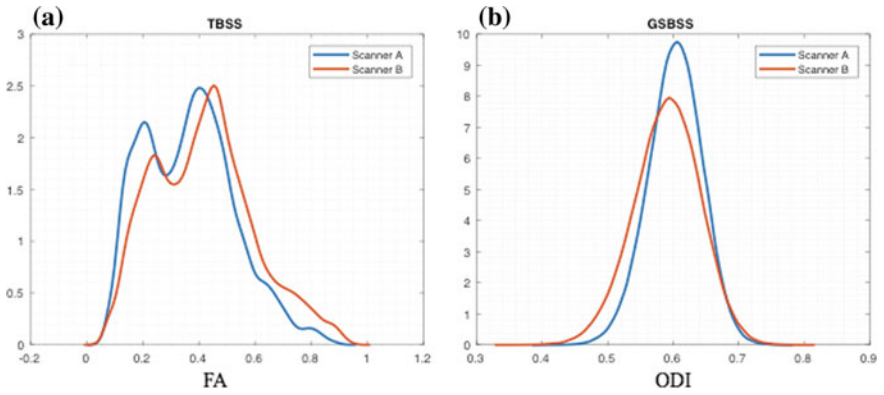


Fig. 1 Scanner differences are illustrated in **a** Study 2: FA in white matter using TBSS between scanners with 1.5T and 3T field strengths. **b** Study 3: ODI in gray matter using GSBSS across two distinct 3T scanners

In this study, we propose three different applications of ComBat for diffusion microstructure. First, we propose to extend the applicability of ComBat in white matter using the tract-based spatial statistics (TBSS) [6]. The number of voxels with significant scanner effects is reduced to zero voxels from 20,500 significant voxels before correction while retaining the age effects (37,508 voxels after correction vs. 37,711 before correction) after ComBat correction. Also with ComBat, both the voxel-wise variance and effect size were reduced by twofold. Second, we evaluate the performance of ComBat for correction of FA maps in white matter across two scanners with different gradient strengths of 1.5T and 3T using TBSS. The number of voxels with significant scanner effects was reduced to zero from 97117 voxels before correction while improving the number of significant voxels with negative effects of age from 71,126 to 72,267 voxels. Also, both the voxel-wise variance and effect size were reduced by fivefold after ComBat correction. Finally, we propose to use ComBat [3] in GM for correcting the scanner variability of orientation dispersion index (ODI) using GSBSS. Significant results of scanner effects are reduced to zero along with reduction in the vertex-wise ODI effect by ~twofold.

2 Methods

2.1 Data

This manuscript examines three independent datasets (noted as Study 1, Study 2, and Study 3). Study 1 included young adults (average age = 26.02 (male, n = 194), 26 (female, n = 223)) ranging from 23 to 34 years old with data acquired from 2 different 3T scanners one from each pair from The Tennessee Twin Study (TTS) (RDoC Constructs: Neural Substrates, Heritability, and Relation to Psychopathology). Study

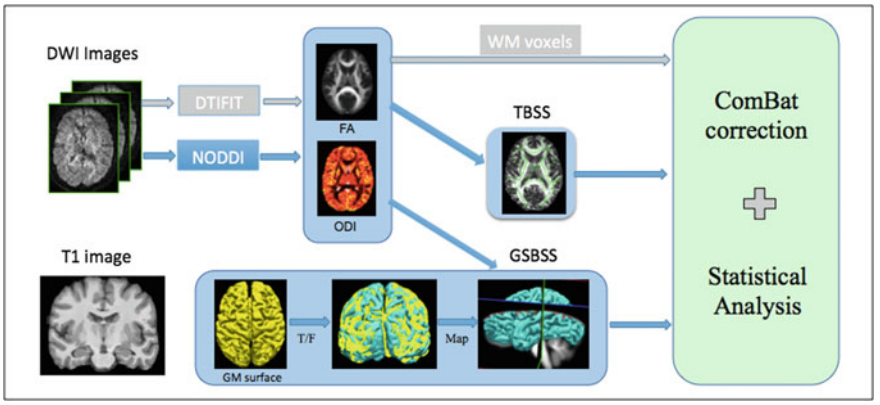


Fig. 2 DWI/HARDI images are processed with DTIFIT and NODDI model to extract FA and ODI. FA maps are then processed using TBSS. T1 image is used for generating GM surface which is transformed to target surface onto which ODI maps are mapped using GSBSS. These measures are passed into ComBat tool for variability correction for statistical analysis. Gray arrows indicate the previous application of ComBat to DTI data

1 has diffusion-weighted imaging scan protocol with b-values of 1000 s/mm^2 with 33 directions. Study 2 included 119 subjects from older adults population (average age = 78.29 (male, n = 45), 71.38 (female, n = 66)) acquired on 2 different scanners with gradient strengths of 1.5T and 3T with b-value 700 s/mm^2 with 30 and 32 gradient directions, respectively, from the Baltimore Longitudinal Study of Aging (BLSA) [7]. Study 3 included 30 healthy subjects (average age = 31.94 (male, n = 18), 35.83 (female, n = 12)) with data acquired from 2 different 3T scanners in an ongoing study of brain connectivity in neuropsychiatric disorders at Vanderbilt. For each subject in Study 3, we have: (1) a T1-weighted anatomical scan (MPRAGE, 1 mm isotropic resolution, TE = 2 ms, TR = 8.95 ms, and TI = 643 ms); (2) a diffusion-weighted imaging scan (2.5 mm isotropic resolution, FOV = 96 × 96, TR = 2.65 s, TE = 101 ms, Gmax = 37.5 mT/m) that included 2 diffusion shells with b-values of 1000 s/mm^2 (24 directions) and 2000 s/mm^2 (60 directions). Diffusion-weighted images (DWI) were affinely registered to b0 using FMRIB Software Library's (FSL 5.0) FLIRT for eddy current correction. The data was then processed using DTIFIT from FSL to obtain diffusion tensor that yields FA maps and other tensor-based metrics. Overall processing flow used in this approach is presented in Fig. 2.

2.2 TBSS and GSBSS Processing

For group analysis, the mean FA image was created for the following registration of individual FA maps to the FMRIB58 template [6]. This image was thinned to represent the mean FA skeleton. Individual FA maps were projected onto this common skeleton.

NODDI processing was carried out using the AMICO package which is a fast implementation of NODDI [8]. Orientation dispersion index (ODI) obtained from the above method was co-registered to the individual structural T1-weighted scan using SPM12 (http://www.fil.ion.ucl.ac.uk/spm). Cortical surfaces were obtained from the MaCRUISE pipeline [9] and initially deformed via nonrigid transformation obtained by template-based volume registration using ANTs SyN [10] registration. Then, surface registration (DSM) [11] was performed to establish cortical shape correspondence across cortical surfaces as mentioned in GSBSS [12]. ODI was mapped onto the individual gray matter (GM) surface and evaluated for cross-subject analysis.

2.3 Variability Correction and Statistical Analysis

FA skeletonized voxels from TBSS and ODI indices along the target surface were then corrected using ComBat as reformulated in the context of DTI images. Additive and multiplicative effects of the site were included in the formulation to account for scanner/site variability as shown in below equation from [3]. In our study, the outcome would be FA in white matter or ODI in GM with scanner effect.

$$y_{ijv}^{\text{ComBat}} = \frac{y_{ijv} - \alpha_v - X_{ij}\beta_v - \gamma_{iv}}{\delta_{iv}} + \alpha_v + X_{ij}\beta_v$$

A design matrix was formed with unpaired two-sample t-tests between the groups (Site A and B) adjusting for age. Permutation-based statistics were performed on skeletonized FA for each method using FSL randomize [13] (5,000 permutations). Significant results ($p < 0.05$) comprising threshold-free cluster environment (TFCE) [14] clustering were presented after accounting for multiple comparisons by controlling for family-wise error (FWE) rate. GM-based vertex-wise spatial statistics were calculated using the permutation analysis of linear models (PALM) [13] using 5,000 permutations followed by a generalized Pareto distribution to the tail of the approximation distribution [15]. p-values < 0.05 (FDR corrected) are considered to be significant.

3 Results and Conclusion

For Study 1 with two scanners having same field strength, Fig. 3 shows significant results obtained between scanner A and scanner B before (20,500 voxels) with an overlay of negative effects of age (37,711 voxels). As noted, the number of significant voxels between scanners is reduced to zero voxels while retaining the age effects (37,508 voxels) after ComBat correction.

For the scanners of different gradient strengths from Study 2, significant results of scanner effects before correction (97,117 voxels) were reduced to zero while

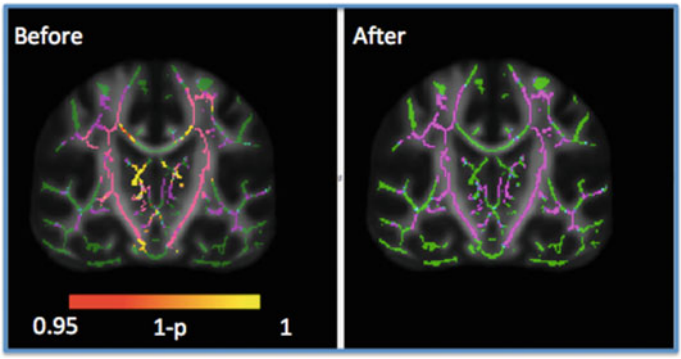


Fig. 3 Study 1: overlay of the significance of scanner effects (red-yellow) with age (magenta) with FWE. The left image shows before correction and right image shows after ComBat correction

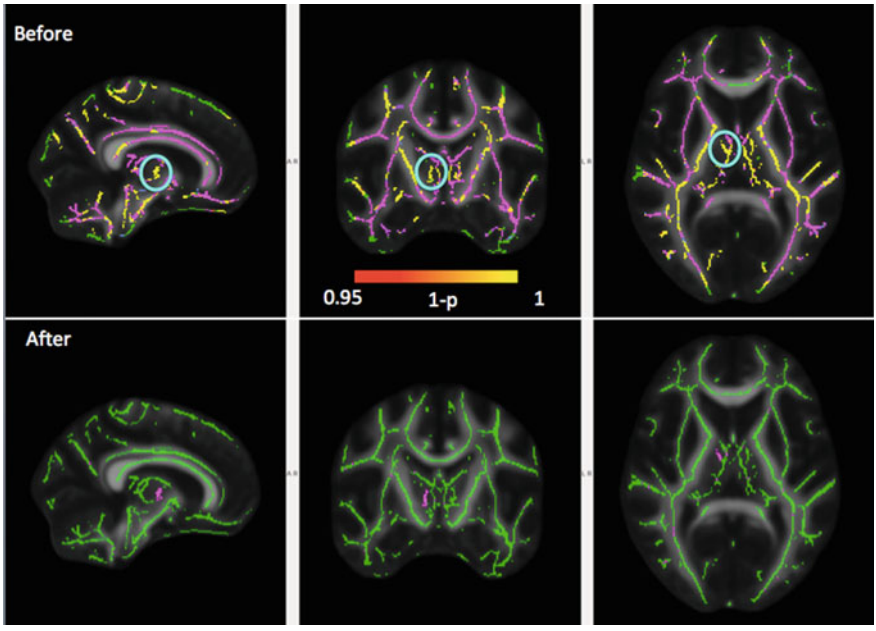


Fig. 4 Study 2: overlay of the significance of scanner effects (red-yellow) with age (magenta) with FWE. The top row shows before correction and lower row shows the *additional* significant effects of age with ComBat correction

improving the number of significant voxels with negative effects of age from 71,126 before correction to 72,267 voxels as shown in Fig. 4. Note that ComBat eliminates the scanner effect, while the age effect is increased.

Finally, in Study 3, scanner differences are shown for ODI in gray matter before and after correction in Fig. 5. Since the population in this study has a narrow distri-

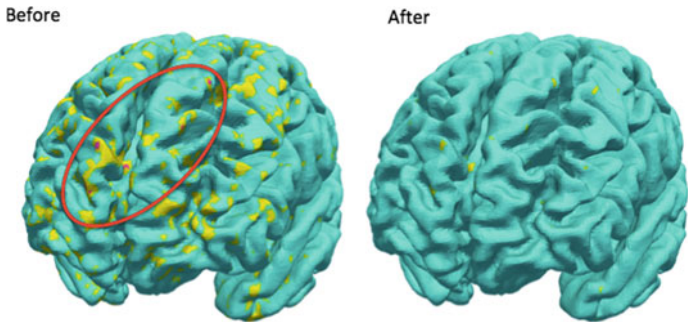


Fig. 5 Study 3: significance of scanner effects is shown in magenta before correction (179 vertices with FDR) overlaid on cluster of uncorrected p-values (yellow)

bution of age range, age-related effects analysis is not included for this study. After ComBat correction for ODI, there is no significant scanner effect.

Figure 6 illustrates the effect size as the difference in mean values of the parameter of interest. Figure 6a shows the effect size of FA values in WM before and after correction from Study 1. Figure 6b shows the effect size of FA values in WM from scanner A and B with field strengths 1.5T and 3T from Study 2. Figure 6c shows the ODI effect size which is shown as the difference in mean ODI values between the two scanners in gray matter from Study 3 with two different 3T scanners. Study 1: prior to correction, there was a ~0.02 FA effect between the two and 3T scanners (blue), while with ComBat, both the voxel-wise variance and size (red) were reduced by twofold. (b) Study 2: prior to correction, there was a ~0.05 FA effect between the 1.5T and 3T scanners (blue), while with ComBat, both the voxel-wise variance and size (red) were reduced by ~fivefold. (c) Study 3: prior to correction, there was a ~0.2 ODI effect between the two 3T scanners (blue), while with ComBat, vertex-wise ODI effect is dropped to ~0.1 with the corresponding reduction in variance. In all the cases, we see the reduction in mean difference between the two groups after applying ComBat. In both cases, we see the reduction in mean difference between the two groups after correction. Mean FA and ODI values between scanner A and scanner B after applying variability correction are shown in Fig. 7.

3.1 Conclusion

Harmonization of WM and GM indices from DW-MRI models using ComBat are shown in (1) Study 1 by applying ComBat to TBSS for correcting FA values in DTI for two different scanners with field strength of 3T. (2) Study 2 characterizing at field strength differences showing the positive effects of ComBat correction while improving the age effect. (3) Study 3 showing the applicability with advanced DW-MRI models like NODDI in GM. In the first two studies, the scanner effect

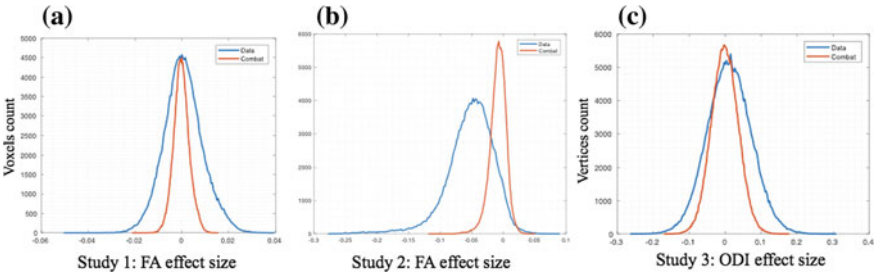


Fig. 6 Effect size values for before and after correction are shown for **a** Study 1 with FA effect size, **b** Study 2 with FA effect size, **c** Study 3 with ODI effect size

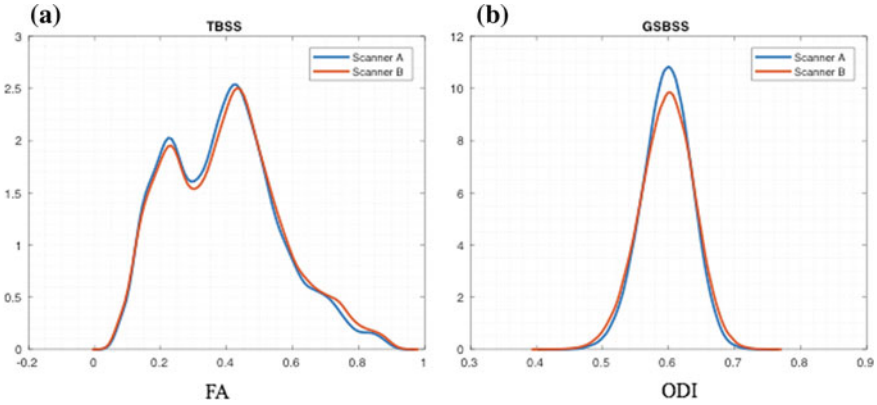


Fig. 7 Mean values between scanner A and scanner B are shown after applying variability correction in **a** FA in white matter using TBSS between 1.5T and 3T gradient strengths. **b** ODI in gray matter using GSBSS within 3T scanners

is corrected while retaining the biological variability with age. Detailed analysis in this domain could provide an opportunity to combine large-scale imaging studies to increase the power of statistical analyses to test the common biological hypothesis of understanding the development of brain microarchitecture.

While we have shown the possibility of application of ComBat technique in three different scenarios, a detailed analysis is needed to interpret the results in a meaningful way. In the first study where we have shown the applicability of the technique to TBSS, further analysis showing the comparison with the original proposed method as baseline is needed, if this approach indeed improves the statistical power.

Acknowledgements This work was supported in part by the Intramural Research Program, National Institute on Aging, NIH and by NIH R01EB017230 & R01MH102266 & Grant UL1 RR024975-01 & Grant 2 UL1 TR000445-06 & Grant UL1 RR024975-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank the participants of the studies.

References

1. O'connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brady M, Brown G (2017) Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 14(3):169
2. Thompson PM, Andreassen OA, Arias-Vasquez A, Bearden CE, Boedhoe PS, Brouwer RM, Buckner RL, Buitelaar JK, Bulayeva KB, Cannon DM (2017) ENIGMA and the individual: predicting factors that affect the brain in 35 countries worldwide. Neuroimage 145:389–408
3. Fortin J-P, Parker D, Tunc B, Watanabe T, Elliott MA, Ruparel K, Roalf DR, Satterthwaite TD, Gur RC, Gur RE (2017) Harmonization of multi-site diffusion tensor imaging data. Neuroimage 161:149–170
4. Fortin J, Cullen N, Sheline Y, Taylor W, Aselcioglu I, Cook P, Adams P, Cooper C, Fava M, McGrath P (2017) Harmonization of cortical thickness measurements across scanners and sites. NeuroImage 167:104–120
5. Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, Soussan M, Frouin F, Frouin V, Buvat I (2018) A post-reconstruction harmonization method for multicenter radiomic studies in PET. J Nuclear Med jnumed. 117.199935
6. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TE (2006) Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage 31(4):1487–1505
7. Shock NW (1984) Normal human aging: the Baltimore longitudinal study of aging
8. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC (2012) NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 61(4):1000–1016
9. Huo Y, Carass A, Resnick SM, Pham DL, Prince JL, Landman BA (2016) Combining multi-atlas segmentation with brain surface estimation. In: Book combining multi-atlas segmentation with brain surface estimation, NIH Public Access
10. Avants BB, Epstein CL, Grossman M, Gee JC (2008) Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12(1):26–41
11. Lombaert H, Sporring J, Siddiqi K (2013) Diffeomorphic spectral matching of cortical surfaces. Inf Process Med Imaging 23:376–389
12. Parvathaneni P, Rogers BP, Huo Y, Schilling KG, Hainline AE, Anderson AW, Woodward ND, Landman BA (2017) Gray matter surface based spatial statistics (GS-BSS) in diffusion microstructure. In: Book gray matter surface based spatial statistics (GS-BSS) in diffusion microstructure. Springer, pp 638–646
13. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014) Permutation inference for the general linear model. Neuroimage 92:381–397
14. Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44(1):83–98
15. Winkler AM, Ridgway GR, Douaud G, Nichols TE, Smith SM (2016) Faster permutation inference in brain imaging. NeuroImage 141:502–516

Deep Neural Architecture for Localization and Tracking of Surgical Tools in Cataract Surgery

Neha Banerjee, Rachana Sathish and Debdoot Sheet

Abstract Over the last couple of decades, the quality of surgical interventions has improved owing to the use of computer vision and robotic assistance. One such application of computer vision, namely, detection of surgical tools in videos is gaining attention of the medical image processing community. The main motivation for detection, localization, and annotation of surgical tools is to develop applications for surgical wsorkflow analysis. Such an analysis can aid in report generation, real-time decision support, etc. Cataract surgery is one of the common surgical procedure where surgeons do have direct visual access to the surgical site. Extremely small tools are used for this procedure and the surgeons observe the surgical site through a surgical microscope. In such cases, detecting the presence of tools can act an additional aid to the surgeon as well as other surgical staffs. We propose a framework consisting of a Convolutional Neural Network (CNN) which learns to distinguish and detect the presence of various surgical tools by learning robust features from the frames of a surgical video. Various deep neural architectures are hence evaluated for the task of detecting tools. The baseline models used for the purpose are pretrained on Imagenet dataset and they render upto 50% prediction accuracy. All the experiments have been validated on the dataset released as part of the Cataracts Grand Challenge. A framework for localization and detection of tools has also been proposed, which is capable of extracting visual features from glimpses of an image, by adaptively selecting and processing only the selected regions at high resolution.

Keywords Cataract surgery · Multiple tool detection · CNN · Deep neural architectures · Class imbalance · Glimpse network

N. Banerjee · R. Sathish (✉) · D. Sheet
Indian Institute of Technology, Kharagpur, Kharagpur, India
e-mail: rachana.sathish@iitkgp.ac.in

N. Banerjee
e-mail: neha.ban2014@gmail.com

D. Sheet
e-mail: debdoot@ee.iitkgp.ac.in

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_4

1 Introduction

Surgical tools have been tracked and their usage recorded by placing markers on the tools or by using additional sensor like magnetic trackers or accelerometers. These additional systems tend to make the surgical tools bulkier and often hampers their free use by the surgeon. Since surgical videos are inherently available during the procedures, vision-based methods offer an alternative for tool tracking and surgical workflow modeling without altering the natural working ambience of the surgeon. The high accuracy of CNN on classification tasks can be leveraged for detection and tracking of objects in video through transfer learning [2]. Twinanda et al. [6] proposed to detect tool presence and surgical phase recognition in surgical videos using CNN which does not require any additional hardware. This method however does not consider the temporal relationship between the various surgical tasks thus leading to a limitation in detecting the tools. Al Hajj et al. [1] proposed a solution based on CNNs for automatic detection of surgical tools which analyzes sequences of consecutive images and the features extracted from each image is fused using the optical flow.

Since the various procedures involved in a cataract surgery are performed using multiple tools in each frame of the surgical video, the task is to detect multiple objects. The challenge in detecting multiple objects in an image lies in the fact that different categories of objects can be located at different positions at different scales and postures. Further due to an increase in label space due to multi-label classification, it requires more number of training samples. Inherently, CNN has a huge number of parameters and thus requires a huge number of training samples which is typically scarce in the domain of medical images. Moreover, a multi-label dataset essentially increases the chances of class imbalance which poses a problem in training of conventional deep learning networks.

2 Problem Statement

The objective at hand, which is detection of presence of multiple tools in the surgical video of a cataracts surgery can be formulated as a multi-label multi-class classification task. Let $T = \{t_1, t_2, \ldots, t_n\}$ represent the set of tools used during the course of the surgery. For each frame f of the video, a subset of tools t_f out of T are to detected which are present in the frame. Single-label image classification assigns one label to an image whereas in multi-label multi-class problem, each image is assigned multiple labels.

3 Methodology

Owing to the success of CNN in single-label classification tasks, a similar approach is used to solve the multi-label classification. There are a total of 21 tools across the videos. Up to three tools were found to be used simultaneously. An additional "No-tool" class was added in the dataset to incorporate those instances which had no tool in them. Thus, finally, there were 22 classes during the classification task. Some baseline models were tested. For each of the models, the last fully connected layer was fine-tuned to contain 22 units corresponding to the classification task. The loss function used for training of the baselines is the *MultiLabelSoftMarginLoss*. Thus for a set of M images, the loss function which is to be minimized is defined as

$$L(\hat{\mathbf{y}}, \mathbf{y}) = \frac{-1}{M} \sum_{i=1}^{M} \left(\mathbf{y}(i) log \left[\frac{exp(\hat{\mathbf{y}}(i))}{1 + exp(\hat{\mathbf{y}}(i))} \right] + (1 - \mathbf{y}(i)) log \left[\frac{1}{1 + exp(\hat{\mathbf{y}}(i))} \right] \right)$$

where $\hat{\mathbf{y}}$ is the prediction and $\mathbf{y}$ is the ground-truth.

The proposed framework consists of two stages.

- **Stage I** The images from the dataset are resized to $3 \times 224 \times 224$ and are used to train a tool counter network, which predicts the number of tools the image may have. This count further guides the number of glimpses that needs to be extracted for identifying the tool type. This stage solves the problem of localization of tools. The tool counter is implemented using ResNet-18 [3].
- **Stage II**, The activation maps from Stage I are expected to contain some hot spots that localize the tool. Using these hot spots in the activation map, three smaller regions of interest of size 50×50 pixels are further considered. These are now used to train a new CNN which takes $9 \times 50 \times 50$ dimensional tensors as inputs and predicts the tool type among the given 22 classes. Three regions are considered

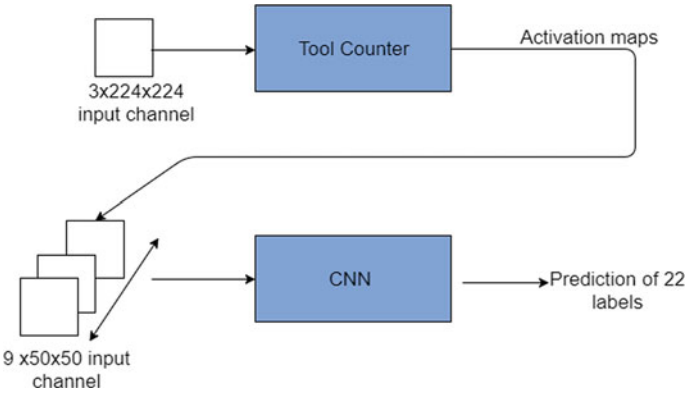


Fig. 1 Building block of proposed model conceptualizing the idea of a glimpse network

as that is the maximum number of tools that could be present in any frame of the video in the dataset used.

Figure 1 shows an overview of the proposed framework.

4 Experiment and Results

4.1 Dataset Description

All baselines and the proposed method are validated on the dataset released as part of the cataracts grand challenge[1]

- The dataset consists of 50 videos of cataract surgeries. The average duration of 10 min and 56 s.
- The tool usage in each videos has been annotated independently by two experts. A tool was considered to be in use whenever it was in contact with the eyeball.
- Training and Test Sets: The dataset was divided into a training set (25 videos) and a test set (25 videos). Division was made in such a way that (1) each tool appears in the same number of videos from both subsets (plus or minus one) and (2) the test set only contains videos from surgeries performed by the renowned expert. The test set was non-annotated.
- After sampling, each image in a video was resized to a tensor of dimension $3 \times 224 \times 224$ representing three channels of inputs. These were the inputs to the deep learning models considered.

4.2 Data Augmentation and Lass Balancing

It was observed that various surgical activities have different duration. Since each task is associated with a particular set of tools, certain tools appeared more frequently across all videos. Due to varying duration of various surgical tasks, certain tools occur more frequently across all the videos resulting in severe class imbalance in the dataset. Therefore, twofold manipulations were required in the dataset:

- Preparing the training set in a way such that each of the 22 classes has an almost equal number of samples.
- Ensuring sufficient data based on complexity of the network architecture used.

The annotations for tool presence in the video is provided at 1 fps. However, the videos are recorded at 30 fps. Therefore, the imbalance in class distribution can be balanced by interpolating to the existing annotation to intermediate frames and

[1] https://cataracts.grand-challenge.org/home/.

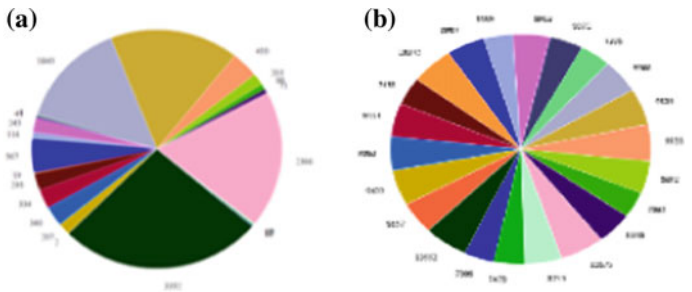


Fig. 2 Tool frequency distribution before and after balancing. **a** Indicates distribution in the raw dataset and **b** indicates distribution in the training dataset

adding them to the training instances. For example, if there are 300 frames in the training set annotated as to contain scissors, it can be increased to 3000 instances by sampling the frames at 10 frames per second (fps), instead of 1.

Moreover, for those tools whose occurrences were too small, data augmentation was performed by horizontal flips, vertical flips and their combination when required to balance them to a particular desired number (say, 10,000 occurrences for each tool). Also, since tools had chances of co-occurrence, it was essential to find co-location matrices for each video, such that sampling could be performed only across the diagonal elements.

Moreover, the deep neural architectures need a huge number of weights to be trained, and they exhibit satisfactory results with sufficient data only. However, loading the entire dataset without considering any sampling technique would not be appropriate as it would require higher computational demands (Fig. 2).

4.3 Evaluation and Performance Metrics

Performance of baselines and proposed methodology is evaluated based on the area under the receiver-operating characteristic (ROC) curve (AUC). It is obtained by varying a cutoff on the confidence levels. The mean AUC value over all tools is also computed. Area under the ROC curve is used, rather than figures of merit based on precision and recall, since a varying number of tools may be used in each frame (zero, one, two, or three). Since, some tools were absent in the test dataset, their auc values were inevitably zero. Therefore, these are ignored and average AUC across tools are computed which are present in both training as well as test sets.

4.4 Training of Baselines

The following baseline architectures were tested using a data set of 25 videos.

- **BL1: AlexNet** The Alexnet [4] architecture contains five convolutional layers and three fully connected layers. Dropout is used in first two fully connected layers without which network shows overfitting when trained with ImageNet. The pooling layers used in AlexNet are used for summarization of outputs of the nearby groups in same kernel map.
- **BL2: VGGNet** The VGGNet 16 [5] has a depth of 16 layers. It considers Maxpooling and ReLU activation layers. Rectified Linear Unit (ReLU) is used to speeds up training process. ReLU has a property that it does not get saturated if the input is not normalized. When some training example produce positive output in ReLU, learning will take place in that neuron.
- **BL3: ResNet-18/50/152** In a ResNet, instead of hoping each few stacked layers directly, the layers are placed using a residual mapping. The under lying mapping is denoted by H(z), the stacked non linear layers fit a mapping $G(z) := H(z) - z$ So $H(z) = G(z) + z$. The formulation of $G(z) + z$ can be realized by a feed forward neural network with shortcut connections, which are connections which can skip one or more layers. They simply perform identity mapping and their outputs are added to that of stacked layers. The network depths considered are 18, 50, and 152 respectively.

All baselines and the proposed model have been trained using the same combination of training and validation sets. The split considered was 17 videos in training set, 5 videos in validation, and 3 videos in test set. Each model has been trained for 100 epochs. All the baseline models are pretrained on the Imagenet[2] dataset (which contains 1.2 million images with 1000 categories). The networks are trained using Stochastic Gradient Descent with momentum 0.9 and weight decay 0.00001. Since the dataset suffered from problem of class imbalance, the loss function was modified by multiplying weights to each class, in inverse proportion of their prevalence in the dataset.

$$w_i = \sqrt{\frac{\max\left\{f_j | 1 \leq j \leq 21\right\}}{f_i}}$$

where fi represents the frequency that the class i appears in the training dataset.

As Stochastic Gradient Descent is used for the training procedure, there could be a possibility for the optimization problem to get stuck at a local minima due to high error rate at the vicinity of a minima. Therefore, for appropriate convergence, the learning rate was decreased by 1/10th of its previous value if the training loss differed by less than 0.0001 throughout the last 10 epochs on an average. The results obtained are shown in Table 1.

[2] http://www.image-net.org/.

Table 1 Performance of baselines

Baseline	Performance score (mean AUC)
ResNet-18	0.65
ResNet-50	0.68
ResNet-152	0.64
Alexnet	0.58
VGGNet	0.59

Table 2 Performance of proposed method

Network	Performance score
Tool counter	0.84 (average accuracy)
Proposed CNN	0.82 (mean AUC)

4.5 Performance of Proposed Method

The tool counter implemented with a ResNet18 to localize the tools present in each image. It was multi-class classification problem having four mutually exclusive classes. (Since either no tool, or a single tool or 2 or 3 tools.) Due to a higher frequency of no-tool class in the raw dataset, proper balancing had to be done to train the network without bias. Finally, the hot spots obtained from the network were used to train a CNN, whose output was a 22×1 vector (each entry being the probability of presence of the tool in each patch). The results obtained are summarized as follows (Table 2).

5 Discussion

When a deep neural network is trained on a dataset with unequal occurrence of individual classes, it tends to get biased to the classes with highest occurrence. This imbalance in the dataset is predominantly present in the case of surgical videos where certain tools are used more frequently than the other. Thus, an appropriate balancing and augmentation of the raw dataset was an important part of the experiment. Unlike other machine learning tasks, where cross validation methods are generally considered to reduce training bias, this experiment required inclusion of all 21 tools in the training set for proper prediction. For example, videos 14, 19 were the only videos where more than three tools were exclusively present. Hence, they had to be present in the training set. The main cause of this problem was that some tools were exclusively present only in some particular videos. Thus, a suitable combination of the training, validation, and test sets had to be chosen.

6 Conclusion

This paper proposes a deep learning based framework for tool detection and tracking. At present, the performance of the baseline and proposed frameworks have been tested off-line using prerecorded videos. Integration of the obtained results in an operating room for online evaluation shall be the next step for verification of performance measures. The results of tool detection can also be used to predict the time after which the next surgical phase is likely to start, i.e., to design the workflow modeling of a surgery. This would allow the surgical staff to plan and prepare ahead of time the tools required for the next phase. Integration of these frameworks to operating rooms can help in developing smart context-aware operating rooms of the future.

References

1. Al Hajj H, Lamard M, Charrière K, Cochener B, Quellec G (2017) Surgical tool detection in cataract surgery videos through multi-image fusion inside a convolutional neural network. In: 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 2002–2005
2. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
3. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
4. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
5. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
6. Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N (2017) Endonet: a deep architecture for recognition tasks on laparoscopic videos. IEEE Trans Med Imag 36(1):86–97

Efficient Segmentation of Medical Images Using Dilated Residual Networks

Lokeswara Rao Bonta and N. Uday Kiran

Abstract Medical image segmentation is an essential part in many medical applications such as automatic measurement of tumour size, volume of organs and content-based image retrieval, etc. Various fully convolutional architectures have been proposed in the literature to tackle this problem. Lack of generalization is one of the main challenge in the field of medical imaging and all existing fully convolutional architectures involve huge number of parameters which make them prone to overfit the data. In this study, we proposed an efficient convolutional architecture called Dilated Residual Network (DRN) for medical image segmentation. By the design of DRN architecture, we have reduced number of parameters involved drastically, making the architecture less prone to overfitting hence by improving the generalization ability. We demonstrate that DRN outperforms the previous state of the art architecture called U-Net in medical image segmentation on various datasets in terms of training time, Dice score and Jaccard score. The source code (based on Keras with Tensorflow as the background) of the DRN and sample train and test image results are available at https://github.com/LokeshBonta/Dilated-Residual-Networks.

Keywords Medical image segmentation · Fully convolutional neural networks Dilated convolutions

1 Introduction and Related Work

Deep convolutional neural networks have been exhibiting better results by outperforming the traditional state of the art algorithms present in the literature for many tasks in computer vision such as object classification, localization, detection and segmentation [1, 3, 6, 7]. The typical use of convolutional networks is on classification

L. R. Bonta (✉) · N. Uday Kiran
Sri Sathya Sai Institute of Higher Learning, Anantapur 515134, Andhra Pradesh, India
e-mail: lrbonta.sssihl@gmail.com

N. Uday Kiran
e-mail: nudaykiran@sssihl.edu.in

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_5

problems [3, 6], in which the output is going to be a single class label. However, in many computer vision tasks, particularly in biomedical computer vision, the output desired must consider localization, i.e. assigning label to the particular pixel where the object present in the image and segmentation, i.e. assigning class label to each and every pixel in the image (foreground/background). So, various fully convolutional architectures [5, 7] have been proposed to solve this problems using convolutions and deconvolutions/transposed convolutions to integrate precise pixel level details with wider, global and context information.

However, the design of these architectures will increase the number of parameters exponentially as we increase the depth of the convolutional architecture. These convolutional networks also progressively reduce the resolution as we go deeper in the network in which the spatial structure would become no longer perceptible. Such a loss in spatial visual acuity can limit segmentation accuracy in applications that use detailed image understanding. To solve this problem, Yu and Koltun [8] proposed dilated convolutions for segmentation. The main applications of dilated convolution are in the situations where the output size of the network is similar to input size such as image localization, segmentation, super resolution, etc.

In simple terms, **dilated convolution** operation is a normal convolution applied to input with definite gaps, dilation rate or factor. Given an input image, dilation factor "d = 1" is normal convolution, d = 2 means skipping one pixel and so on. In general, d = n means skipping n − 1 pixels.

Receptive field is defined as the implicit region captured on actual input (at first layer) by every input unit to its next layer. Dilation convolution increases the global view (receptive field size) of the CNN exponentially while maintaining the linearity in the number of parameters involved. Figure 1 shows exponential growth in the receptive field for the dilation rate 1, 2 and 4, respectively. With this advantage, it

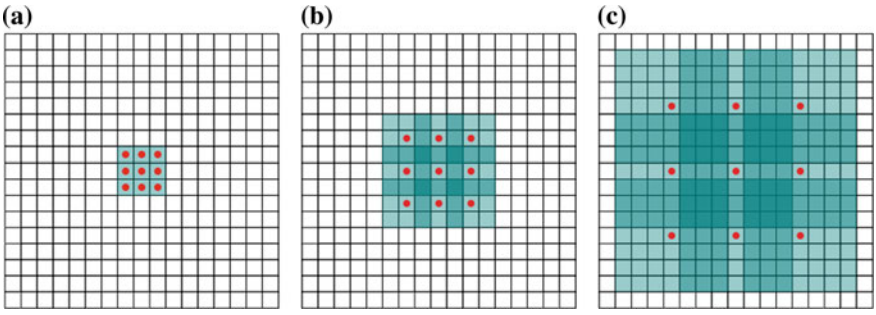


Fig. 1 Expansion of receptive view exponentially using dilation without any loss of coverage or resolution. **a** F1 is produced from F0 by a 1-dilated convolution; each element in F1 has a receptive field of 3 × 3. **b** F2 is produced from F1 by a 2-dilated convolution; each element in F2 has a receptive field of 7 × 7. **c** F3 is produced from F2 by a 4-dilated convolution; each element in F3 has a receptive field of 15 × 15. The receptive field size increases exponentially while the number of parameters increases linearly [8]

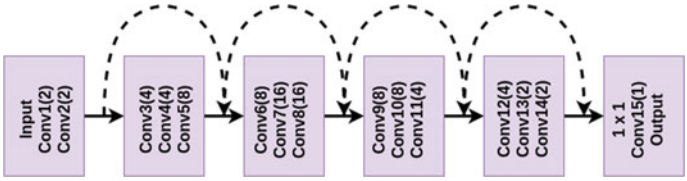


Fig. 2 Proposed dilated residual network (DRN), dilation factor is shown in parenthesis

finds usage in applications which care more about combining cognition of global context with less cost.

Dilated convolution has three main advantages

- Detection of finer details by processing inputs in higher resolutions.
- Broader receptive view captures more global or contextual information.
- Faster run time with less parameters.

2 Proposed Architecture

In this work, we propose a new architecture called "**DRN—Dilated Residual Network**", see Fig. 2, this architecture is very uniform and elegant as the number of channels in all the intermediate convolutions is 16. This architecture has only 41 thousand parameters. This architecture is inspired from the work done in Yu and Koltun [8] for segmentation using dilated networks and residual blocks from He et al. [1] to avoid vanishing gradient problem in deeper networks.

3 Datasets Used

SSSIHMS—Kidney Dataset

We have used kidney image dataset from Sri Sathya Sai Institute of Higher Medical Sciences (SSSIHMS), Prasanthigram. Renal dataset consists of Computed Tomography (CT) images in native format. These are then converted to JPEG format for the training purposes. This dataset is collected by us under the direct supervision of professional radiologist Dr. Gurudatt.[1] This data is*** collected with all legal agreements and protocols signed. The dataset is comprising of 1500 images and task is to segment kidney region in the images. The radiology experts have manually created ground truth images for all the images.

[1]Radiologist, SSSIHMS, Prasathigram, e-mail: s.gurudatt@gmail.com.

SunnyBrooke—Cardiac Dataset [4]
As part of MICCAI 2009s 3D Segmentation in the Clinic: A Grand Challenge, Sunnybrook Health Science Centre, Toronto, Canada presented the Cardiac MR Left Ventricle Segmentation Challenge from short-axis MRI of cardiac data. The Sunnybrook dataset comprises a variety of pathologies and ventricular shapes of cine MRI from 45 patients, or cases, having a blend of cardiac conditions: healthy, hypertrophy, heart failure with infarction, and heart failure without infarction. Expert manual segmentation contours for the endocardium, epicardium and papillary muscles are provided for basal through apical slices at both End-Diastole (ED) and End-Systole (ES) phases. This dataset was made available as part of the MICCAI 2009 challenge on automated left ventricle (LV) segmentation from short-axis cardiac MRI. The Sunnybrook dataset is available through the Cardiac Atlas Project with a public domain license.

Ultrasound Nerve Data-Kaggle [2]
In this work, we have also use dataset from Kaggle's ultrasound nerve segmentation challenge. The task in this contest is to segment a group of nerves known as the Brachial Plexus (BP) in ultrasound images. We were provided with a large amount of rich training set of images in which the BP has been manually annotated by human experts. Annotators were trained by experts and instructed to annotate images where they felt certain about the existence of the BP position. The datasets contain a mix of images where BP is present and not present. The dataset has 5635 training images and ground truth and 5508 test images of size 420×580.

4 Experiments, Results and Discussion

4.1 Metrics Used

Although accuracy is the simplest, it is not a good way to measure the quality in medical imaging because of class imbalance problem. So, we use more sophisticated metrics such as Dice score and Jaccard score. The results which were empirically observed in medical image segmentation domain (for example in Tran [7]) using Dice score and Jaccard score are much better than the ones computed through traditional losses like multinomial logistic loss or softmax loss.

Let G and P be set of ground truth (manual) pixels and predicted (automated) pixels for minority classes (foreground pixels) respectively. Then,

DSC—Dice Score Coefficient of two sets is defined as

$$D(G, P) = 2\frac{G \cap P}{G + P} = \frac{2TP}{2TP + FN + FP} \tag{1}$$

The Dice index is always in [0, 1], 1 when there is perfect match and zero when total mismatch happens.

Jaccard Coefficient is defined as

$$J(G, P) = \frac{G \cap P}{G \cup P} = \frac{G \cap P}{G + P - (G \cap P)} = \frac{TP}{TP + FN + FP} \tag{2}$$

The Jaccard index also varies from zero to one. One when perfect match with the ground truth and zero if there is total mismatch.

4.2 Experimental Framework and Implementation Details

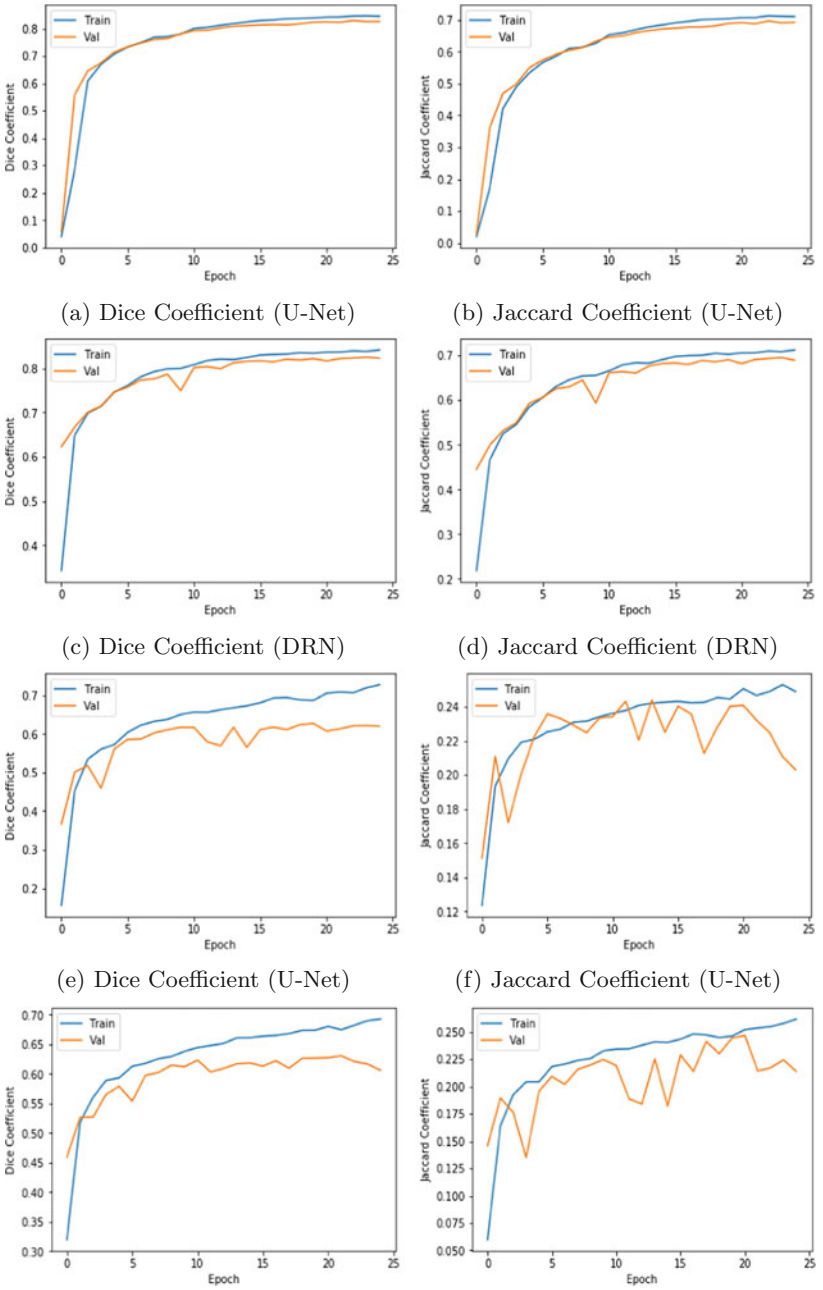
We have trained our networks on NVIDIA TITAN X GTX GeForce GPU with 3072 cores, clock rate of 1000 MHz and with a 12 GB memory. We have used Keras with Tensorflow background as our deep learning framework because of its prevalence in deep learning research community.

Working with full size of training data in the datasets incurred memory and computation issues. To avoid these issues, we have resized all the images to 192×192. In hyper-parameter tuning, we have used a batch size of 32 with learning rate 0.0001 in case of U-Net and 0.001 in DRN case. We have used Adam as the optimizer for training our networks and Dice loss (negative of Dice Score) as the loss function and calculated Dice and Jaccard score after every epoch. We trained our architectures on 80% of the training dataset and 20% remaining dataset for validation. We compared the results with respect to U-Net and DRN for 25 epochs. In case of DRN experiments, we have used he-normal initialization for the parameters.

4.3 Results and Discussion

The Dice score and the Jaccard score value for Ultrasound nerve dataset and SSSIHMS kidney dataset is plotted against number of epochs in Fig. 3. Where the blue coloured curve represents the results with respect to training data and orange colour curve is for validation data. We have depicted similar kind of results for Sunnybrooke left ventricle cardiac dataset in Fig. 4. In the both cases, we have done the analysis with the U-Net architecture and the proposed Dilated Residual Network. The summary of the comparative study of U-Net and DRN architectures using different datasets is illustrated in Table 1. We have compared these networks using the parameters such as time taken per epoch, number of learning parameters involved, Dice score and Jaccard score.

From the results aforementioned, we observed that the proposed DRN outperforms previous state-of-the-art architecture U-Net by all means. DRN involves only 41 thousand parameters whereas U-Net has about 7.7 million parameters which makes the U-Net architecture more prone to the overfitting. Because the number of learning parameters is less in case of DRN, it has taken less time per epoch while



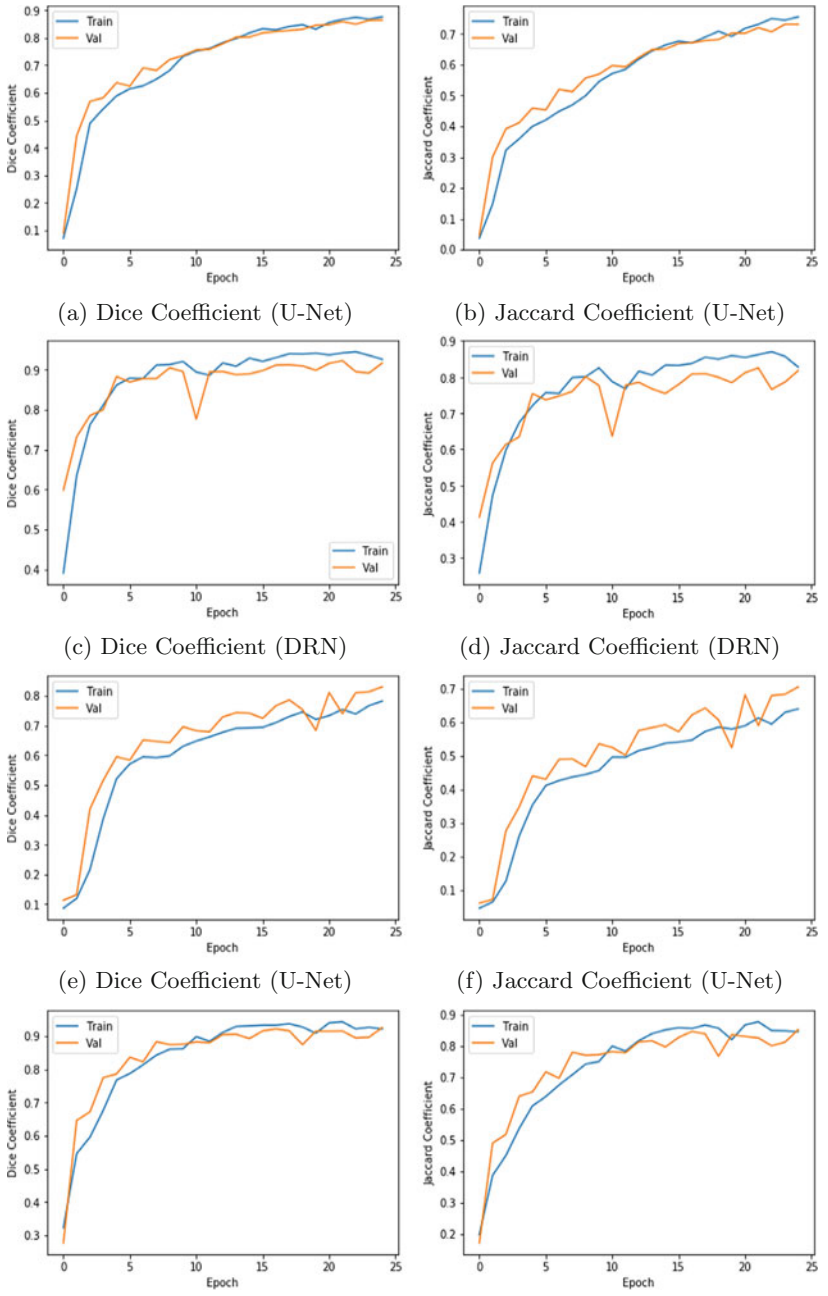
(a) Dice Coefficient (U-Net) (b) Jaccard Coefficient (U-Net)

(c) Dice Coefficient (DRN) (d) Jaccard Coefficient (DRN)

(e) Dice Coefficient (U-Net) (f) Jaccard Coefficient (U-Net)

(g) Dice Coefficient (DRN) (h) Jaccard Coefficient (DRN)

Fig. 3 Results on SSSIHMS—Kidney dataset and Kaggle's Ultrasound Nerve data using U-Net and DRN. **a**, **b**, **c**, **d** for SSSIHMS Renal dataset and **e**, **f**, **g**, **h** are for ultrasound nerve dataset



(a) Dice Coefficient (U-Net) (b) Jaccard Coefficient (U-Net)

(c) Dice Coefficient (DRN) (d) Jaccard Coefficient (DRN)

(e) Dice Coefficient (U-Net) (f) Jaccard Coefficient (U-Net)

(g) Dice Coefficient (DRN) (h) Jaccard Coefficient (DRN)

Fig. 4 Results on left ventricle segmentation using U-Net and DRN **a**, **b**, **c**, **d** are for endocardium (inner cardium) and **d**, **e**, **f**, **g** are for epicardium (outer cardium)

Table 1 Comparison of DRN and U-Net architectures on 25 epochs

Model	Dataset	Time/Epoch (s)	Parameters	Dice score	Jaccard score
U-Net	Kaggle's ultrasound nerve dataset	75	7.7 M	0.68	0.25
	SSSIHMS Kidney dataset	35		0.77	0.70
	Sunnybrooke cardiac (Endo)	12		0.82	0.76
	Sunnybrooke cardiac (Epi)	6		0.88	0.72
DRN	Ultrasound nerve	75	41 K	0.70	0.29
	Kidney	28		0.83	0.72
	Cardiac (Endo)	8		0.94	0.87
	Cardiac (Epi)	4		0.92	0.85

training. We have also observed there is an improvement in both in Dice score and Jaccard score for DRN over U-Net in almost all the cases which makes DRN an ideal fully convolutional architecture in case of medical image segmentation problem. We have also observed that adding any further layers in U-Net architecture will lead in exponential increase in the number of parameters because of the design of U-Net. Whereas DRN can be extended further because the number of parameters increase linearly.

5 Conclusions and Future Work

Extending networks in "conv–deconv" fully convolutional way increases number of parameters exponentially and makes the model overfit the dataset and training process slow. So, we proposed an efficient architecture called Dilated Residual Networks. The advantage of this network is twofold. First, it has very less parameters compared to the other networks which makes it less likely to overfit. Second, it outperformed the existing architectures in terms of both accuracy and speed. We would like to extend DRN to 3D medical image segmentation problem. We believe that DRN has a capability to work well even in the other computer vision domains like object localization and detection not only in medical imagery but also in natural imagery as well. Integrating proposed models into 3D slicer, thereby helping the open source community.

References

1. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
2. Kaggle (2015) Ultrasound nerve segmentation data
3. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
4. Radau P, Lu Y, Connelly K, Paul G, Dick A, Wright G (2009) Evaluation framework for algorithms segmenting short axis cardiac MRI. MIDAS J-Card MR Left Ventricle Segm Chall 49
5. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
6. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
7. Tran PV (2016) A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv:1604.00494
8. Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. arXiv:1511.07122

Non-rigid Registration of Brain MR Images for Image Guided Neurosurgery Using Cloud Computing

D. Preetha Evangeline and P. Anandhakumar

Abstract Image guided Surgery (IGS) has become familiar on enhancing it with computerized technologies which makes processing much easier. These computerized systems are highly sophisticated and improvised which orients the surgeon to perform the surgery with three dimensional images of the patients MR images. This work presents the Non Rigid Registration of Brain MRI with the help of Cloud computing. There exist geometric ambiguities among individuals due to the varying shape of the cortex and this is a challenging task when it comes to registration process. Intensity based registration is not alone enough to tackle the issues hence anatomical knowledge is highly important. Segmentation and labelling of cortex sulci with non-parametric approach that enables capturing of its shape and topology. The intensity and feature points are matched using the registration energy. A linear combination of compact smooth linear field and branched basic function is formed. The addition of the sulci from the upper border till its bottom part makes the registration process even more efficient. Cloud computing has been used in two scenarios, Firstly, the complete workflow of registration is enabled at the cloud to handle hypothetical computations and obtain precise results. Secondly, storage and computational nodes are used to handle large amount number of images nearly 7TB. The Deformable registration is evaluated and its feasibility has been analysed using cloud resources which resulted in timely execution of complex components during registration. The result obtained has proved that Cloud Computing provides practical and cost effective support for telemedicine while performing Image guided Surgery. The accuracy level has been improved upto 62% while performing Non Rigid Registration.

D. Preetha Evangeline (✉)
Vellore Institute of Technology, Vellore, Tamil Nadu, India
e-mail: Preethaevangeline.d@vit.ac.in

P. Anandhakumar
Anna University, Chennai, Chennai, Tamil Nadu, India
e-mail: anandh@annauniv.edu

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_6

1 Introduction

Image Guided Surgery has become popular in recent days and used in various ranges of Telemedicine applications [1]. Deformable registration is one of the key technologies for enabling image-guided procedures during surgery. Imaging data are obtained from various sources across time points and these images are integrated into a procedural workflow for Image Guided Surgery (IGS). Image Guided Neurosurgery is one of the vital applications in the field of telemedicine. Non-Rigid Registration leads to accurate removal of tumours. It enables patients to regain from unhealthy state by masking the pre-processing procedures over the intraprocedural imaging. A clinical Deformable Registration must include factors such as speed of the process, preciseness and robustness. The registration should be carried out within the speculated time period in coordination with the procedural workflow constraints which is highly important to send timely reports to the surgeons. The results of this deformable registration should not be liable to factors influencing the quality of images or imaging artifacts. The registration realistically reflects the physical warping of the tissue. Recently, ITK filters are used for Physics-based Non-Rigid Registration (PbNRR) [2] which has acceptable registration accuracy.

Regardless of the improvements obtained by PbNRR, there can be still more improvement in the registration accuracy, and minimal execution time is vital. Existing methods [3] suggests that varying the values of the input parameters used for registration may result in higher accuracy but the problem here is finding an optimal value of these parameters is at high risk. There is not any concordance projecting the exact value for the properties of live tissue [4]. Similarly, while registering block matching components, proper metrics which matches accurately with the intraoperative procedures have to be followed which is quite difficult [5]. The accuracy of the registration is affected due to the block matching parameters such as block size and block window, respectively. The properties of the images are very important as the optimum value of the block size and block window must be obtained from it. The registration problem is viewed as a computational procedure, which aims at improving the accuracy rate. Here, several PbNRR instances are fed with different input parameters and metrics to obtain the maximum registration accuracy [6]. In the conventionalities of NRR, there is an important time gap between the procurement of the initial intraoperative scan and the MRI. This conventional time gap is used to identify the similarity metrics and compare them. This comparison of the similarity metrics required high-speed computing environments. This is where the large-scale computing is needed.

This research uses large-scale computing environment to find out optimal values of the input parameters for improvising the accuracy rate. Selecting optimal parameters still exist as one of the challenges, which involves speculative execution as the situation may become uncertain. Iterations are performed with multiple registrations trying out for different parameter settings of the same input images. Variations in the obtained results show the sensitivity of certain metrics which is needed for assur-

ing the accuracy of the registration. This speculative execution requires large-scale computation environment where storage and computational nodes are mandatory.

Cloud computing is one of the hot topics for the past few years among researchers because of its enormous availability of resources, on the fly computation and scalability. Cloud computing is used for large-scale intensive and data sensitive applications [7] among which medical image processing and telemedicine are some of the critical applications where time and computational accuracy is a must. The Image Guided Neurosurgery uses the storage and computational resources from the cloud to provision computational platform for Physical-based Non-Rigid Registration.

The organization of the paper is as follows: Sect. 2 reviews about the role of cloud computing played in the field of medical image processing, Sect. 3 describes about the Non-Rigid Registration in detail, Sect. 4 explains about the speculative executions and its various constraints, Sect. 5 presents the results of the physical-based Non-Rigid Registration and its experimental set up, Sect. 6 concludes the contribution of the paper and its future works.

2 Literature Review on Cloud-Based Medical Image Processing

MammoGrid [10] and BIRN [11] Biomedical Informatics Research Network used grid computing for medical image processing where the lack of central management exists and the application is made to run on loosely connected resources. TeraGrid [3] had few limitations where the resources were not enough to handle the computations of NRR involved in MIP. References [12, 13] explains about the generalized grid used for image processing but the drawback here was the absence of overall resource maintenance and management which made it difficult to handle the processing and one more issue was the non-reservation of resources for on the spot computation accuracy.

In the intraoperative scenario, the timing and the special constraints play a major role and the criticality of the application is unpredictable. Few other methodologies [14–22] specified solutions using grid computing, but the issues and challenges related to network latency and batch scheduler proved that grid cannot be suited for Image Guided Neurosurgery. Cloud computing [7] can sufficiently provide resources required for efficient Medical Image Processing. Figure 2 illustrates cloud computing being employed for NRR in Image Guided Surgery. MRI images are classified at two levels, (i) Prior operative images and (ii) Endooperative images. These images are stored in the cloud and processed. Non-Rigid Registration is performed in the speculative mode and results with higher accuracy are obtained which is forwarded to the surgeon performing Image Guided Surgery (IGS) [23]. The successive registrations are performed using the parameters that provided maximum accuracy.

Deployment of telemedicine applications and processing them using cloud provides several advantages. Cloud provisions resources on demand, where reservation

of resources is not required. The processing is made easy and cost of infrastructure set up is comparatively low where resources can be rented based on the demand. Scaling of resources becomes possible and is adaptable based on the situations for pre- and post-surgery procedures.

3 Methodology

The methodology of Non-rigid registration especially the intensity-based technique have proven its efficiency in the case of large volumetric images. The proposed intensity-based registration algorithm used the following registration energy.

$$E(H, S) = E_s(\mathrm{A}, \mathrm{B}, \mathrm{H}) + \sigma\|\mathrm{H} - \mathrm{S}\|^2 + \sigma\lambda E_r(\mathrm{S}) \quad (1)$$

where A and B are the images that are used for registration, H and S are the Non-parametric transformations,

H Pair of homogeneous points identified as per the similarity measures

S Smooth estimation of NRR

E_{S} Intensity similarity energy

E_{r} Physical model

When two different brains are considered for NRR, the intensity constraint is not enough anymore especially when the cortex part is considered and the inexactness issue is a major challenge when the sulci have to be matched with the given sulcus. When anatomical structures are involved, the intensity-based similarity measures are not much efficient. To overcome these challenges, geometric features are combined along with the intensity matching. The labelled segmentations at the border and bottom layer of the sulci are known to be geometric features.

3.1 Extraction of Sulci

The initial anatomical extraction from the brain MRI is the sulci. The steps involved in segmenting the sulci are (i) Non-uniform partial correction, (ii) Separation of white matter (Chronic conditions), (iii) Skeletonization based on thinning technique and (iv) Surface splitting. Figure 1 shows the extracted one-dimensional feature of the sulci from brain MRI. The surfaces are then labelled based on neural network which used the information such as the depth, size, the target space, etc., of the sulci, which is compared with the minimum distance and the relational number of neighbours to the other sulcus. One-dimensional images are extracted illustrating the sulci bottom which defines the depth and the sulci border which defines the boundary

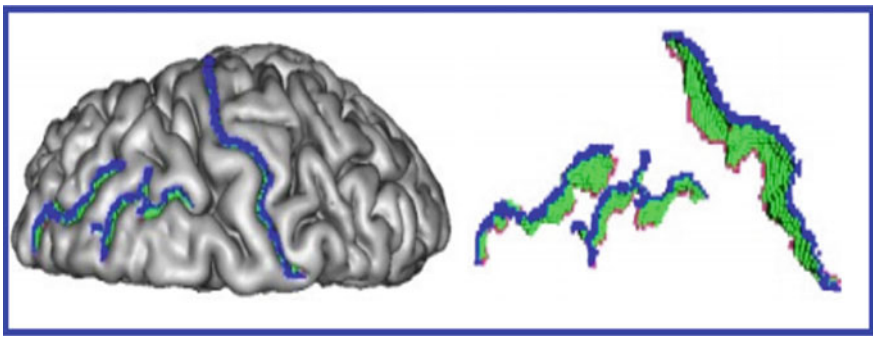


Fig. 1 Region in Green is the sulci, the Blue region is the sulci border and the Red region is the sulci bottom

of the brain. A sulcus has been known for its complex topology structure due to its cortical folds and shallow creases. Hence results in similar labelling of cortical parts. The bottom line of the sulci are included for geometric considerations. These lines are considered to be more precise anatomical landmarks. The other parts of the sulci are hard to match across individuals because of the delimitations. The local extents of the folds impact border localization of the sulci which greatly varies with individuals hence matching becomes difficult for other parts of the sulci. The reason behind considering one-dimensional images is the reduced computation complexity when processing the algorithm as the number of feature points will be drastically reduced. The sulci border helps in matching the entire structure of the sulcus in the case of intensity-based computations.

Where Label 'L', $1 \leq L \leq 90$, $S_L(A)$ be the set of points of image A, which has been distinguished as the feature points belonging to the sulci. $S_L^{\downarrow}(A)$ be the subset of feature points of $S_L(A)$ which makes up the sulcul bottom, similarly $S_L^{\uparrow}(A)$ makes up the sulcul border.

$$S(A) = \cup_L S_L^{\downarrow}(A) \cup S_L^{\uparrow}(A) \tag{2}$$

3.2 NRR Using Combined Geometric Features and Intensity-Based Algorithm

(i) Brain MRI images A and B are said to have imperfect registration of sulci $S_L^{\downarrow}(A)$ and $S_L^{\uparrow}(A)$ because both may have varying topologies, different folds and creases. There are possibilities that a part of sulci can belong to one part of the brain and the rest of the part may belong to the other brain, the accuracy of achieving the exact match will require additional folds which results in discontinuities during deformation of the cortex. Hence the best solution is to get the features closer rather than mapping the points one to one.

(ii) There occurs some errors due to labelization, hence $S_L^{\downarrow}(A)$ may not match with $S_L^{\downarrow}(B)$.

Inorder to inculcate the sulcus matching in the proposed algorithm, the generalization of registration energy given in Eq. (1), takes place by introducing conformities H_2, in-between the points of the two images which seems to be located with the same label 'L'.

$$\forall x \in S_L^{\downarrow}(A),\ H_2(x) \in S_L^{\downarrow}(B) \tag{3}$$

$$\mathrm{E}(\mathrm{H}_1, \mathrm{H}_2, \mathrm{S}) = \mathrm{E}_{\mathrm{sim}}(\mathrm{A}, \mathrm{B}, \mathrm{H}_1) + \sigma\|\mathrm{H}_1 - \mathrm{S}\|^2 + \sigma\lambda\|\mathrm{H}_2 - \mathrm{S}\|^2 + \sigma\lambda E_{reg}(S) \tag{4}$$

(iii) On minimizing $\mathrm{E}_{\mathrm{sim}}(\mathrm{A}, \mathrm{B}, \mathrm{H1}) + \sigma \cdot \|\mathrm{H1} - \mathrm{T}\|2$ w.r.t H1, i.e. and finding dense conformities H1 between voxels based upon the intensity information. 2. Next is to minimize $\|\mathrm{H2} - \mathrm{S}\|2$ w.r.t. H2, i.e. and then find conformity points H2 between sulci with the same label closest to S with a closest point algorithm.

(iv) Fit S to H1 and H2 by minimizing $\|\mathrm{H1} - \mathrm{S}\|2 + \gamma \cdot \|\mathrm{H2} - \mathrm{S}\|2 + \lambda \cdot \mathrm{Ereg(S)}$ w.r.t to S. S is thus of the form $\mathrm{S(x)} = \alpha \mathrm{Z} * \mathrm{H}_1(\mathrm{t}) + \sum \mathrm{ti} \in \mathrm{S(A)}\alpha \mathrm{iZ(t - ti)}$

where Z is the response of the impulse associated with $\mathrm{E}_{\mathrm{reg}}$.

(v) Repeat step 1 until convergence.

The final estimate transformation is then a weighted average of the fitting of intensity and feature points

$$\mathrm{S(t)} = \lambda \mathrm{Z} * \mathrm{H1(t)} + (1 - \lambda) \cdot \sum \mathrm{i}\, \alpha \mathrm{i} \cdot [\mathrm{H2(ti)} - \mathrm{S(ti)}] \cdot \mathrm{Z(t - ti)}2 / \sum \mathrm{i}\, \mathrm{Z(t - ti)} \tag{5}$$

where λ is considered as the co-efficient between feature based and intensity matching.

Anisotropic diffusion was applied to images priory and bias was removed carefully. The whole registration process was carried out in a multiresolution scheme which prevents it from local minima.

4 Experimental Setup

A set of five labelled brains were used for multisubject registration, out of which one brain was selected to be the reference brain and other brains were registered to it using affine registration. The experiments were carried out with the proposed algorithm with and without sulcus matching. Results were obtained for the proposed algorithm and the comparison has been illustrated in Figs. 2 and 3.

The sulcus matching has been carried out without interrupting the smoothness of the transformation. Some negotiable occlusion problem occurs due to affine transformation. Figure 4 depicts the quality of the transformation for four registrations.

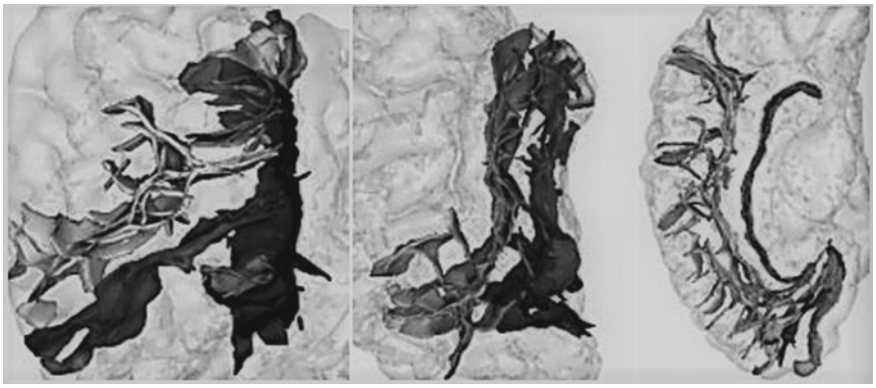


Fig. 2 Position of the sulci after NRR (without sulcus matching)

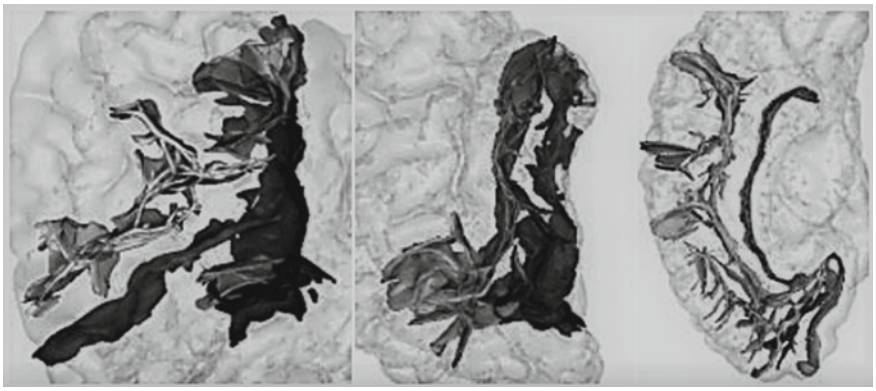


Fig. 3 Position of the sulci after NRR (with sulcus matching)

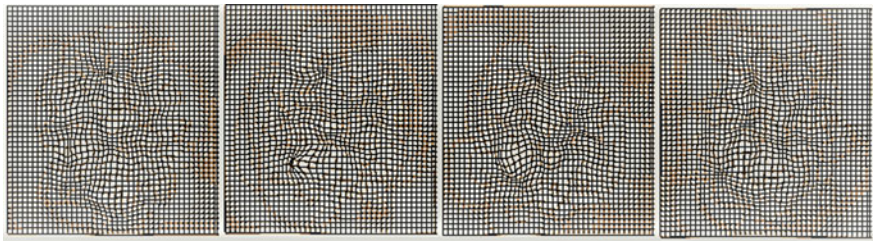


Fig. 4 Quality of transformations during registration of 4 multisubject images

Practically high computational environment is needed for performing registration. Approximately it takes up to 12 h to compute 100 k registration points for sequential block matching in an Intel Xeon 3.7 GHz processor. When there is a need for considering more than three similarity metrics, for the process of rejecting the outlier in

the solver and to evaluate the accuracy for each NRR will result in maximum time delay which may take days to complete.

Private cloud has been set up with 100 Virtual instances where four servers were used to accommodate the VM's. Each server consists of two CPU sockets and 82.4 GHz per socket. The overall memory needed to run the application was 116 GB RAM where each VM used 4 GB RAM. Memory deduplication technique was used to reduce the consumption of physical memory. Jobs of NRR were submitted parallel using a Directed Acyclic Graph in Hadoop Spark Cluster.

Microsoft Azure has been used for providing a platform to run the registration application. This Azure platform works as a High-Performance Computing platform where the application manager manages and monitors the application, the proxy nodes are the VM instances which balance the load among the computational nodes. The Head node assigns the virtual machines to run the incoming jobs. Clusters were set up having 3, 6, 12 and 24 computational nodes. The overhead of running the jobs in the clusters in evaluated. The number of computational nodes and storage nodes are varied while executing the NRR. The overhead of the jobs is calculated by obtaining two values, (i) Runtime (T1) of NRR and (ii) The total Execution time (T2), the time calculated from the submission of the job by the head node till its completion time. This overhead time includes both the queuing time and the time taken to obtain the output.

The use of Cloud Resources is used in two operations, (i) speculative execution of intraoperative data and (ii) time-critical computations and suboptimal parameter settings for NRR. During surgery, the computational nodes are used to run PbNRR with a parameter to obtain optimal results with accuracy and correctness. Performance evaluation has been carried out on a private cloud set up at MaaS Laboratory, MIT Campus, Anna University. OpenStack private cloud was set up for infrastructure as a service for providing computational and storage node. Windows Azure was used as a platform for running the processing application with a number of virtual instances.

5 Results and Discussion

The process of sulci surface labelling was carried out with a set of 90 labels. The error rate of the proposed algorithm was 14% due to ambiguity of the labelization process on some part of the cortex. The image datasets were collected from public BRATS Dataset, a multimodal Brain Tumour segmentation challenges and Surgical Planning Laboratory. The registration accuracy of NRR has been shown in Table 1.

Figure 5 shows the graph for Average Overhead Time for various number of tasks. The experiment was carried out using virtual machines with configuration of 8cores, 14 GB RAM and 2 TB of Hard disk. A cluster consists of 30 computational nodes for the execution of NRR.

Table 1 Evaluation of registration error

Case	Affine registration	NRR without sulcus extraction	NRR with sulcus extraction	% of improvement in accuracy
1	24.880	21.019	5.4569	71.875
2	9.210	4.490	2.2161	52.333
3	9.333	5.325	2.2161	58.423
4	9.665	7.020	2.2161	68.036
5	6.608	4.223	2.2161	44.765
6	8.162	3.505	2.2161	38.972
7	11.583	7.010	2.7884	60.594
8	14.346	9.849	2.9000	70.846
Average				62.986

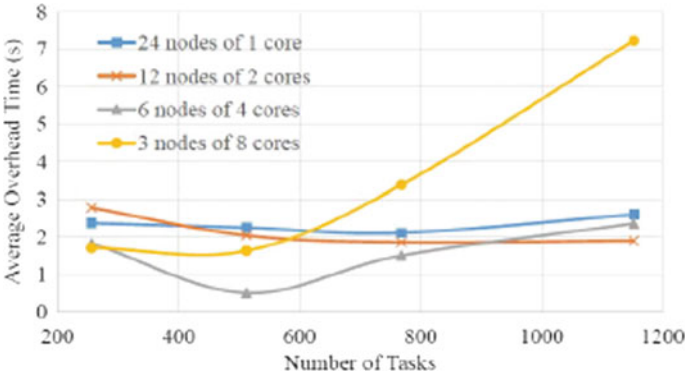


Fig. 5 Overhead Time for various tasks

6 Conclusion

This research work has combined intensity point based and geometric features to arrive at a registration energy which is used for extracting the sulcul bottom and border of the brain. The extraction is done by a non-parametric way and then it is labelled for all parts of folds and creases. The information that is used for registration process leads to a smooth transformation without any deformity caused. The extension part of this work will consider sulcus root part which corresponds as a connectivity to the foetal brain that will lead to accurate matching pattern and the folds which remains as a challenge in the current work.

References

1. Jolesz FA (2014) Intraoperative imaging and image-guided therapy. Springer, Berlin
2. Liu Y. Kot A, Drakopoulos F, Yao C, Fedorov A, Enquobahrie A, Clatz O, Chrisochoides NP (2014) An ITK implementation of a physics-based non-rigid registration method for brain deformation in image-guided neurosurgery. Front Neuroinformat 8(33)
3. Fedorov A, Clifford B, Warfield S, Kikinis R, Chrisochoides N (2009) Non-rigid registration for image-guided neurosurgery on the teragrid: a case study. Technical report WM-CS-2009-05, Department of Computer Science
4. Delingette H, Ayache N (2004) Soft tissue modeling for surgery simulation, vol XII of hand book of numerical analysis: special volume: computational models for the human body, 1st edn. Elsevier, Netherlands, pp 453–550
5. Skerl D, Likar B, Pernuˇs F (2008) A protocol for evaluation of similarity measures for nonrigid registration. Med Image Anal 12(1):42–54
6. Ino F, Kawasaki Y, Tashiro T, Nakajima Y, Sato Y, Tamura S, Hagihara K (2006) A parallel implementation of 2-d/3-d image registration for computer-assisted surgery. Int J Bioinformat Res Appl 2(4):341–357
7. Gannon D, Fay D, Green D, Takeda K, Yi W (2014) Science in the cloud: lessons from three years of research projects on microsoft azure. In: Proceedings of the 5th ACM workshop on scientific cloud computing, Science Cloud '14. ACM, New York, pp 1–8
8. Mell PM, Grance T (2011) Sp 800–145. The NIST definition of cloud computing. Technical report, Gaithersburg, MD, United States
9. Foster I (2002) What is the grid? A three point checklist, white paper. http://wwwfp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf
10. MammoGrid Project (2008) http://www.cems.uwe.ac.uk/cccs/project.php?name=mammogrid
11. Biomedical Informatics Research Network (2008). http://www.nbirn.net
12. Dong S, Insley J, Karonis NT, Papka ME, Binns J, Karniadakis G (2006) Simulating and visualizing the human arterial system on the TeraGrid. Future Gener Comput Syst 22:1011–1017
13. Manos S, Zasada S, Mazzeo MD, Haines R, Doctors G, Brew S, Pinning R, Brooke J, Coveney PV (2008) Patient specific whole cerebral blood flow simulation: a future role in surgical treatment for neurovascular pathologies. In: Proceedings of Teragrid'08
14. Archip N, Clatz O, Whalen S, Kacher D, Fedorov A, Kot A, Chrisochoides N, Jolesz F, Golby A, Black PM, Warfield SK (2007) Non-rigid alignment of pre-operative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery. Neuroimage 35:609–624
15. Chrisochoides N, Fedorov A, Kot A, Archip N, Clatz O, Kikinis R, Warfield SK (2006) Toward real-time image guided neurosurgery using distributed and grid computing. In: Proceedings of the ACM/IEEE Conference on Supercomputing
16. Majumdar A, Birnbaum A, Choi DJ, Trivedi A, Warfield SK, Baldridge K, Krysl P (2005) A dynamic data driven grid system for intra-operative image guided neurosurgery. Proc ICCS 2005:672–679
17. Clatz O, Delingette H, Talos IF, Golby AJ, Kikinis R, Jolesz FA, Ayache N, Warfield SK (2005) Robust non-rigid registration to capture brain shift from intra-operative MRI. IEEE Tran Med Imag 24(11):1417–1427
18. Foteinos P, Chrisochoides N (2014) High quality real-time Imageto-Mesh conversion for finite element simulations. J. Parallel Distrib Comput 74(2):2123–2140
19. Liu Y, Foteinos P, Chernikov A, Chrisochoides N (2012) Mesh deformation-based multi-tissue mesh generation for brain images. J Eng Comput 29(4):305–318
20. Drakopoulos F, Foteinos P, Liu Y, Chrisochoides NP (2014) Toward a real time multi-tissue adaptive physics-based nonrigid registration framework for brain tumor resection. Front Neuroinformat 8(11)
21. Vangel M, Fedorov A, W. W. III, Tempany C (2012) Statistical framework for characterization of deformable registration performance. Tech. report, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

22. Commandeur F, Velut J, Acosta O (2011) A VTK algorithm for the computation of the hausdorff distance. VTK J
23. Microsoft (2013) "Microsoft azure for research overview". http://research.microsoft.com/enus/projects/azure/windows-azure-for-research-overview.pdf
24. Talos IF, Archip N (2007) Volumetric non-rigid registration for MRI-guided brain tumor surgery, Tech. Report, Surgical Planning Laboratory, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
25. Kraut A, Moretti S, Robinson-Rechavi M, Stockinger H, Flanders D (2010) Phylogenetic code in the cloud–can it meet the expectations? Stud Health Technol Informat 159:55–63

A Hybrid Fusion of Multimodal Medical Images for the Enhancement of Visual Quality in Medical Diagnosis

S. Sandhya, M. Senthil Kumar and L. Karthikeyan

Abstract In the field of medical imaging, Multimodal Medical Image Fusion (MIF) is a method of extracting complementary information from diverse source images from different modalities such as Magnetic Resonance Imaging, Computed Tomography, Single Photon Emission Computed Tomography, and Positron Emission Tomography and coalescing them into a resultant image. Image fusion of multimodal medical images is the present-day studies in the field of medical imaging, biomedical research, and radiation medicine and is widely familiar by medical and engineering fields. In medical image fusion, single method of fusion is not proficient as it always lags in information while comparing with other available techniques. Hence, fusion for hybrid image is used to perform the image processing by applying multiple fusion rules. The integration of these results was obtained together as a single image. In proposed system, Shearlet Transform (ST) and Principal Component Analysis (PCA) are used to apply integrated fusion. The fusion technique is applied for CT that is Computed Tomography and Magnetic Resonance Imaging (MRI) images, where these images are first transformed using the Shearlet Transform and PCA is applied to the transformed images. Finally, the fusion image is acquired using Inverse Shearlet transform (IST). The proposed system performance is evaluated by using specific metrics, and it is demonstrated that the outcome of proposed integrated fusion performs better when compared to existing fusion techniques.

Keywords Shearlet Transform (ST) · Principal Component Analysis (PCA) Image fusion · Medical image

S. Sandhya (✉)
IT, Valliammai Engineering College, Chennai, India
e-mail: sanrsn@gmail.com

M. Senthil Kumar · L. Karthikeyan
CSE, Valliammai Engineering College, Chennai, India
e-mail: senthilkumarm.cse@valliammai.co.in

L. Karthikeyan
e-mail: karthikeyanl.cse@valliammai.co.in

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_7

1 Introduction

1.1 Medical Image Fusion

Due to the rapid development in the field of medical imaging, use of the medical images is prevalent in healthcare and biomedical applications like monitoring, diagnosis of disease, treatment, and research. Different modalities represent different information on the organs and tissues and with application ranges. This type of information cannot be obtained from single modal since, for example, structural images like Magnetic Resonance Imaging (MRI) provides soft tissues information and lacks in boundary information, Computed Tomography (CT) image shows bone structures and lacks in providing information about the tissues, Positron Emission Tomography (PET) image provides precise information in blood flow but lacks in boundary information and so on. Ultrasonography (USG) and magnetic resonance angiography (MRA), etc., provide high-resolution images with excellent structural detail and exact localization proficiency [1]. Hence, every single modality images have their limitations because those images are recorded with varying radiation levels. To overcome this, images from different modalities are in need. To integrate, two or more images are fused together to provide high information than the original input image from different modalities in Image Fusion (IF) technique. Therefore, combining images from multimodalities have gained the focus in the field of medical image processing [2].

The process or stages involved in any traditional image fusion technique are Image registration and Image Fusion. Image registration is performed where the input images are registered which is done by mapping the images which are taken as input in correspondence to the reference image. In the process of mapping, the corresponding images are matched based on certain features to support in the image fusion. Image registration turns to be difficult due to the existence of missing features, inter-image noise outliers in the images. Image fusion involves feature fusion by identifying and selecting the features with an emphasis on the significance of the selected features for the assessment purpose.

2 Related Works

Image fusion is performed through many fusion techniques. The fusion method is categorized as spatial domain and transform domain. In transform domain, images undergo multiple levels of resolutions, followed by various manipulations on the transformed images whereas the spatial domain methods work directly on the pixel values. Fusion methods like simple averaging, simple maximum fusion, Principal Component Analysis and Intensity Hue Saturation falls under spatial domain [1]. The sharpness and contrast of the images are found to be reduced in this method. Techniques such as wavelet transform, curvelet, contourlet transform, shearlet,

surfacelet, etc. fall under transform domain. Regardless of the worthy performance, transform domain methods still have some shortcomings.

Traditional image fusion methods lack the facility to get a high-quality image when combining two different modal images. This creates a necessity for an integrated image fusion technique. This hybrid fusion of images is done by combining spatial domain with transform domain technique for increasing the performance and quality of the fused image. Alternative way of hybrid fusion is to apply two stage transformations to the input images before the fusion process [2]. Thus, in overall hybrid fusion, techniques improve the visual quality and decreasing the image artifacts and noise.

El-Hoseny et al. [1] described the various hybrid fusion methods using Discrete Wavelet Transform/Additive Wavelet Transform along with PCA, Additive Wavelet Transform and Non-subsampled Contourlet Transform, AWT/NSCT and DT-CWT (Dual-Tree Complex Wavelet Transform), and DWT along with Curvelet Transform. The performance measures were compared against each of the proposed hybrid fusion algorithms. Rezaeifar and Saadatmand-Tarzjan [3] proposed a pixel level fusion algorithm for the multimodal medical images fusion based on the surfacelet transform. Menon et al. [4] compared the performance of transform domain based fusion algorithms namely Wavelet Transform, Contourlet Transform, and Non-subsampled Contourlet Transform. With the evaluation metrics, PSNR and RMSE values, NSCT was found to perform better than Wavelet and Contourlet Transform. Biswas et al. [5] implemented the hybrid fusion of MRI and PET images using the Shearlet Transform (ST) and Singular Value Decomposition (SVD) where the input images are converted into shearlet images and then SVD is applied, and finally Inverse Shearlet Transform (IST) is carried out to obtain the fused image. Harinkhede and Mishra [2] presented the fusion of images using DWT and PCA and utilized 2D filter for enhancing the quality of the fused images, where their fusion algorithm works in two stages. Rajalingam and Priya [6] presented the performance of traditional fusion algorithms using PCA, DWT, and PCNN against the hybrid model of DWT and PCA. During the performance assessment, it was given that hybrid model yields better results than the traditional fusion techniques. Gautam and Datar [7] described techniques for fusion of images such as DWT, PCA, DWT-PCA, DCT, and Curvelet Transform and through the evaluation results, it was observed that Curvelet Transform performs better. Sarala et al. [8] represented a multimodal image fusion using wavelet transform along with contourlet transform. Parameters like entropy, MSE and PSNR were used for estimating the efficacy of the hybrid model against prevailing techniques. Mazaheri et al. [9] projected the fusion of DWT and PCA for echocardiography images. Different metrics were employed to evaluate the performance of the system. Singh and Patel [10] gave a review of medical image fusion using wavelet transform. In this paper, multiresolution fusion of medical images had been discussed. Guruprasad et al. [11] constructed a hybrid DWT-DCT-PCA and its fusion outcomes are analyzed using PSNR and RMSE with the existing system. Nandeesh and Meenakshi [12] acquired the hybrid technique by integrating Principal Component Analysis and Stationary Wavelet Transform (SWT) where SWT is a time-invariant transform.

3 Proposed Method

The proposed method tries in improving the development of hybrid fusion technique by integrating PCA with Shearlet Transform, the two well-known image fusion methods. PCA method can preserve a better resolution, but it distorts spectral features with various degrees as well. ST is an evolving multi-scale geometric analysis tool as it augments the visual quality of the fused image. The prominent merit of ST is they work in multiple directions.

3.1 Principal Component Analysis (PCA)

PCA is a well-known procedure to perform dimension reduction, data visualization, and feature extraction. In general, PCA is defined by transforming high dimensional vector space into low dimensional space which helps to reduce the size of huge volume of medical image data without any loss of vital data. Many correlated variables are transformed into uncorrelated variables called principal components during the process of principal component analysis. Each and every obtained component is considered along the direction of maximum change which lies perpendicular to one another in the subspace [6, 7, 12].

3.2 Shearlet Transform (ST)

ST efficiently encode anisotropic features in a multivariate problem. ST is the extension of wavelet transform to incorporate the multivariate functionalities. The fusion method based on shearlet performs the decomposition of the input images, then the fusion rules are adopted and inverse shearlet transform is applied finally to reconstruct the image for obtaining the fused image.

Figure 1 depicts the architecture diagram of the proposed system. Here, the two images are taken as input on which both the combined spatial and transform techniques are applied to acquire the high-quality fused image.

The input images MRI and CT are considered and these images decomposed into low-frequency and high-frequency coefficients based on the shearlet transform. Then, the PCA is used to obtain the feature level information from input images. Finally, the inverse shearlet transform is applied to obtain the fused image.

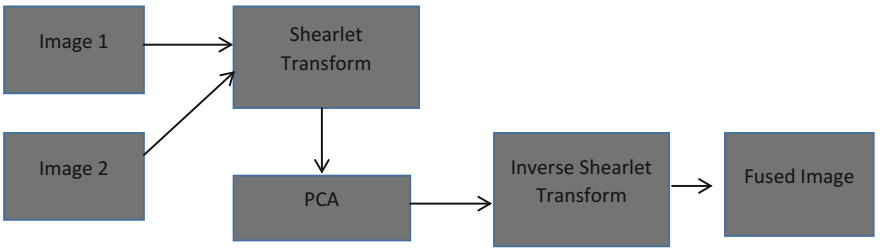


Fig. 1 The proposed architecture diagram

4 Evaluation Metrics

The apparent degradation of image is obtained by a feature known as image quality [13]. The legitimacy of the fused image leads to the requirement of image quality measurement. The fused image quality results are evaluated using various quality metrics like Entropy, MSE, PSNR Mutual Information, Mean, Standard Deviation Average, and Spatial Frequency.

For the comparison of the performance of the proposed system, few of the quality measures are taken into consideration. They are Entropy (EN), Standard Deviation (STD), Average Gradient (AVG), Image Quality Index (IQI), and Root Mean Squared Error (RMSE). Entropy and Standard Deviation gives the higher value for the proposed system as entropy contains more information from the image. Average Gradient (Avg) calculates the performance of fused image and gives the clarity of the fused image. IQI is higher, and RMSE is least for the proposed technique.

a. **Entropy** is the amount of information that the image contains and it can be defined as in Eq. (1) in which $p(x_i)$ is the probability of the occurrence of x_i.

$$E = -\sum_{i=0}^{n} p(x_i) \log p(x_i) \tag{1}$$

b. **Standard Deviation** (**STD**) is the quantity of deviation of the data from the average or mean value. If the image is clear, then the value of STD is higher. STD is defined as in Eq. (2), where M and N represent the dimensions of image f(i, j) and μ represents the mean value.

$$\mathrm{STD} = \sqrt{\frac{\sum_{i=1}^{M} \sum_{j=1}^{N} |f(i,j) - \mu|^2}{MN}} \tag{2}$$

c. **Average Gradient** (**g**) denotes the extent of texture difference in the image. It is calculated as in Eq. (3) where R and S are the image dimensions of the images x and y.

$$g = \frac{1}{(R-1)(S-1)} \sum_{i=1}^{(R-1)(S-1)} \frac{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial x}\right)^2}}{2} \tag{3}$$

d. **Image Quality Index** (**IQI**) is to measure the correlation between two images x and y, and it is given as follows in the Eq. (4), where $\bar{x}$, $\bar{y}$ gives the mean value of the image x, y. $\partial x^2 + \partial y^2$ is the variance of image x and y. ∂xy is the covariance between x and y.

$$Q_0(x, y) = \frac{\partial xy}{\partial x \partial y} \cdot 2\frac{\bar{x}\,\bar{y}}{\bar{x}^2 + \bar{y}^2} \cdot 2\frac{\partial x \partial y}{\partial x^{2+\partial y^2}} \tag{4}$$

e. **Root Mean Squared Error** (**RMSE**) relates to the pixels in the reference image and fused image. The RMSE value is equal to zero when the reference and fused image are alike and the value increases when the dissimilarity between the images increases. The formula to calculate RMSE is given in Eq. (5)

$$\text{RMSE} = \sqrt{\frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} \left(I_r(x, y) - I_f(x, y)\right)^2} \tag{5}$$

To evaluate the performance of the proposed system, the evaluation metrics are assessed for the input images CT and MRI given in Figs. 2 and 3. The metrics are compared for the fusion techniques PCA, ST and the combined fusion algorithm (PCA-ST) (Fig. 4).

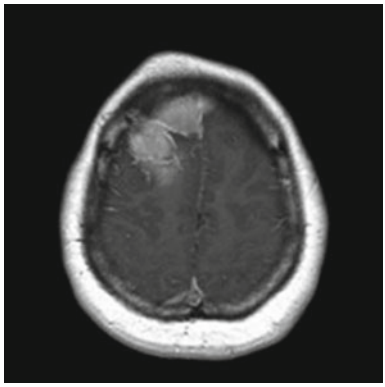


Fig. 2 Image (i) MRI

Fig. 3 Image (ii) CT

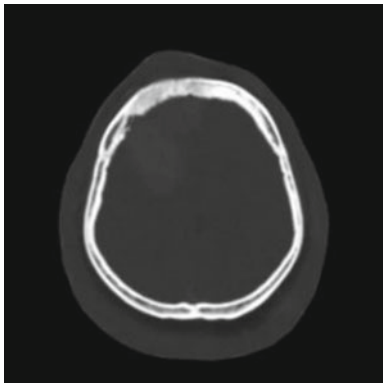
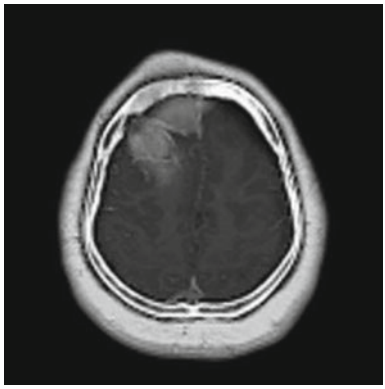


Fig. 4 Final fused image



5 Experimental Setup

To perform the proposed technique, two images MRI and CT are used in order to carry out the simulation results using MATLAB 2012. Then the simulated results are used to compare and analyze the results of image fusion with other fusion techniques. The experimental results are shown in Table 1. The performance metrics are examined for the fusion techniques DWT and CT along with the proposed method.

Table 1 Comparison of evaluation metrics for various fusion techniques

Metrics	DWT	CT	Proposed method
EN	6.519	6.882	6.946
STD	42.008	43.196	44.293
AVG	0.041	0.043	0.052
IQI	0.217	0.238	0.345
RMSE	0.314	0.242	0.213

Figure 5 shows the comparison of entropy measure for the techniques DWT, CT, and the proposed system. The entropy measure for the proposed system yields the highest value as the fused image contains more information.

Figure 6 gives the comparison of standard deviation where the value is higher for the proposed system comparatively with DWT and CT.

Figure 7 represents the comparison of AVG, and it shows the highest value for the proposed system due to clarity of the fused image and it is depicted in Fig. 4.

In Fig. 8, the Image Quality Index is compared, and the proposed system yields the highest value.

The performance metric RMSE is shown in Fig. 9 and the error value for the proposed system is the lowest.

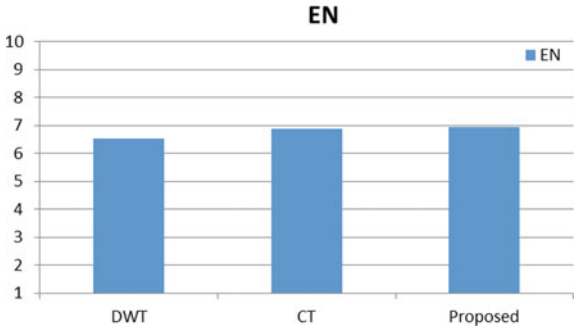


Fig. 5 Comparison of entropy measure

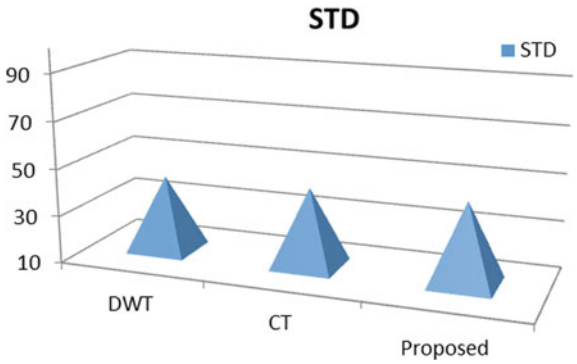


Fig. 6 Comparison of standard deviation

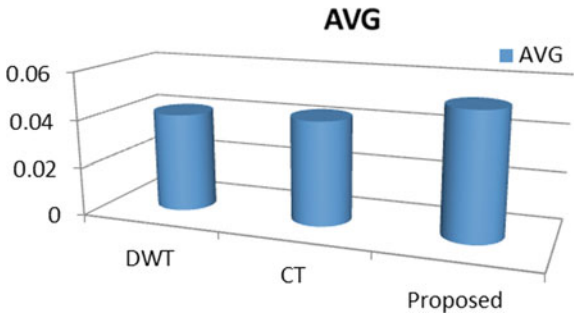


Fig. 7 Comparison of Average Gradient Measure

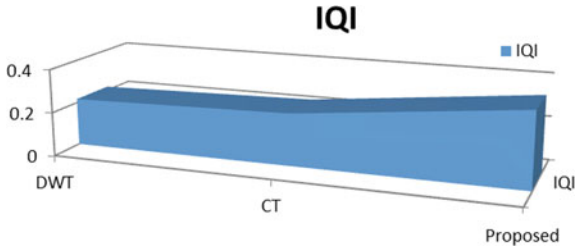


Fig. 8 Comparison of Image Quality Index measure

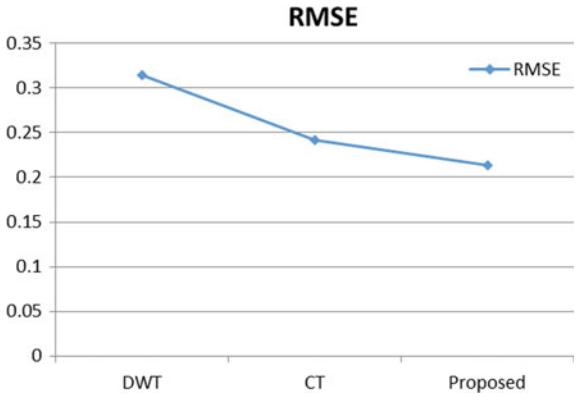


Fig. 9 Comparison of Root Mean Squared Error measure

6 Conclusion

In this paper, combined domain fusion algorithm is proposed for integrating the Computed Tomography and Medical Resonance Images by spatial and transforms domain techniques. The fusion results are compared with the methods DWT and CT along with the proposed system. While evaluating the performance, the proposed method shows better results for the medical image fusion. The proposed method preserves visual quality of the fused image obtained from multimodalities which will be helpful in the field of medical diagnosis. We will take effort in utilizing other image processing techniques based on Shearlet Transform in our future work.

References

1. El-Hoseny HM, Rabaie ESME, Elrahman WA, El-Samie FEA (2017) Medical image fusion techniques based on combined discrete transform domains. In: NSRC 2017, pp. 471–480
2. Harinkhede S, Mishra S (2018) A comparatively analysis of various hybrid image fusion techniques. Int J Eng Sci Res Technol. ISSN: 2277-9655, CODEN: IJESS7, pp. 51–55
3. Rezaeifar B, Saadatmand-Tarzjan M (2017) A new algorithm for multimodal medical image fusion based on the surfacelet transform. In: ICCKE 2017, pp. 396–400
4. Menon A, Arunvinodh C, Davis AM (2015) Comparative analysis of transform based image fusion techniques for medical applications. In: ICIIECS'15
5. Biswas B, Ghoshal S, Chakrabati A (2017) Medical image fusion by combining SVD and shearlet transform. In: ICCKE 2017, pp. 396–400
6. Rajalingam B, Priya R (2017) Multimodality medical image fusion based on hybrid fusion techniques. Int J Eng Manuf Sci 7(1):22–29
7. Gautam R, Datar S (2017) Application of image fusion techniques on medical images. Int J Curr Eng Technol 7(1):161–167
8. Sarala N, Lavanya M, Lavanya M (2017) A novel approach to multi-modal hybrid image fusion using wavelet and contourlet transform for medical diagnosis applications. Imperial J Interdisc Res 3(5):331–338
9. Mazaheri S, Sulaiman PS, Wirza R, Dimon MZ, Khalid F, Moosavi Tayebi R (2015) Hybrid pixel-based method for cardiac ultrasound fusion based on integration of PCA and DWT 2015:1–16
10. Singh BB, Patel S (2017) Review of medical image enhancement using wavelet based image fusion. Int J Eng Trends Technol 48:102–105
11. Guruprasad S, Kurian MZ, Suma HN (2015) Fusion of CT and PET medical images using hybrid algorithm DWT-DCT-PCA
12. Nandeesh MD, Meenakshi M (2015) A novel technique of medical image fusion using stationary wavelet transform and principal component analysis
13. Raut N, Paikaro PL, Chaudhari DS (2013) A study of quality assessment techniques for fused images. Int J Innov Technol Exploring Eng 2(4)

An Amplifying Image Approach: Non-iterative Multi Coverage Image Fusion

K. Elaiyaraja, M. Senthil Kumar and L. Karthikeyan

Abstract Better information can be obtained from sharp images than blurry images. Focused images are not possible to acquire in some circumstances from modalities. Some regions in the images may be blurred and out of focus. In order to overcome these drawbacks, taking more number of images of a scene (that is the same position) and applying fusion techniques to all images by merging to retrieve the best informative fused image. One of those fusion techniques is known as Multi-exposure Image Fusion. This technique synthesizes an LDR image (Low Dynamic Range image). The final result of an image will be more informative. The proposed Non-iterative Multi Exposure Image fusion is one of the promising techniques to enhance the images. Multiscale transform method and Sparse-Representation methods are popular nowadays. In Multi-scale transform methods, the images are divided into layers. In these layers, low pass band and the high pass band exist. In the low pass band, it is possible to lose some vital information when compared to its original image from source. This problem arises due to illumination in various levels of an image. Max-abs (Maximum Absolute) value can be used for the high-pass band to merging image. But for the low-pass band, the even averaging rule cannot be applied. Instead of choosing multi-scale transform and SR representation techniques, we are going to operate directly in all images. In existing system, MEF technique is used as an iterative approach. That means the origins of space from all images are iteratively increased up to convergence level. The proposed Non-iterative MEF context is to form better MEF for image fusion.

Keywords Image fusion · Medical image · Non-iterative · Multi-coverage · MEF

K. Elaiyaraja (✉)
Department of IT, Valliammai Engineering College, Chennai, India
e-mail: k.elaiyaraja@rediffmail.com

M. Senthil Kumar · L. Karthikeyan
Department of CSE, Valliammai Engineering College, Chennai, India
e-mail: senthilkumarm.cse@valliammai.co.in

L. Karthikeyan
e-mail: karthikeyanl.cse@valliammai.co.in

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_8

1 Introduction

MRI and CT scan are the common modalities used to diagnose diseases. Images from different modalities are fused together and overcome the deficiency of the single image pattern. The images from modalities are deteriorated with noise [13]. Image quality depends upon the noises affected in the image and preserves edges [19]. Apart from noise, blurring affects image quality. Noise can be classified into Gaussian, Salt pepper, Speckle, Poisson and Blurred [16]. Image enhancement is required to improve quality of an image for the betterment of diagnosing diseases by avoids these kinds of noise [18]. Low pass, average or median techniques are used to prevent noise.

Image Fusion technique or method combines multiple images into a single fused image. These fused images are retrieved from different modalities [7]. The main agenda of image fusion is to get an increased resolution image. The fused images will have more information or details than source images that is non-fused images [3].

Image fusion methods are classified as Pixel, Decision and feature level. In pixel-level image fusion, pixel transformations are not used [4]. In decision level, fuzzy techniques, voting techniques are used to preserve information of an image [6]. And finally in the feature level, edge details, texture details are concentrated.

In this proposed work, fusion concept is applied to underexposed images as well as overexposed images. Noises are present in underexposed images. Blurs exist in overexposed images. The MEF process consist more information by reducing noise and preserves sharp edges. This technique also reserves contrast and intensity.

An iterative method which selects various ways based on fusion used to retrieve information. In this iterative method begins with segmentation of an image. Next regions and its energy are calculated. By using iterative, the most significant energy region is applied. This largest energy region is called as “main” region.

In the proposed work, the variations in energy levels are calculated. Low energy level regions are selected by comparing with source images. Then fusion concepts are applied to small region energy levels of an image to retrieve information. That is image restoration.

2 Materials and Methods

MEF (Multi-Exposure Fusion) image is an economical and simple method. In this method mainly concentrate Low-pass band areas to fuse significant information [1]. So many MEF algorithms are available in the market to ensure the quality of an image [14]. However, still cannot be achieved significant detailed information as the existing methods have not optimized as well [12]. For instance, Structural Similarity index (MEF-SSIM) method exposed a weak correlation in showing space of given images [9]. The measurement of this technique is based on the luminance of two images, contrast and resemblances of two images.

Mertens et al. [10] specify the fusion can be done using the multi-resolution technique. The filters like Bilateral, recursive and guided are used to obtain edge information of an image [17]. Gu et al. [5] elaborated pixel-level gradient for structure tensor and smoothening edges to compute the weight of *i*th pixel of *k*th exposure image. Zheng and Rahardja have given a similar gradient-based Multi Exposure Fusion method.

Debevec and Malik [2] introduced a model to find average mapping value of the pixel from scene radiance. Shen et al. [3] developed the High Definition Range scene radiance with the help of retinex theory. Robertson et al. [11] noted a probabilistic formulation for the estimation of the response function. Goshtasby [3] recommended block-based scheme. In this scheme, images are partitioned into uniform blocks. Then this scheme selects only the informative block and applying blending functions on selected images.

Kao et al. [8] stated an empirical method for MEF to find pixels which are unstable. Shen et al. [15] noted probabilistic model for MEF to achieve balanced local contrast and consistency of colors. Jacobs et al. [7] shown the composite multi-exposed images derived from gradient information. Jacobs et al. [7] explored the construction of High Definition Rage image through the calculation of the weighted average value of irradiance.

In an existing algorithm, summated weight framework specified by MEF as follows.

$$y = \sum_{k=1}^{K} w_k X_k \tag{1}$$

where, K => Number of Exposure levels, X_k => Co-located pixels from the X_k-th image. y => corresponding pixel if fused image Y. w_k => information of X_k. Here, the decomposition scheme is applied to find distortions, exposedness, and gradients. The enhanced fused image is obtained from this MEF algorithm. Gradient direction concepts are used to face the misalignment issues of given images. The median filter was applied to avoid moving objects. Patch-wise matching algorithm was used to handle object motions. Random walker method was used to calculate a weight for this patch.

The formulated MEF for quality criteria is

$$\text{Yopt} = \text{argy max Q}(\{\text{Xk}\}, \text{Y}) \tag{2}$$

An iterative MEF-SSIM algorithm for Gradient ascent at the *i*th iteration as follows.

$$\text{Yi} + 1 = \text{Yi} + \lambda \nabla \text{YQ}(\{\text{Xk}\}, \text{Y}) | \text{Y} = \text{Yi} \tag{3}$$

Similarly, the preferred mean intensity for a fused image is calculated by

$$i = \frac{\sum_{k=1}^{K} \mu(\mu k, lk) lk}{\sum_{k=1}^{K} \mu(\mu k, lk)} \tag{4}$$

3 Proposed Work

Initially, the images are decomposed, and this decomposition is obtained as a steerable form of pyramid decomposition [16]. Then quality maps are accessed from these decomposition subbands. SSIM is used to define these subbands. After finishing these steps, quality maps are fused together for the given image. Finally, Weight strategy based on contrast is used to retrieve the superior value of subband.

The exact quality map at location x and y in sub-band 's' of the Low Dynamic Range images are computed from the given images by the assumption which the collection of LDR images are preserved. The exact mapping is determined with the help of values retrieved from the location x and y in subband 's'. Now Structural Similarity Index value [13] is used to calculate the image information (Fig. 1).

The two kinds of parameters are taken for the proposed system. One is the index parameter and the second one is the model parameter. The model parameter values are obtained from the existing publications.

$$\text{Spread parameter } \sigma_g = 0.2 \text{ and } \sigma_l = 0.2$$

Stabling constants $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ where the K_1 value is 0.01, and K_2 value is 0.03. L value is assumed to be an operating range of source image. For

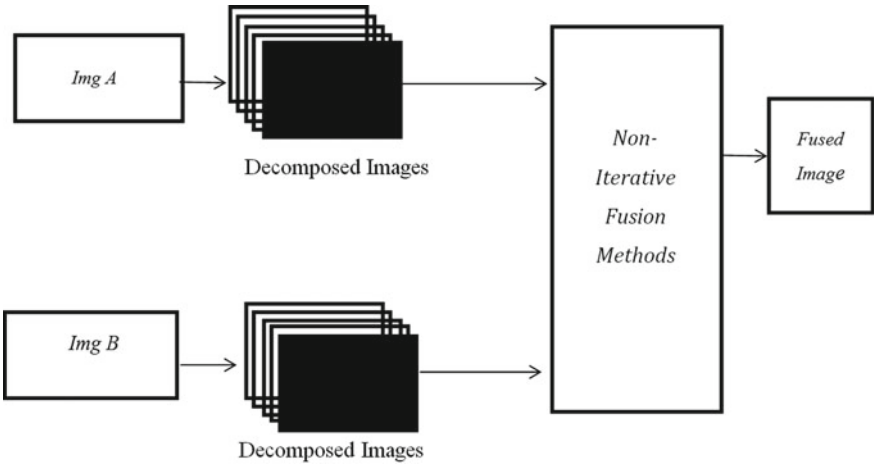


Fig. 1 Architecture diagram of iterative fusion method

an eight bit sequence, L value is 255. Our non-iterative MEF directly calculates the convergence speed as well as stability.

The degree of distortion [SAMDOD] is calculated as:

$$D_s \triangleq \sqrt[q]{\frac{1}{L}\sum_{l=1}^{L}\left|Q\left(\widehat{G_1}, P\right) - Q\left(\widehat{G_1}, P\right)\right|q} \tag{5}$$

where the P value is obtained from filtering with an LDR endpoint at the resolution ratio D_s attains minimum if both the vectors are equal.

Spectral Angle Mapper (SAM) is calculated as follows:

$$\text{SAM} = \arccos\left(\frac{\left\langle V, \hat{V}\right\rangle}{\|V\|2 \cdot \left\|\hat{V}\right\|2}\right) \tag{6}$$

SNR is calculated as follows:

$$SNR_{max2} = \frac{\sum_{x=0}^{N-1} y^2 w^2(x) + w^2(x-2) + 2yw(x)w(x-2)}{\sum_{x=0}^{N-1}\left(1 + 2y^2 - 2y\sqrt{1+y^2}\right)w^2(x)} \tag{7}$$

Algorithm for Non-iterative

Step 1: Read 'n' images $I_1, I_2, \ldots, I_n$ and priority index of each image.
Step 2: Read Priority index p1, p2, …, pn.
Step 3: Making 'fis" file. In this file, 'Gaussian' function is applied.
Step 4: To obtain the final image, the pixel values are applied with n input and the output column.
Step 5: If pn > 1, image should be fused.
Step 6: Training data is made.
Step 7: Checking data of given images.
Step 8: Show fused image.

For the deployment of this algorithm, n = 2 and t1 and t2 values are assumed as 2. Images used for testing is shown in Figs. 2 and 3.

4 Experimental Setup

The proposed algorithm is deployed using R Studio R 3.3.2 in windows operating systems. The experiment is done with four images taken in sequential order. Here, LDR and HDR information samples are selected from the origin of each source image manually. Based on the selection rules, 40–60% of pixels selected in LDR regions for training samples in this experiment.

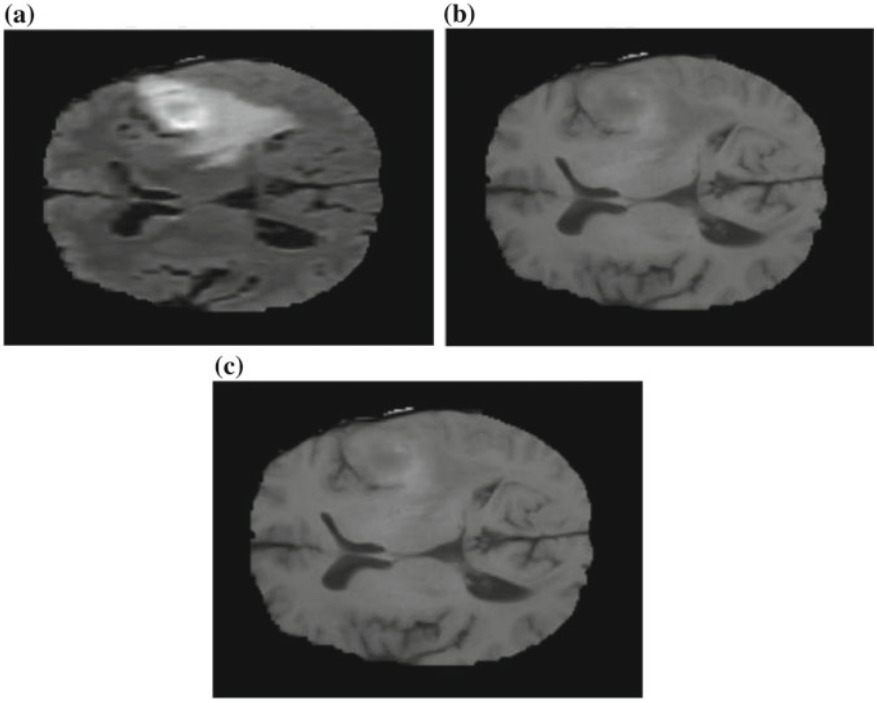


Fig. 2 Source images with its corresponding fused image (first pair). **a** and **b** Different source image of the same scene. **c** Fused image

The Non-iterative MEF algorithm is initialized on these images. Two operators namely x and y denoted for weighting factors linearly associate with given images. The real fusion results default settings are specified in the table are obtained from widely available.

To confirm the supremacy of the proposed fusion system, two other fusion algorithms namely MAP and Wavelet have chosen. First source images are decomposed. Then extraction of vector sets is obtained. Finally, Non-iterative MEF concept is applied.

5 Performance Evaluation

In this section, three different kinds of measures namely SAM, DOD, and RSNR are used to calculate the performance of proposed systems. Tests are done on two pairs of given images (refer Figs. 2 and 3).

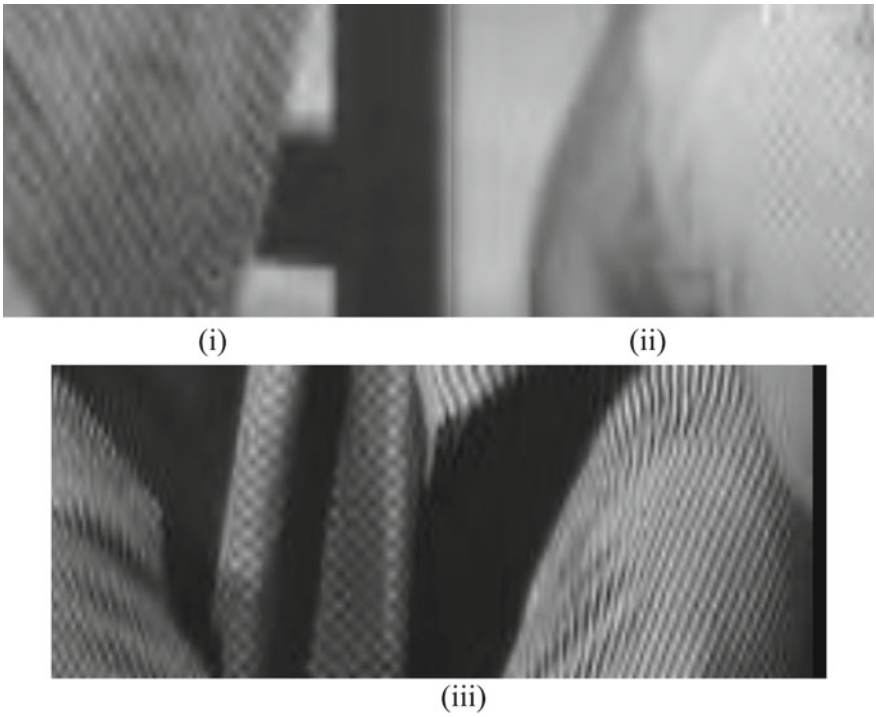


Fig. 3 Source image and corresponding fused image (second pair). (i) and (ii) different source image of the same scene. (iii) Fused Image

Table 1 Parameters evaluation for fused image using different types of techniques

Methods	Parameters		
	Spectral angle mapper	Degree of distortion	Reconstruction SNR
MAP	5.05	4.87	23.33
Wavelet	3.98	3.89	25.53
Proposed	3.33	3.6	27.03

The values of SAM, DOD, and RSNR of the proposed system are 3.33, 3.6 and 27.03 respectively. The fused image obtained by this proposed system has better quality with a lot of necessary information. Table 1 represents the comparisons between proposed and existing fusion methods.

The SAM parameter value of the proposed system is reasonably reduced when comparing with another existing system. Figure 4 shows the comparison results of SAM parameter.

The next parameter DOD of this proposed system is analyzed with MAP and Wavelet method. The result of the comparison is shown in Fig. 5.

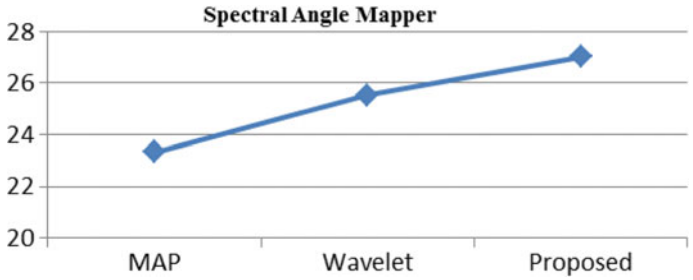


Fig. 4 Comparison of SAM with other approaches

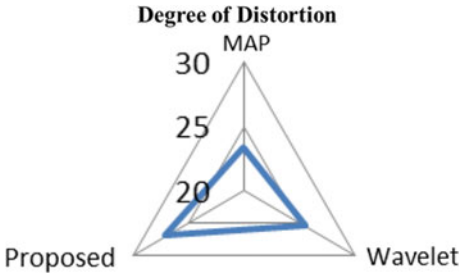


Fig. 5 Comparison of DoD with other methods

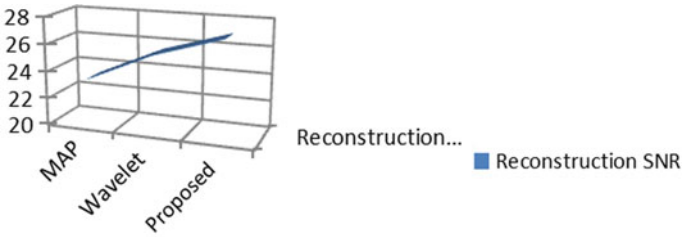


Fig. 6 Comparison of RSNR with other methods

The third parameter Reconstruction SNR value is high compared to other techniques. This comparison is represented in Figs. 6 and 7.

6 Conclusion

The values obtained from this proposed method are accurate and efficient when compared to other fusion methods. Here, the given images were decomposed and then applied non-iterative MEF algorithm. SAM, Distortion ratio and RSNR are computed with existing fusion methods and proposed methods. Then those values

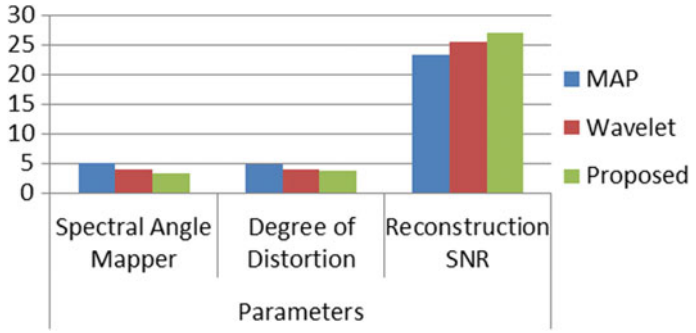


Fig. 7 Comparison chart of proposed techniques with different techniques

are compared, and the results and its corresponding graphs are shown. However, our algorithms can be improved or optimized further by implementing batch processing in this proposed method.

References

1. Burt PJ, Kolcznski RJ (1993) Enhanced-image capture through fusion. In: Proceedings of the IEEE international conference computer vision
2. Debevec PE, Malik J (2008) Recovering high dynamic range radiance maps from photographs. In: ACM SIGGRAPH 2008 classes. ACM, p 31
3. Goshtasby AA (2005) Fusion of multi-exposure images. Image Vis Comput 23(6):611–618
4. Goshtasby A, Nilolov S (2007) Image fusion—advances in the state of art. Inf Fusion
5. Gu B, Li W, Wong J, Zhu M, Wang M (2012) Gradient field multi-exposure images fusion for high dynamic range image visualization. J Vis Commun Image Represent
6. He K, Sun J, Tang X (2012) Guided-image filtering. Pattern Anal Mach Intell 35:1397–1409
7. Jacobs K, Loscos C, Ward G (2008) Automatic high-dynamic range image generation for dynamic scenes. IEEE Comput Graph Appl 28(2)
8. Kao WC, Hsu CC, Chen LY, Kao CC, Chen SH (2006) Integrating image fusion and motion stabilization for capturing still images in high dynamic range scenes. IEEE Trans Consum Electron 52(3):735–741
9. Ma K, Li H, Yong H, Meng D, Wang Z, Zhang L (2017) Robust multi-exposure image fusion: a structural patch decomposition approach. IEEE Trans Image Process 26(5):2519–2532
10. Mertens T, Kautz J, Van Reeth F (2009) Exposure fusion: a simple and practical alternative to high dynamic range photography. In: Computer graphics forum
11. Robertson MA, Borman S, Stevenson RL (2003) Estimation-theoretic approach to dynamic range enhancement using multiple exposures. J Electron Imaging 12(2):219–228
12. Rong H, Hui W, Xiaoxu L (2017) Observation of CT-MRI image-fusion in post operative precise radio therapy for Glimas. Chin J Radiat Oncol Biol
13. Sampat MP, Wang Z, Gupta S, Bovik AC, Markey MK (2009) Complex wavelet structural similarity: a new image similarity index. IEEE Trans Image Process 18:2385–2401
14. Shen F, Zhao Y, Jiang X, Suwa M (2009) Recovering high dynamic range by multi-exposure retinex. J Vis Commun Image Represent 20(8):521–531
15. Shen R, Cheng I, Shi J, Basu A (2011) Generalized random walks for fusion of multi-exposure images. IEEE Trans Image Process 20(12):3634–3646

16. Simoncelli EP, Freeman WT (1995) The steerable pyramid: a flexible architecture for multi-scale derivative computation. In: Proceedings of the IEEE International Conference on Image Processing, pp 444–447
17. Zhang W, Cham WK (2012) Gradient-directed multi-exposure composition. IEEE Trans Image Process 21(4):2318–2323
18. Zhao Y, Zhao A, Hao A (2014) Multi-modal medical image-fusion using improved multichannel PCNN. Biomed Mater Eng
19. Zhu J, Che J, Guo Z (2014) Curvelet-algorithm of remote-sensing image-fusion based on local-mean and standard-deviation. J Ningxia Univ

U-Net Based Segmentation and Multiple Feature Extraction of Dermascopic Images for Efficient Diagnosis of Melanoma

D. Roja Ramani and S. Siva Ranjani

Abstract Skin cancer is found to be one of the most common types of deadly cancers among human beings in recent years. Computational-based techniques are developed to support the dermatologists for the early diagnosis of skin cancer. Computational analysis of the skin lesions in the dermascopic images is a challenging task due to the difficulties such as low-level of contrast between the lesion and surrounding skin regions, irregular and vague lesion borders, artifacts and poor imaging conditions. This paper presents a U-Net based segmentation and multiple feature extraction of the dermascopic images for the efficient diagnosis of skin cancer. The input dermascopic image is preprocessed to remove the noise and hair in the skin image. Fast Independent Component Analysis (FastICA) is applied to the skin images for obtaining the melanin and hemoglobin components. The U-net segmentation is applied to the dermascopic image to separate the cancer region from the background of the skin image. Different features such as vascular features, color features, texture features, RGB features, and depth features are extracted from the segmented image. RVM classification is applied to classify the normal and abnormal images. With the efficient segmentation and extraction of multiple features, our proposed work yields better performance than the existing segmentation and feature extraction techniques.

Keywords Melanoma · Independent component analysis · U-Net segmentation Vascular features

D. Roja Ramani (✉) · S. Siva Ranjani
Department of Information Technology, Sethu Institute of Technology, Pulloor, India
e-mail: rosevsroja@gmail.com

S. Siva Ranjani
e-mail: ssivaranjani@sethu.ac.in

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_9

1 Introduction

Skin is the largest organ that covers the muscles, bones and all parts of the human body. A minor change in the skin function may affect other parts of the body. As the skin is exposed to the outer environment, it is highly prone to disease and infections. The spot on the skin is called as a lesion area. Skin lesions are highly useful in the diagnosis of disease such as chickenpox, keratodermia, melanoma, etc. [1]. Malignant melanoma is found to be one of the rapidly increasing cancers with estimated new diagnosis rate of 91,270 (55,150 in men and 36,120 in women) and death rate of about 9,320 (5,990 males and 3,330 females) in the United States in 2018 [2]. The survival rate of the advanced stage cancer is around 10–20% and survival rate in the early stages of cancer is above 95% [3].

Early detection of the skin disease is highly complex to the inexperienced dermatologist. The Computer-Aided Diagnosis (CAD) systems facilitated early diagnosis of skin disease without requiring physical contact with the skin [1]. Computerized image analysis methods are developed for improving the visual interpretation of the dermoscopic images. Computer-based analysis of the skin lesion images is highly significant in skin cancer prevention [4]. Korotkov and Garcia [5] presented a review of the computer-aided diagnosis systems for the pigmented skin lesions (PSLs). Smith and MacNeil [6] conducted a review of the existing non-invasive imaging techniques used for the diagnosis of skin lesions.

The fundamental steps in the computer-based image analysis are preprocessing, segmentation, feature extraction, and classification. Preprocessing techniques such as color space transformation [7], contrast enhancement [8], and artifact removal [9] are important for efficient segmentation of the skin lesions. Segmentation enables extraction of Region of Interest (RoI) from the macroscopic image under analysis [4]. Automatic segmentation of melanoma from the surrounding skin regions is an essential step in the computer-based dermascopic image analysis [10, 11]. Due to the variations in the size, shape, and color of melanoma along with different types of skin and texture and presence of artifacts such as hairs, frames, blood vessels, and air bubbles, segmentation becomes a trivial task. In some cases, the lesions have irregular and fuzzy borders and the contrast between the skin lesions and surrounding regions is low [12].

This paper presents U-Net based segmentation and extraction of color, texture, depth, RGB and vascular features for the efficient diagnosis of melanoma. Initially, the dermascopy image dataset is preprocessed to remove the noise and hair present in the input dermascopic image. FastICA is applied to the skin images for obtaining the melanin and hemoglobin components. The U-net segmentation is applied to the dermascopic image to separate the ROI from the background of the skin image. Different features such as vascular features, color features, texture features, RGB features, and depth features are extracted from the segmented image. RVM

classification is applied to classify the normal and abnormal images. Texture analysis assists in the effective discrimination of the benign and malignant skin lesions by evaluating the roughness of the skin structure [13]. When compared with the Support Vector Machine (SVM) classifier, the RVM classifier could considerably reduce the computational complexity with better detection accuracy. The RVM classifier requires only minimum number of training vectors and yield maximum accuracy. As RVM is based on the probabilistic prediction, it can improve the classification accuracy. The RVM-based algorithms use significantly fewer kernel functions and do not need to estimate any parameters associated with the learning or the optimization technique to be used [14]. Vascular structures play a vital role in the pathological assessment of skin diseases. The morphology of the vascular in the skin lesions will serve as a biomarker for the diagnosis of malignant tumors. The vascular formation and angiogenesis are an indicator of the development and growth of tumor [15]. Hence, detection of the cutaneous blood vessel provides acute information during the evaluation of skin lesions [16]

The paper is structured in the following way: Sect. 2 presents a brief overview of the recent dermascopic skin image segmentation approaches. The proposed work including preprocessing, FastICA, U-Net based segmentation, multiple feature extraction, and RVM classification is explained in Sect. 3. Section 4 illustrates the performance analysis including feature-based analysis, segmentation-based analysis, and classification-based analysis of the proposed work and existing techniques. The concluding remarks of the proposed work are described in Sect. 5.

2 Related Works

Emre Celebi et al. [17] developed a method for detecting the skin lesion border in the dermascopy images by combining four different thresholding methods. The proposed method is easy to implement, robust and extremely fast without requiring sophisticated post-processing. However, the proposed method may not perform well on the dermascopic images with hair or bubbles. Peruch et al. [18] presented a novel technique for the segmentation of melanocytic lesion by combining the thresholding scheme with the optimizations. The proposed scheme is fast, robust, and accurate than the existing segmentation techniques, even during the presence of artifacts such as hairs or air bubbles, skin deformations, and illumination variations. The cognitive process of the dermatologists is reduced. Zhou et al. [19] proposed a Mean Shift based Gradient Vector Flow (MSGVF) segmentation algorithm for the segmentation of two-dimensional (2D) and three-dimensional (3D) images. The drawbacks of the proposed algorithm are the need for a large amount of computation to achieve convergence and more time consumption. Sadri et al. [20] introduced a novel approach for segmenting the skin lesions based on the Wavelet Network (WN). The image is seg-

mented and the exact boundary of the lesions is determined accurately. Xie and Bovik [21] developed a new segmentation algorithm by combining a Self-Generating Neural Network (SGNN) and Genetic Algorithm (GA). The proposed algorithm provided more accurate segmentations. However, it is difficult to obtain a priori information about the number of clusters in the dermoscopy images, due to the complexity of skin lesions. Celebi et al. [22] presented a complete overview of methods for detecting border in the spectroscopy images, based on their color space.

Deep learning mechanisms can learn a set of high-level features from the low-level features and yield high classification accuracy without the need for extracting handcrafted features. Nasr-Esfahani et al. [23] applied Convolutional Neural Network (CNN) for the detection of melanoma. Better detection accuracy is achieved and the benign and melanoma cases are distinguished efficiently by training the CNN with a large number of samples. Automated classification of the lesions using the images is a challenging task due to the fine-grained variability in the appearance of lesions. Deep CNN is applied for the classification of malignant carcinomas and melanoma and benign seborrheic keratosis and nevi. The deep CNN shows reliable cancer classification when tested on a larger dataset [24]. Jafari et al. [25] proposed a method for accurately extracting the lesion region based on the deep neural networks. The proposed method achieved high detection accuracy and sensitivity than the existing methods. Jafari et al. [26] developed a method for accurate extraction of the skin lesion region based on the deep learning approaches. A segmentation mask showing the lesion region is generated and refined with post-processing operations. The segmentation accuracy is improved. Li and Shen [27] proposed two fully convolutional residual networks for efficient segmentation and classification to detect the melanoma. The proposed network achieved best average precision and sensitivity for dermoscopic feature extraction than the existing deep learning frameworks. Kawahara and Hamarneh [28] applied a pre-trained FCNN to detect the dermoscopic features from the skin lesion images. The feature maps from several layers in the network are interpolated to match the size of the input image. The resized feature maps are concatenated and the network is trained to reduce a smoothed negative F1 score. The proposed network achieved highest averaged area under the receiver operator curve.

3 Proposed Work

3.1 Preprocessing

The overall flow diagram of the proposed work is shown in Fig. 1. Dermoscopy is an imaging technique used for examining the lesions in the skin with a dermatoscope. This process is performed by using oil immersion between the skin and optics. The lens of the dermatoscope is placed directly in contact with the skin for illuminating the sub-surface structures to magnify the skin. This reveals the pigmented structures

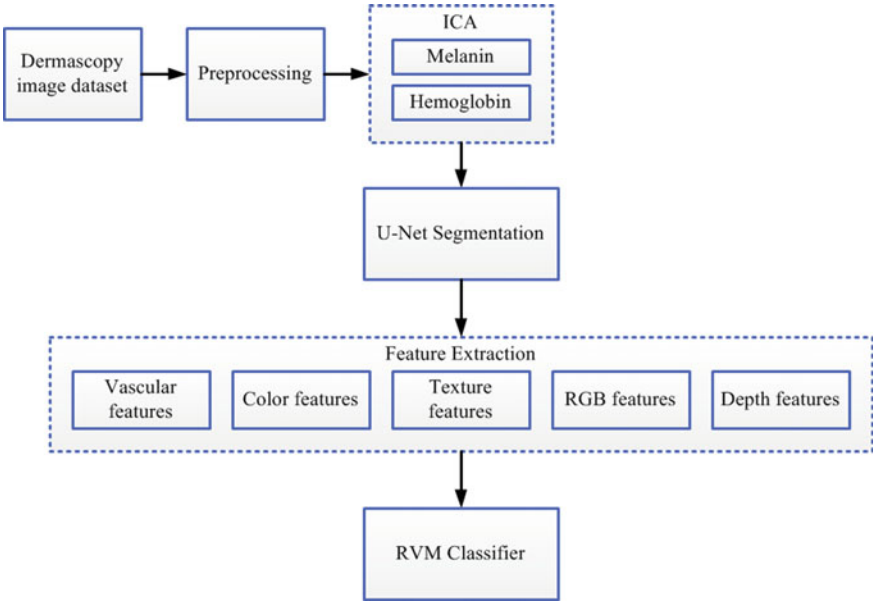


Fig. 1 Overall flow diagram of the proposed work

and color shades that are invisible to the naked eye. This enables direct visual-based analysis of the epidermis. The input dermascopic image contains noises in the form of hairs. This causes imprecision in the classification process. Preprocessing is performed to remove the noise in the image. In this work, dull razor is applied for removing hair in the image. Fast median filtering of 5×5 window size is applied for filtering the noise in the image. The shapes and edges of the images are enhanced using the contrast enhancement technique for improving the segmentation accuracy [29]. Input images obtained from the public ISBI 2016 dataset [30] and dataset obtained from Atlas of dermoscopy by Argenziano [16] are preprocessed. U-net based segmentation is applied, features are extracted and the images are classified as normal and melanoma images.

3.2 FastICA

Human skin is a multi-layered structure including epidermal and dermal layer with melanin and hemoglobin components [16]. Melanin is responsible for the brown color of the skin and hemoglobin is responsible for the color of blood circulated within the blood vessel and purple color of the skin. The melanin and hemoglobin absorb light in the visible spectrum for recognizing the underlying skin structures by using the skin color information. The melanin and hemoglobin components are the

most dominant factors in skin color. The amount of components in skin is mutually independent from each other. FastICA is proposed for the analysis of facial color image [31]. The skin image is decomposed into melanin and hemoglobin channels for detecting the hemoglobin channel and analyzing it to cluster the erythematous areas and segment the vasculature using the shape information.

FastICA is a technique for separating a multivariate signal from the constructing melanin and hemoglobin components. It is applied to the skin images for obtaining the melanin and hemoglobin components. For solving the problem of skin pigmentation that obstructs the appearance of blood vessels, this approach is based on the extraction of melanin and hemoglobin components. The hemoglobin channel is analyzed for segmenting the cutaneous vasculature.

The total amount of absorbed light is stated as a linear combinations of the densities of the underlying substances, based on the modified Beer–Lambert law. The color density $I_{x,y}$ in the optical density domain is defined as

$$I_{x,y} = \left[-\log r_{x,y}, -\log\ g_{x,y}, -\log\ b_{x,y}\right] \tag{1}$$

where $r_{x,y}$, $g_{x,y}$, $b_{x,y}$ are regularized color values of the RGB color space. By considering the abovementioned assumptions, the density vector is modeled as a linear combination of melanin and hemoglobin quantities

$$I_{x,y} = c^m q^m_{x,y} + c^h q^h_{x,y} + \Delta \tag{2}$$

where c^m and c^h are pure color per unit density of melanin and hemoglobin, respectively. q^m and q^h are the relative amount of the melanin and hemoglobin components in each pixel (x, y); Δ is the stationary column vector caused by the skin pigments and structures. In the conventional ICA, c^m and c^h are considered as mixing signals, q^m and q^h are considered as source signals and $I_{x,y}$ is considered as an observed signal. Principal Component Analysis (PCA) is implemented to the three-dimensional input to attain best results. First, two principal components are chosen for the analysis while preserving 99% of the total variance of data. ICA is applied for extracting the relative amount of melanin and hemoglobin components and estimating the pure color per density vectors $\tilde{C} = \left[\widetilde{c^m}, \widetilde{c^h}\right]$. The non-Gaussianity in the estimated melanin and hemoglobin components is maximized to ensure independency among the components. If it is assumed that the minimum amount of each of the components in any pixel is zero, then

$$\min_{x,y}\left\{\tilde{C}^{-1} I_{x,y}\right\} - \min_{x,y} \tilde{C}^{-1}\Delta = 0 \tag{3}$$

where $\tilde{C}$ represents the estimated pure color density vector. E is defined as

$$E = \min_{x,y}\left\{\tilde{C}^{-1} I_{x,y}\right\} \tag{4}$$

The relative amount of melanin and hemoglobin components is obtained as

$$\left[q^m_{x,y}, q^h_{x,y}\right]^t = \tilde{C}^{-1} I_{x,y} - E \tag{5}$$

The skin composition is obtained by

$$I'_{x,y} = \tilde{C}\left(K\left[q^m_{x,y}, q^h_{x,y}\right]^\tau + jE\right) + j\Delta \tag{6}$$

$I'_{x,y}$ is the synthesized skin color. 'K' and 'j' are the control parameters for controlling the amount and effect of the stationary signal, respectively. The melanin and hemoglobin components of the color image are extracted by setting $j = 0$ and $K = diag[1, 0]$ and $K = diag[0, 1]$. But, FastICA involves ambiguity on permutation and does not determine which of the melanin and hemoglobin components correspond to each chromophore. After the skin decomposition and extraction of the densities of the melanin and hemoglobin components, an expert user determines which component represents the chromophore. As an empirical solution considering that the hemoglobin deals with the erythema and redness of the skin, the hemoglobin channel is correlated with the redness. It is known that the a^* channel in the $L^*a^*b^*$ color space can show the redness of the image, the hemoglobin component comprises higher correlation with the a^* channel. This concept is modeled in terms of the following equation:

$$q^h = arg\max_i\left(corr\left(q^i, a^*\right)\right), i \in m, h \tag{7}$$

$$corr\left(q^i, a^*\right) = \frac{Cov\left(q^i, a^*\right)}{\sqrt{Var\left(q^i\right)Var(a^*)}} \tag{8}$$

Three clusters within skin: normal skin (n), pigmented skin (p) and erythema (e) are developed. A reference value vector is learned for each cluster through a set of selected pixels among the images. FastICA is applied to the selected images and the hemoglobin channel is obtained. Then, mean values along with the standard deviations of red, green and blue channels of the hemoglobin component of the three clusters are calculated as $\bar{I}^n$, $\bar{I}^p$, $\bar{I}^e$, respectively.

After obtaining the reference values and storing it as a priori information, a thresholding framework is designed based on the Mahalanobis distance on the hemoglobin component for classifying the red regions. The Mahalanobis distance from the three clusters is calculated for each pixel. The pixel is classified into the group with the closest distance

$$E^{i}_{x,y} = \sqrt{\frac{\left(r_{x,y} - r_i\right)^2}{\sigma_{ri}} + \frac{\left(g_{x,y} - g_i\right)^2}{\sigma_{gi}} + \frac{\left(b_{x,y} - b_i\right)^2}{\sigma_{bi}}}, where\, i \in n, p, e \quad (9)$$

where $E^{n}_{x,y}, E^{p}_{x,y}, E^{e}_{x,y}$ are the Mahalanobis distances of the pixel (x, y) from the three clusters. r_i, g_i and b_i are the RGB values of the clusters. If $E^{e}_{x,y} = \min\left(E^{n}_{x,y}, E^{p}_{x,y}, E^{e}_{x,y}\right)$, the pixel is classified as belonging to the erythematous cluster. Each pixel is classified into one of the clusters. As a result, a mask image is generated for segmenting the red areas of the skin [16].

3.3 Segmentation

The main objective of the segmentation process is to separate the Region of Interest (ROI), i.e., cancerous region from the background of the input skin image [32]. In our proposed work, U-Net segmentation as shown in Fig. 2 is applied for segmenting the input dermascopic image.

The U-Net architecture involves contracting path at the left side and expansive path at the right side. The contracting path follows the general architecture of a convolutional network. It includes the recurrent application of two 3×3 convolution processes, each followed by a Rectified Linear Unit (ReLU) and a 2×2 max-pooling

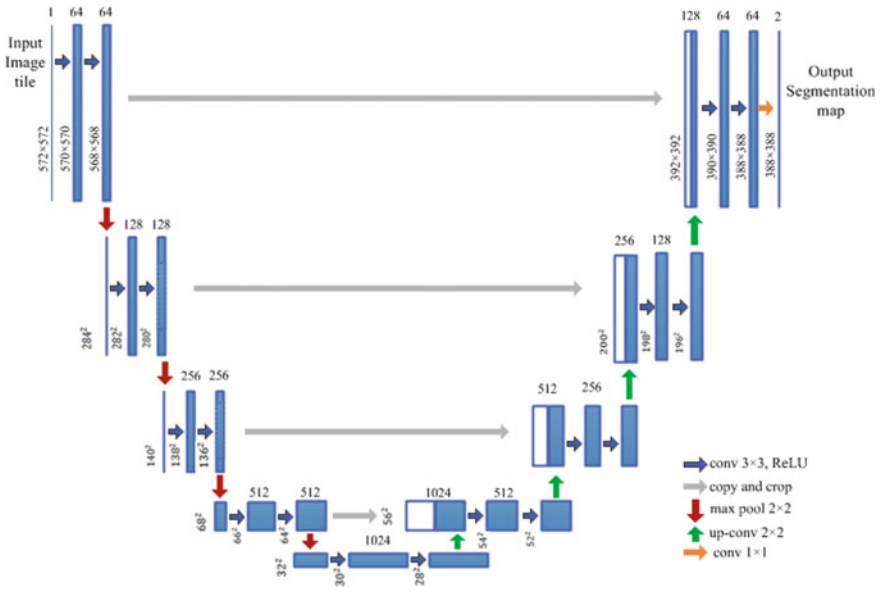


Fig. 2 U-Net architecture

operation with stride 2 for downsampling. The number of feature channels is doubled up at each downsampling step. Each step in the expansive path involves upsampling of the feature map followed by a 2×2 convolution. This minimizes the number of feature channels to half, a concatenation with the collected feature map from the contracting path and a couple of 3×3 convolutions, each followed by a ReLU. Collection of the feature map is essential due to the loss of border pixels in the image during each convolution step. Each 64-component feature vector is mapped to the required number of classes by applying a 1×1 convolution at the final layer. There are 23 convolutional layers in the fully convolutional network. It is important to choose the size of the input title for the continuous tiling of the output segmentation map, such that all 2×2 max-pooling operations are applied to a layer with a uniform x-size and y-size.

The main idea in a fully convolutional network [33] is to balance a typical network by the consecutive layers. The pooling operators are substituted with the upsampling operators. The consecutive layers in the network increase the resolution of the output from the network. The high-resolution features from the contracting path are merged with the upsampled output for achieving localization. A consecutive convolution layer learns to collect a high precise output, depending on this information.

This U-Net architecture comprises a large number of feature channels in the upsampling section to distribute the context information to the higher resolution layers. The expansive path and contracting path are found to be similar to each other. This network architecture does not require fully connected layers. The valid section of each convolution layer is used. Hence, continuous segmentation of the large images is enabled using an overlap-tile strategy. The input image is mirrored and missing context is concluded for predicting the pixels located in the border region of the image.

This tilting strategy is highly significant to apply the network to large images, as the resolution is limited by the memory of the Graphic Processing Unit (GPU). With the accessibility of small amount of training data, extra data augmentation is used by applying elastic deformation to the training images. Thus, the network architecture can learn invariance to the deformations, without the need for viewing the transformations in the amount of annotated image.

The input images and respective segmentation maps are used for training the U-Net structure with the implementation of stochastic gradient descent. The size of the output segmented image is lesser than the input image by a constant border width. This size variation is due to the unpadded convolutions. To reduce the overhead and efficient utilization of the GPU memory space, the large input titles are preferred over a large batch size. Thus, the large batch is reduced to a single image.

The energy function is evaluated using a pixel-based soft-max layer over the last feature map. This map is integrated with the cross entropy loss function. The soft-max is defined using the following equation:

$$p_k(x) = exp(a_k(x))/\left(\sum_{k'=1}^{K} \exp(a_{k'}(x))\right) \tag{10}$$

where $a_k(x)$ represents the activation in the feature channel 'k' at the pixel position $x \in \Omega$ with $\Omega \subset \mathbb{Z}^2$. K indicates the number of classes and $p_k(x)$ denotes the approximated maximum function, i.e. $p_k(x) \approx 1$ for the classes 'K' having maximum activation. $p_k(x) \approx 0$ for all other classes 'K'. The cross entropy corrects the deviation of the function $p_{l(x)}(x)$ from 1 using

$$E = \sum_{x \in \Omega} w(x) \log\big(p_{l(x)}(x)\big) \tag{11}$$

where $l: \Omega \rightarrow \{1, \ldots, K\}$ denotes the true label of each pixel and $w: \Omega \rightarrow \mathbb{R}$ represents a weight map introduced to provide significance to the pixels in the training. The weight map is calculated for segmenting each ground truth image. This compensates different frequency of the pixels from a specific class in the training dataset. The network is applied for learning the minor separation borders between the touching cells. The morphological operations are applied to compute the separation border. The weight map is calculated as

$$w(x) = w_c(x) + w_0 \cdot exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right) \tag{12}$$

where $w_c: \Omega \rightarrow \mathbb{R}$ indicates the weight map to balance the class frequencies, $d_1: \Omega \rightarrow \mathbb{R}$ represents the distance to the border of the nearest cell and $d_2: \Omega \rightarrow \mathbb{R}$ denotes the distance to the border of the second adjacent cell.

3.4 Feature Extraction

Feature extraction extracts the significant and relevant features from the segmented image, to distinguish the benign and malignant melanoma images.

(1) Vascular features

Basal Cell Carcinoma (BCC) is one among the common types of non-melanocytic skin cancer. The main cause of the BCC is the presence of small dilated blood vessels near the skin surface in the lesion. Epidermis is a pigmented layer of skin with the normal melanin amount that interferes with the visibility of the blood vessel. In addition, there are certain types of pigmented BCC. The vessel mask obtained from the segmentation process is used to differentiate the skin cancer from the benign lesions. A set of 12 vascular features are defined and extracted from each skin lesion among a set of BCC and non-BCC lesions [16]. Table 1 shows the vascular features.

(2) Color features

Color is another important feature descriptor for the classification of skin cancer. The color of the skin lesion region changes, if the particular region is affected. Relative color histograms in different color spaces are created to identify the skin cancer. The

Table 1 Vascular features

Feature	Description
Max_Length	Maximum length of all vessel segments in the lesion
Mean_Length	Average length of all vessel segments in the lesion
SD_Length	Standard deviation of length of all vessel segments in the lesion
Max_Width	Maximum width of all vessel segments in the lesion
Min_Width	Average width of all vessel segments in the lesion
SD_Width	Standard deviation of width of all vessel segments in the lesion
Max_Area	Maximum area of all vessel segments in the lesion
Mean_Area	Average area of all vessel segments in the lesion
Area_Ratio	Ratio of vessel area to the lesion area
Num_Branch	Number of vascular branches
Branch_Ratio	Ratio of number of branches to lesion area

Three-Dimensional (3D) histogram is created for the color spaces such as Red, Green, and Blue (RGB), LAB, Hue Saturation Value (HSV), HUE, and Opponent Color Space (OPP). The mixture coefficients of the red, green and blue colors represent the color components. The drawback of the color spaces is not uniform. It provides a high correlation among these color channels. Different color representations are proposed to mitigate these drawbacks. The edges are reinforced in the biologically inspired color spaces. For these histograms $8 \times 8 \times 8 = 512$ color bins are generated and considered as a single feature vector. This vector is applied as an input to the classifier [34].

(3) Texture features

Two-Dimensional (2D) wavelet transformation in LUV color space for each block within an image is applied to extract the texture features. The algorithm describes the steps for extracting the features of a single region. The segmented region is padded with the black pixels to form a minimum rectangle. The padding image is disintegrated into $m * m$ pixel blocks. The black blocks are removed completely for strengthening the captured features of the generated regions. The algorithm is a modified version of the method [35], as it does not cluster on the subspace (t1, t2, t3) to find the regions. Only the 'L' portion of an input region is considered in the 2D wavelet transformation. Each returned band coefficient is a quaternary vector. The moments of all blocks of LH, HL and HH bands are computed. The mean value of them is effective for stating the texture features [36]. Other features (f7, f8, f9) are computed to define the shape features. The shape and texture features are extracted efficiently. As per block, wavelet transformation is performed, the blank blocks are rejected easily to obtain more accurate representations for the irregular regions [37].

Feature Extraction Algorithm

Input: Image of padding region in RGB color space IM; Block size 'm'
Feature vector for input image, F
Step 1: Divide IM_LUV into blocks of size $m * m$
Step 2: $f1 = mean(IM\_LUV.L)$ //Feature 1
Step 3: $f2 = mean(IM\_LUV.U)$ //Feature 2
Step 4: $f3 = mean(IM\_LUV.V)$ //Feature 3
Step 5: $IM_{LUV} = colorspace('RGB \Longrightarrow LUV', IM)$;
Step 6: **for** each block 'b' in IM_LUV **do**
Step 7: **if** b is all black **then**
Step 8: discard this block
Step 9: **end if**
Step 10: $[LL, LH, HL, HH] = 2D - DWT(b, L)$
Step 11: $t1 = (1/4 * \sum_{i=1}^{4} LH_i^2)^{1/2}$
Step 12: $t2 = (1/4 * \sum_{i=1}^{4} HL_i^2)^{1/2}$
Step 13: $t3 = (1/4 * \sum_{i=1}^{4} HH_i^2)^{1/2}$
Step 14: $t1, t2, t3$ *add to* S_LH, S_HL *and* S_HH
Step 15: **end for**
Step 16: $f4 = mean(S\_LH)$ //Feature 4
Step 17: $f5 = mean(S\_HL)$ //Feature 5
Step 18: $f6 = mean(S\_HH)$ //Feature 6
Step 19: f7, f8, f9 are normalized inertia of order 1, 2, 3
Step 20: $F = (f1, f2, f3, f4, f5, f6, f7, f8, f9)$
Step 21: **return** F;

(4) RGB histogram features

An image histogram describes the tonal distribution in an image, in a graphical format. The tonal differences and the number of pixels in a particular tone are depicted in the horizontal axis and vertical of the histogram, respectively. The color histogram is used for describing the occurrence rate of each color in an image. In order to obtain the RGB histogram features, the number of bins is fixed to 16 for individual subbands in the image. The total number of bins is 4096, due to the presence of red, green and blue components in the image [1].

(5) Depth features

3D reconstruction is vital for estimating the depth of the skin lesions. The depth of the skin is computed using the defocus occurrence at the edges. The input skin images lesions are reblurred using a Gaussian function. The edge defocus is obtained as a ratio between the gradient magnitude of the input image and the reblurred version. The blur observed at the edges is propagated to the whole image. This enables calculation of the depth maps. An ideal step edge model considering the edge is located at $x = 0$. This model is defined as

$$\varepsilon(x) = A\mathcal{S}(x) + \mathcal{O} \tag{13}$$

where

A denotes the amplitude and
$\mathcal{O}$ represents the offset.
$\mathcal{S}(x)$ indicates the step function.

The sharp input image is convoluted with a point spread function to obtain defocus blur. A Gaussian function $g(x, s)$ is used for approximating the spread function. The standard deviation 's' in the Gaussian function is directly proportional to the circle of confusion (c). It is defined as $s = kc$.

The blurred edge obtained using the edge model and Gaussian function is defined as $\mathcal{B}(x) = \varepsilon(x) \otimes g(x, s)$. 'I' represents the input skin lesion image. A 2D isotropic Gaussian function $I^{\mathcal{B}}$ is used for reblurring the input image. The gradient magnitude of the input image along the 'x' and 'y' directions is defined as

$$\|\nabla I(x, y)\| = \sqrt{\nabla I_x^2 + \nabla I_y^2} \tag{14}$$

where ∇I_x and ∇I_y denote the gradient along the 'X' and 'Y' directions, respectively.

The gradient magnitude of the blurred image is calculated in the same way. The gradient magnitude ratio between the edge locations of the input image and reblurred image is $\frac{\|\nabla I(x)\|}{\|\nabla I^{\mathcal{B}}(x)\|}$. A sparse depth map $\overline{D(x)}$ is created by estimating the occurrence of the blur scale at each location of the edge. The inaccurate blur estimates at the certain edge locations are removed using a combined bilateral filter and input image as a reference. The resulting sparse depth map $\overline{D(x)}$ is used to obtain a full depth map $D(x)$ and the defocus blur estimates are propagated to the entire image. The full depth map is obtained by using the image interpolation based on matting Laplacian approach [38]. The optimal depth map 'D' is obtained as

$$(\mathcal{L} + \lambda\mathcal{M})D = \lambda\mathcal{M}\overline{D} \tag{15}$$

where $\mathcal{L}$ denotes the matting Laplacian, $\mathcal{M}$ represents the diagonal matrix and λ indicates the scalar balance factor. D and $\overline{D}$ are the vector representations of the full and sparse depth maps, respectively [39]. Table 2 shows the vascular features. Table 3 depicts the texture and color features. The RGB histogram features and depth features are presented in Tables 4 and 5.

3.5 RVM Classifier

The RVM classifier models the probability distribution of the class label of an input vector 'x' using the logistic regression defined as follows:

Table 2 Vascular features

S. no.	Features	Normal	Abnormal
1	Max_Length	0.8625	0.5348
2	Mean_Length	0.5421	0.2654
3	SD_Length	0.0917	0.0087
4	Max_Width	0.8456	0.5741
5	Min_Width	0.4328	0.2365
6	SD_Width	0.0854	0.0041
7	Max_Area	0.7541	0.5482
8	Mean_Area	0.3254	0.1756
9	Area_Ratio	0.0254	0.0154
10	Num_Branch	24	31
11	Branch_Ratio	0.5435	0.3214

Table 3 Texture and color features

S. no.	Features	Normal		Abnormal	
		Min	Max	Min	Max
1	Contrast	0.44	0.49	0.27	0.47
2	Correlation	0.97	0.97	0.98	0.99
3	Dissimilarity	0.16	0.21	0.06	0.10
4	Energy	0.32	0.40	0.44	0.51
5	Entropy	1.48	1.76	0.75	1.09
6	Homogeneity	0.92	0.94	0.97	0.99
7	Maximum probability	0.53	0.62	0.59	0.63
8	Inverse difference moment	0.99	0.99	1.00	1.00

Table 4 RGB histogram features

S. no.	Features	Normal		Abnormal	
		Min	Max	Min	Max
1	RGB histogram (mean value)	0.0254	0.5684	0.0751	0.4325

Table 5 Depth features

S. no.	Features	Normal		Abnormal	
		Min	Max	Min	Max
1	3D depth features (mean value)	0.0358	0.2489	0.0576	0.6871

$$p(d = 1|x) = \frac{1}{1 + exp(-f_{RVM}(x))} \tag{16}$$

where $f_{RVM}(x)$ represents the classifier function described as

$$f_{RVM}(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) \tag{17}$$

where $K()$ denotes a Kernel function and $x_i, where\, i = 1, 2, \ldots, N$ represents the training samples. The parameters α_i in the classifier function are determined using the Bayesian estimation and are assumed to be statistically independent. A sparse prior is introduced on α_i. These parameters follow a zero-mean Gaussian distribution with variance λ_i^{-1}. A hyper-prior is assumed on the variance for enforcing the parameters to be highly concentrated around zero. This leads to few non-zero terms in the classifier function. The parameters are obtained by improving the posterior distribution of the class labels for the given input vectors. This is equal to the maximization of the objective function defined as follows:

$$J(\alpha_1, \alpha_2, \ldots, \alpha_N) = \sum_{i=1}^{N} \log p(d_i|x_i) + \sum_{i=1}^{N} \log p\left(d_i|\lambda_i^*\right) \tag{18}$$

where the first summation term relates to the likelihood of the class labels and second term relates to the prior on the parameters. λ_i^* indicates the maximum aposteriori estimate of the hyper-parameter λ_i. The samples associated with the non-zero coefficients α_i called as relevance vectors can only contribute to the decision function.

The Kernel function is used for forming the expansion basis functions for the classifier function. The Mercer's condition [40] states that the Kernel function should be a positive integral operator. For every square-integrable function $g(\cdot)$ defined on R^n, the Kernel satisfies the condition described in the following equation

$$\iint K(x, y)g(x)g(y)dxdy \geq 0 \tag{19}$$

The polynomial Kernel $P_K(x, y)$ and Radial Basis Function (RBF) $R_K(x, y)$ are used to satisfy the Mercer's condition. They are defined as

$$P_K(x, y) = \left(x^T y + 1\right)^p$$

where p>0 is a constant that defines the order of the kernel.

The RBF kernel is calculated as

$$R_K(x, y) = exp\left(-\frac{||x - y||^2}{2\sigma^2}\right) \tag{20}$$

where $\sigma > 0$ is a constant that denotes the width of the Kernel [41].

4 Performance Analysis

4.1 Experimental Results

Figure 3 represents skin lesion input image. This lesion image is decomposed into hemoglobin and melanin channel as shown in Fig. 4 and Fig. 5, respectively using the Fast ICA algorithm. After that, melanin channel is applied to U-Net segmentation, resulting in red area segmentation leads to lesion as demonstrated in Fig. 6. This segmented ground truth image is overlaid on original image to better visual understanding of this detection, which is described in Fig. 7.

4.2 Feature-Based Analysis

In the feature-based analysis, our proposed work is compared with the melanoma classification using texture and color features [34] using 150 dermoscopy images. The sensitivity, specificity, and classification accuracy of the Gray-Level Co-occurrence

Fig. 3 Input image

Fig. 4 Hemoglobin channel

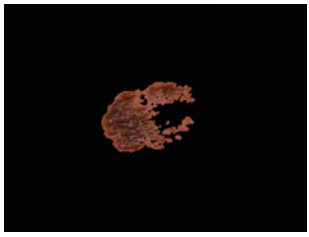


Fig. 5 Melanin channel

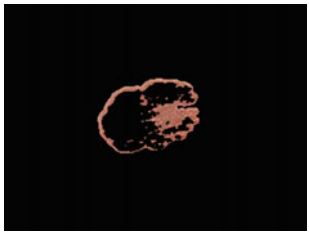


Fig. 6 Binary mask

Fig. 7 Lesion detection

Matrix (GLCM) that extracts texture features are 72.9%, 71.1%, and 76%, respectively. The classification accuracy of the RGB and HSV color space is 92%. The sensitivity and specificity of the RGB and HSV color space are 86.2% and 100%, respectively. The sensitivity, specificity and classification accuracy of OPP is 84.2%, 95.3%, and 89%, respectively. The combination of texture feature and RGB color space yields better sensitivity, specificity and classification accuracy of 88.2%, 85.5% and 93.1%, respectively. The combination of multiple features such as vascular, color, texture, depth and RGB features provides maximum classification accuracy of 95.2% and sensitivity of 89%. Figure 8 shows the comparative analysis of the feature extraction techniques.

4.3 Segmentation-Based Analysis

In the segmentation-based analysis, our proposed work is compared with the segmentation techniques such as SegNet (basics) [42], SegNet (VGG), FCN, FCN-MRF [33], and Deep Convolutional (DC) Network [43]. The proposed work is evaluated using the public ISBI 2016 dataset [30] comprising 900 training dermoscopic images and 379 testing images. The size of the skin images varies from 1022×767 to 4288×2848 pixels, and the ground truth image is provided by the medical experts. This dataset contains 727 benign and 173 melanoma images. Figure 9 shows the comparative analysis of the existing and proposed segmentation techniques. The performance metrics such as Jaccard Coefficient (JC), Dice Similarity Coefficient (DSC), Average Precision (AP) and Accuracy (ACC) calculate the spatial overlap between the ground truth and the obtained segmented binary mask. The U-Net based segmentation yields maximum JC, DSC, AP, and ACC of about 82.4%, 91.2%, 94.1%, and 93.5%, respectively.

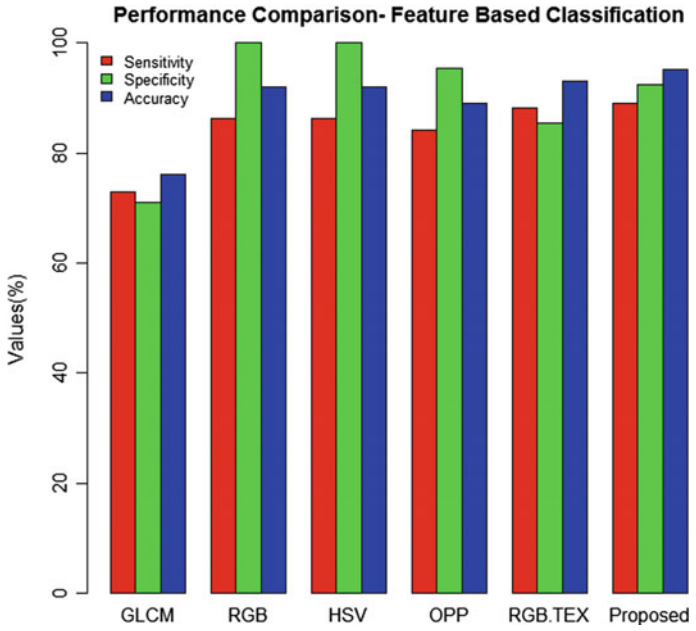


Fig. 8 Comparative analysis of feature extraction techniques

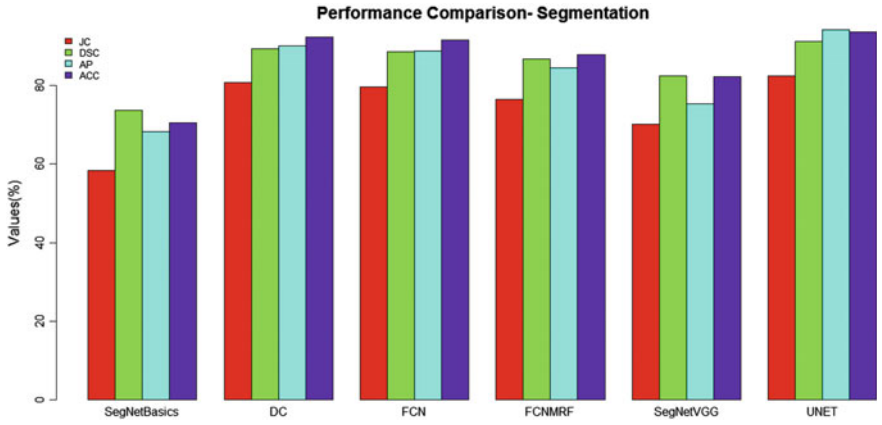


Fig. 9 Comparative analysis of existing and proposed segmentation techniques

4.4 Classification-Based Analysis

In the classification-based analysis, vascular feature with RVM-based classification is compared with the vascular features with Basal Cell Carcinoma (BCC) classification [16]. Figure 10 illustrates the comparative analysis of vascular features with RVM-

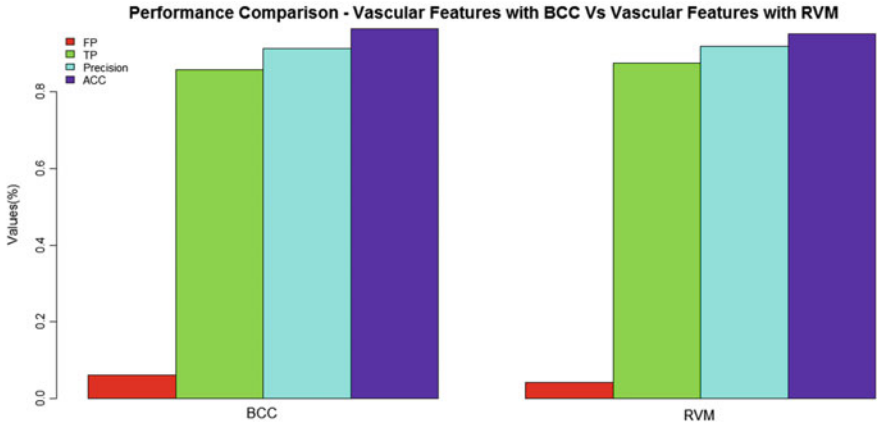


Fig. 10 Comparative analysis of BCC and RVM classifiers

based classification with the vascular features with Basal Cell Carcinoma (BCC) classification using the dataset obtained from Atlas of dermoscopy by Argenziano [16], University of Missouri and Vancouver Skin Care Centre. The Atlas, University of Missouri and Vancouver Skin Care Centre comprise images of 768 × 512 pixels, 1024 × 768 pixels, and 1930 × 1779 pixels. The remaining 659 images comprise 299 BCC versus 360 non-BCC lesions. The False Positive (FP), True Positive (TP), precision and accuracy of the existing classifier are 0.061, 0.859, 91.4% and 96.5%. The vascular feature with RVM classification yields maximum TP, precision, and accuracy of about 0.875, 92% and 95.2% and minimum FP is 0.042.

5 Conclusion

In this paper, a fully automatic segmentation of dermoscopic image and extraction of multiple features for efficient diagnosis of melanoma is presented. Compared to the previous methods, this method extracts shape, color, and texture and vasculature information from the input dermascopic image to yield better diagnosis results. The combination of multiple features such as vascular, color, texture, depth and RGB features provides maximum classification accuracy and sensitivity. The U-Net based segmentation yields maximum Jaccard Coefficient, Dice Similarity Coefficient, Average Precision and Accuracy than the existing segmentation techniques. The vascular feature with RVM classification yields maximum TP, precision and accuracy and minimum FP. Hence, our proposed work is a potential tool for the quantification of vasculature features and can be applied in a wide range of dermatology applications.

References

1. Sumithra R, Suhil M, Guru D (2015) Segmentation and classification of skin lesions for disease diagnosis. Procedia Comput Sci 45:76–85
2. A. C. Society (2018) Key statistics for melanoma skin cancer. https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html
3. A. C. Society (2018) Survival rates for melanoma skin cancer, by stage. https://www.cancer.org/cancer/melanoma-skin-cancer/detection-diagnosis-staging/survival-rates-for-melanoma-skin-cancer-by-stage.html
4. Oliveira RB, Marranghello N, Pereira AS, Tavares JMR (2016) A computational approach for detecting pigmented skin lesions in macroscopic images. Expert Syst Appl 61:53–63
5. Korotkov K, Garcia R (2012) Computerized analysis of pigmented skin lesions: a review. Artif Intell Med 56:69–90
6. Smith L, MacNeil S (2011) State of the art in non-invasive imaging of cutaneous melanoma. Skin Res Technol 17:257–269
7. Abbas Q, Celebi ME, Garcia I (2012) A novel perceptually-oriented approach for skin tumor segmentation. Int J Innov Comput Inf Control 8:1837–1848
8. Schaefer G, Rajab MI, Celebi ME, Iyatomi H (2011) Colour and contrast enhancement for improved skin lesion segmentation. Comput Med Imaging Graph 35:99–104
9. Abbas Q, Garcia IF, Emre Celebi M, Ahmad W, Mushtaq Q (2013) A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images. Skin Res Technol 19
10. Ganster H, Pinz P, Rohrer R, Wildling E, Binder M, Kittler H (2001) Automated melanoma recognition. IEEE Trans Med Imaging 20:233–239
11. Celebi ME, Kingravi HA, Uddin B, Iyatomi H, Aslandogan YA, Stoecker WV et al (2007) A methodological approach to the classification of dermoscopy images. Comput Med Imaging Graph 31:362–373
12. Yuan Y, Chao M, Lo Y-C (2017) Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance. IEEE Trans Med Imaging 36:1876–1886
13. Cavalcanti PG, Scharcanski J (2013) Macroscopic pigmented skin lesion segmentation and its influence on lesion classification and diagnosis. In: Color medical image analysis. Springer, pp 15–39
14. Majumder SK, Ghosh N, Gupta PK (2005) Relevance vector machine for optical diagnosis of cancer. Lasers Surg Med 36:323–333
15. Benazzi C, Al-Dissi A, Chau C, Figg W, Sarli G, de Oliveira J et al (2014) Angiogenesis in spontaneous tumors and implications for comparative tumor biology. Sci World J 2014
16. Kharazmi P, AlJasser MI, Lui H, Wang ZJ, Lee TK (2017) Automated detection and segmentation of vascular structures of skin lesions seen in dermoscopy, with an application to basal cell carcinoma classification. IEEE J Biomed Health Inform 21:1675–1684
17. Emre Celebi M, Wen Q, Hwang S, Iyatomi H, Schaefer G (2013) Lesion border detection in dermoscopy images using ensembles of thresholding methods. Skin Res Technol 19
18. Peruch F, Bogo F, Bonazza M, Cappelleri V-M, Peserico E (2014) Simpler, faster, more accurate melanocytic lesion segmentation through meds. IEEE Trans Biomed Eng 61:557–565
19. Zhou H, Li X, Schaefer G, Celebi ME, Miller P (2013) Mean shift based gradient vector flow for image segmentation. Comput Vis Image Underst 117:1004–1016
20. Sadri AR, Zekri M, Sadri S, Gheissari N, Mokhtari M, Kolahdouzan F (2013) Segmentation of dermoscopy images using wavelet networks. IEEE Trans Biomed Eng 60:1134–1141
21. Xie F, Bovik AC (2013) Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm. Pattern Recognit 46:1012–1019
22. Celebi ME, Wen Q, Iyatomi H, Shimizu K, Zhou H, Schaefer G (2015) A state-of-the-art survey on lesion border detection in dermoscopy images. Dermoscopy Image Anal 97–129

23. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K et al (2016) Melanoma detection by analysis of clinical images using convolutional neural network. In: 2016 IEEE 38th annual international conference of the engineering in medicine and biology society (EMBC), pp 1373–1376
24. Esteva A, Kuprel B, Thrun S (2015) Deep networks for early stage skin disease and skin cancer classification. Project report, Stanford University
25. Jafari MH, Nasr-Esfahani E, Karimi N, Soroushmehr S, Samavi S, Najarian K (2016) Extraction of skin lesions from non-dermoscopic images using deep learning. arXiv:1609.02374
26. Jafari MH, Karimi N, Nasr-Esfahani E, Samavi S, Soroushmehr SMR, Ward K et al (2016) Skin lesion segmentation in clinical images using deep learning. In: 23rd international conference on pattern recognition (ICPR), pp 337–342
27. Li Y, Shen L (2018) Skin lesion analysis towards melanoma detection using deep learning network. Sensors 18:556
28. Kawahara J, Hamarneh G (2017) Fully convolutional networks to detect clinical dermoscopic features. arXiv:1703.04559
29. Jaleel JA, Salim S, Aswin R (2013) Computer aided detection of skin cancer. In: International conference on circuits, power and computing technologies (ICCPCT), pp 1137–1142
30. Gutman D, Codella NC, Celebi E, Helba B, Marchetti M, Mishra N et al (2016) Skin lesion analysis toward melanoma detection: a challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). arXiv: 1605.01397
31. Tsumura N, Haneishi H, Miyake Y (1999) Independent-component analysis of skin color image. JOSA A 16:2169–2176
32. Agrawal P, Shriwastava S, Limaye S (2010) MATLAB implementation of image segmentation algorithms. In: 2010 3rd IEEE international conference on computer science and information technology (ICCSIT), pp 427–431
33. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
34. Kavitha J, Suruliandi A (2016) Texture and color feature extraction for classification of melanoma using SVM. In: International conference on computing technologies and intelligent data engineering (ICCTIDE), pp 1–6
35. Chen Y, Wang JZ (2004) Image categorization by learning and reasoning with regions. J Mach Learn Res 5:913–939
36. Gersho A (1979) Asymptotically optimal block quantization. IEEE Trans Inf Theory 25:373–380
37. Zhang G, Shu X, Liang Z, Liang Y, Chen S, Yin J (2012) Multi-instance learning for skin biopsy image features recognition. In: 2012 IEEE international conference on bioinformatics and biomedicine (BIBM), pp 1–6
38. Levin A, Lischinski D, Weiss Y (2008) A closed-form solution to natural image matting. IEEE Trans Pattern Anal Mach Intell 30:228–242
39. Satheesha T, Satyanarayana D, Prasad MG, Dhruve KD (2017) Melanoma is skin deep: a 3D reconstruction technique for computerized dermoscopic skin lesion classification. IEEE J Trans Eng Health Med 5:1–17
40. Ripley BD (1996) Pattern recognition via neural networks. A volume of Oxford graduate lectures on neural networks, title to be decided. Oxford University Press. http://www.stats.ox.ac.uk/ripley/papers.html
41. Wei L, Yang Y, Nishikawa RM, Wernick MN, Edwards A (2005) Relevance vector machine for automatic detection of clustered microcalcifications. IEEE Trans Med Imaging 24:1278–1285
42. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39:2481–2495
43. Bozorgtabar B, Sedai S, Roy PK, Garnavi R (2017) Skin lesion segmentation using deep convolution networks guided by local unsupervised learning. IBM J Res Dev 61:6: 1–6:8

Secured Transmission of Medical Images in Radiology Using AES Technique

Pavithra Prabhu and K. N. Manjunath

Abstract Medical imaging technology results in more than thousands of images per patient and transmitting them securely over an insecure network is a challenging task still. Maintaining the data integrity against intruder is important. This paper discusses a hybrid method which combines image processing and cryptography to make sure that all images are securely transmitted. In the proposed method, a Digitally Reconstructed Radiograph (DRR) is generated from all the images of a 3D volume of the patient, and then divided the DRR image into four equal quadrants. The zigzag pattern was applied to all these sixteen quadrants. Each quadrant was separately encrypted in block cipher mode using AES algorithm. At the receiver side, the DRR was regenerated from all the transmitted images and was compared with the deciphered blocks using histogram comparison of each block. The method was applied to CT images of 40 patients of brain tumor, CT colonography, and nasopharynx dataset. The image injection techniques were applied and tested the result with the addition of image, deletion of an image and through modification of Hounsfield values.

Keywords DICOM · DRR · Zigzag, CT scan · AES · Block cipher

1 Introduction

Medical images play a vital role in radio diagnosis and telesurgery. The patient scan images are frequently accessed in *intra-hospital* (through PACS server) and *inter-hospital* networks. Exchange of medical images over a public or private network is always vulnerable. Ensuring the security of exchanged medical images is made

P. Prabhu · K. N. Manjunath (✉)
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India
e-mail: kn.manjunath@ieee.org

P. Prabhu
e-mail: pavithraprabhu104@gmail.com

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_10

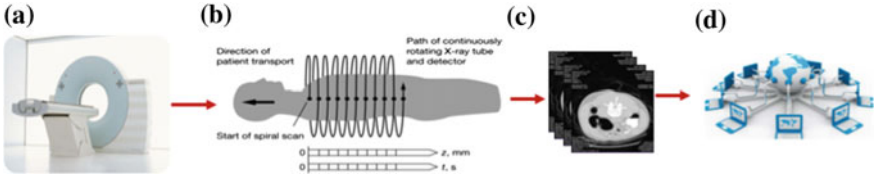


Fig. 1 The transmission of medical images over the insecure network after image acquisition (*Image Source* Wili Kalender [2])

compulsory by the government in the form of Health Insurance Portability and Accountability Act (HIPAA). Security implementation of an application requires three main characteristics. This includes **Confidentiality**: a concept of ensuring that the data is not disclosed to unauthorized people, **Integrity**: preventing the data modification, **Availability**: Making the data available at any time when a query is placed. At present, digital watermarking and cryptography are the two major methods adopted by the medical image transmission.

The Digital Imaging and Communications in Medicine (DICOM) standard is a technical framework and protocol for storing and exchange of radiology images [1]. It includes different modalities like Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and ultrasound images. For example, in CT Colonography cases, the patient scan (Fig. 1a, b) results in more than 3000 images in both Supine and Prone position (Fig. 1c). And these numbers depends mainly on the selection of Slice Thickness (ST). The security of all those "n" number of images is very important when transmitted over the insecure network. Because image falsification or tampering by the adversary (I) results in the incompleteness of patient data which is unethical and it may mislead the diagnosis.

At present, the transmission of each image is being confirmed by checksum calculation at the image level. At sender side, checksum value is calculated and attached to the image, at receiver, it is recalculated for the transmitted image and the value is compared against the stored checksum in the received image. Still, it does not guarantee that all the images of a patient are transmitted or not. In this paper, we discuss maintaining the data integrity of all images at the patient level.

2 Literature Review

The current state of secure transmission of medical images is discussed in this section. The methodology discussed by Hiba [3] involves a combination of two technologies. The technique simultaneously embeds two watermarks which involves one in the spatial domain and one in the encrypted domain. The algorithm makes use of reversible data hiding based on the Histogram Shifting method to prevent any distortion in the medical image to ensure the right diagnosis. Al-Haj et al. [4] have proposed an algorithm which uses internally generated symmetric keys and hash values, which

was aimed at overcoming the limitations of Kobyashi's [5] scheme. The drawback of Kobyashi's scheme was, only pixel data was encrypted, and not the header details. Natsheh et al. [6] also have proposed an encryption approach for multi-frame DICOM images where it encrypted pixel data along with the header information. The method included an XOR based AES encryption algorithm. The first encrypted image was used as a key to the rest of the images in the frame. This resulted in less encryption–decryption time along with ensuring confidentiality. Another crypto watermarking technique was implemented by Puech and Rodrigues [7] which is based on the combination of public–private keys and secret key ciphering, and watermarking. It uses RSA algorithm which is a public key cryptosystem. The encryption procedure along with a secret key was applied to the image data and encrypt the secret key with an encryption method based on public–private keys. They generated the public and private keys with the algorithm. But when huge data is involved RSA key generation might not be a good approach. To overcome the issue regarding the confidentiality, Blackledge [8] has provided a Stegacryption method which includes both encrypting the data and using steganographic methods of hiding it in an image. The proposed method removes private data, encrypts it, and then embeds it into the DICOM image in an undetectable way.

The standard algorithms such as AES, DES consumes lot of execution time and memory. The author [9] introduced a quaternion technique along with modular arithmetic operation and counter mode encryption which reduced computational time. Dalel [10] has proposed for watermarking both LSB replacement and Quantization Index Modulation (QIM) was used along with $RC4$ algorithm. In quantifying the components of one image according to a set of codebooks based quantizers on inserting a message. The author in [11] implemented the same technique where along with $RC4$, and the algorithm was also used for better encryption output.

In spite of well-established crypt techniques at the image level, we did not find any literature which discusses the transmission of a complete set of images at the patient level which guarantees secured transmission. We need a formal method in CIA which should ensure the correct transmission and receive at the other end. The main aim of the proposed methodology is to assure the integrity of all the medical images transmitted over an unsecured network.

3 Materials and Methods

The required patient dataset for the study purpose was downloaded from National Cancer Institute, the USA [11]. The image acquisition parameters are 120 kVP, 200 mA, and 512×512 image resolution. The abdomen scan, brain tumor, and nasopharynx cases were considered. The work was implemented in C# (.NET framework 4.5) with Microsoft visual studio 2012 express edition. The cryptography library from C# and the image processing libraries (for DICOM image validation and the creation of DRR image) has been considered from the recently published work by Manjunath et al. [12].

4 Methodology

The proposed methodology (Fig. 2) for achieving the objective of data integrity involves two phases. First is the creation of Digitally Reconstructed Radiograph (DRR) from the 3D volume and in the second, application of cryptographic techniques on this image. It is known that the pixels of CT images are stored in Hounsfield units. Any kind of processing on the image must be done only on Hounsfield units. Only for the display purpose, they are converted into greyscale using window center and window width.

4.1 Step 1: DICOM Image Validation

In the first step, each DICOM image is validated for its header attributes as per the standards defined in [PS 3.3]. Primarily, type 1 and type 2 attributes are checked. This ensures the data completeness of all the images. If any attributes are missing, such dataset is not considered for further steps. After DICOM validation, the 2D slices are sorted according to the slice location, and then the 3D volume is reconstructed in a patient coordinate system using (x, y, and z axis in Fig. 2) linear interpolation technique to achieve cubic voxels.

4.2 Step 2: DRR Image Creation

DRR is an artificial 2D X-ray image generated from the 3D volume. It is one representative candidate of the entire 3D volume which shows the anatomy distribution on a 2D image. It is extracted from 3D volume (Fig. 2) in Anterior–Posterior direction (coronal view), Left–Right direction (sagittal view) and in a Superior–Inferior direction (axial view) (Fig. 3). Coronal and sagittal view images are visually more

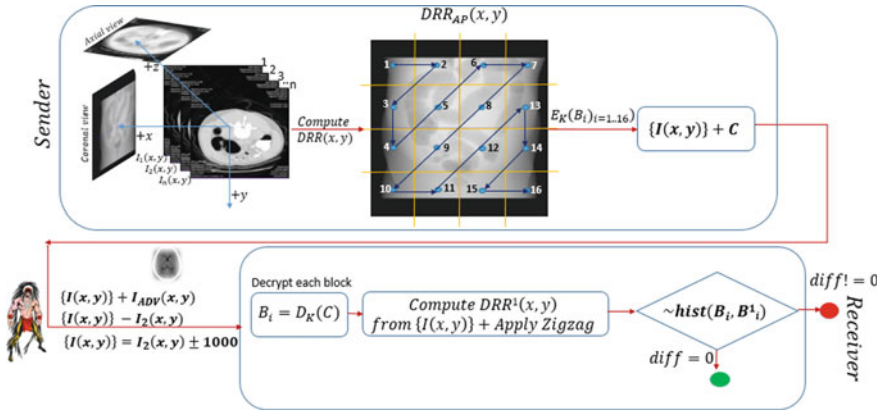


Fig. 2 The computation of DRR and applying the encryption scheme

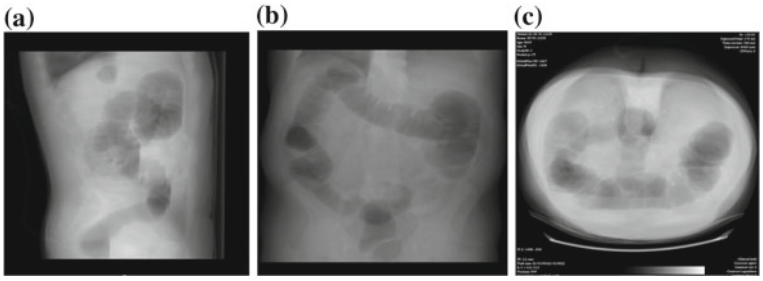


Fig. 3 The DRR image extracted in (a) LR direction, (b) AP direction, and (c) Superior to inferior direction

appealing compared to axial. The coronal image is used for encryption. Following pseudocode is used for computing DRR image.

```
  // for each row
  for r=0 to r<imageheight;

  // for each slice in z axis
  for slicebegin = zAxisBegin to zAxisEnd

  // consider the HU along the ray
  CTNumbers.Add (HU[r, stackbegin + stackindex,
slicebegin]);

  // calculate the average of pixel intensities
  DRR [r, stackbegin + stackindex] = Con-
vert.ToInt32 (aCTNumbers.Average ());
```

Similar to digital watermarking if this DRR image can be fused to all the CT images and then encrypt the pixel content, this approach would give better results as the DRR image is generated from these set of images, and the DRR pixels are fused in turn to the same CT images. Since the intruder always tries to decode the key information to decrypt the information, if he does so, then also it is not possible to decode the logic of how the image pixels are modified. But this technique is computationally expensive.

4.3 Step 3: Apply Zigzag Pattern for Blocks

Block level scrambling is done by applying the zigzag pattern on DRR image. Application of zigzag pattern ensures that blocks are not sent in sequential order. The DRR image is divided into 16 equal blocks (4 rows and 4 columns) of the same size. This

results in small blocks of size 128 ∗ 128 (i.e., m/4 * n/4) in case of 512 ∗ 512 images where m and n are the rows and columns of the image. Most of the radiology images have an image resolution of 512 ∗ 512 (in CT) or 256 ∗ 256 (in MRI) or even 1024 ∗ 1024 (high-resolution CT) also. Irrespective of the image dimension, the logic of 16 blocks works fine. There is no specific logic behind splitting the entire image into 16 blocks. If more accuracy is required, then even 64 blocks can also be considered. For each of the obtained block, the encryption is applied in the next step. The data structure and its contents for storing these blocks are shown below.

```
IDictionary<int,int[,]>TheBlocks

Example:

    TheBlocks [0] = {1, SubImage [128,128]}
    TheBlocks [1] = {2, SubImage [128,128]}
    TheBlocks [2] = {5, SubImage [128,128]}
```

and so on. Here, index 2 of the array corresponds to the 5th block in the zigzag pattern and this block size is 128 × 128 as image of 512 × 512 is divided into 16 quadrants.

4.4 Step 4: Encryption Algorithm

For each of the block obtained in step 3, the AES algorithm is applied with user-defined key value. Through this feature, it will be a hard problem for the attacker to trace the algorithm. The output of the encryption is stored in a dictionary with the following data structure.

```
//Data type
IDictionary<int,string>TheCipher
Int Counter=1;


//Logic
foreach (int[] Block_i in TheBlocks)
{
     Counter++;
     TheCipher.Add (Counter, AES(key, Block_i));
}

Example:
TheCipher[0] = {1,”AAAA2SDSDOWOEW23O”}
TheCipher[1] = {2,”AAAA2TR2Y544FRRRTO1”}
```

The block number and the ciphertext *C* is stored in this dictionary. These 16 encrypted blocks along with all the images of the dataset are then sent to the receiver.

4.5 Step 5: Transmission

Now all the images along with the cipher C are transmitted to the intended party over a network by assuming the network as insecure, and there always exists an adversary I. During transmission, the following three adversarial effects are implemented. These are the perfect cases of data integrity violation.

- Inserting an image $I_{ADV}(x, y)$ to the dataset {I(x,y)}. Here, I_{ADV} can be either the duplicated Image IOD from the existing dataset or it is an image from some other patient dataset.
- Deleting an existing image $I_2(x, y)$ from {I(x,y)}.
- Modifying the pixel intensities of an image $I_2(x, y)$ from the set of images {I(x, y)}. This is done by changing few pixel intensities randomly in the range [−1072 to +3096] as this is the HU range of a CT image.

In all these three cases, if we regenerate the DRR image, we can notice the difference in its pixel intensities.

4.6 Step 6: Decryption

Now the images along with C are received by the receiver. At the receiver side, the encrypted blocks are decrypted using the same key K used for Encryption. From the received set of images, the DRR image is regenerated, and the zigzag pattern is applied again. Now, these block of size 128×128 is compared individually with the decrypted blocks through histogram comparison (following pseudo code) and RMSE. If the **diff(hist1, hist2)** is equal to zero, or if the RMSE = 0, then the data is preserved, and there is no data integrity violation by I. Or else it proves histogram comparison has failed due to a mismatch in the DRR images. This mismatch is due to either modification of pixels or addition of I_{ADV} to the {I(x, y)} or due to deleting an image from the existing collection. DRR image concept is heavily used in many image processing applications to analyze the results after processing.

```
bool comparison=true;
foreach(Block B in Block_i)
{
    If hist(Block_i)!=hist(Block_i(M))
           Comparison&=false;
}
```

If the comparison flag is true by the end of all blocks comparison, then the data transmitted is secure. Since the histogram difference at image level may fail, we compared the histograms of all 16 blocks separately and computed the sum of differences.

5 Results and Discussion

The method was applied to forty datasets. The dataset included the CT images of brain tumor, nasopharynx, and CT colonography cases. The testing scenarios include injecting an image, deleting the existing image, and modifying the pixel values of an image. With any of these operations by the intruder, the blockwise comparison will fail, and modification of the data can be identified. Another advantage of sending the image blockwise is, there is a possibility that the intruder can modify the histogram of the image itself to fool the receiver. Through this, the attacker will make the receiver to believe that there is no data modification. To overcome this issue, the image is divided into blocks before sending to the receiver, and also the DRR is recomputed at the receiver for the set of images what it receives. Ultimately, the histograms are compared. This comparison resulted in a difference of zero since there was no modification of the data was done. In some cases, where all three adversarial effects were tested, the comparison resulted in a non–zero value which infers the data integrity violation. We did not check the header information as already the checksum calculation involves the header details. We preferred more on checking the integrity of pixel values. The different anatomy dataset and the SQL injection techniques results are shown in Fig. 4.

In all the cases, as part of the image injection mechanism, an axial image is added or removed (column 3, Fig. 4) from the existing dataset. Before injection lets us call the DRR computed as DRR^1 (column 1, Fig. 4), and after injection, let us call it as DRR^2 (column 2, Fig. 4). As explained in methodology, the DRR is computed at block level (of size 128×128) after applying the zigzag pattern. Since the DRR image pixel values are heavily depending on 2D axial slices, a slight modification in the collection of slices results in the modification of DRR and hence leads to $\mathrm{DRR}^1_{(\mathrm{Bi})} ! = \mathrm{DRR}^2_{(\mathrm{Bi})}$ (column 4, Fig. 4). Since the human visual system is limited to identify only few gray levels (16 levels), it is difficult to perceive the modified DRR in Column 4 of Fig. 4. The only possibility is checking the modified DRR through statistical comparison, i.e., through histogram comparison at block level.

Only in row 3, Fig. 4, where there was no injection, the DRR remains the same at the destination due to which the histogram difference result is zero.

Limitations: We did not test this method in a network environment. Instead, we modeled the adversary effects like the addition of image, deletion of image, and pixel modification and then checked the output within a simple client–server problem and logged all the message related to the results for the analysis purpose.

6 Conclusion

Many other literatures have discussed the state of the art encryption techniques at the image level. In this paper, we have discussed a block cipher algorithm to achieve data integrity while transmitting the complete set of images of a patient over an

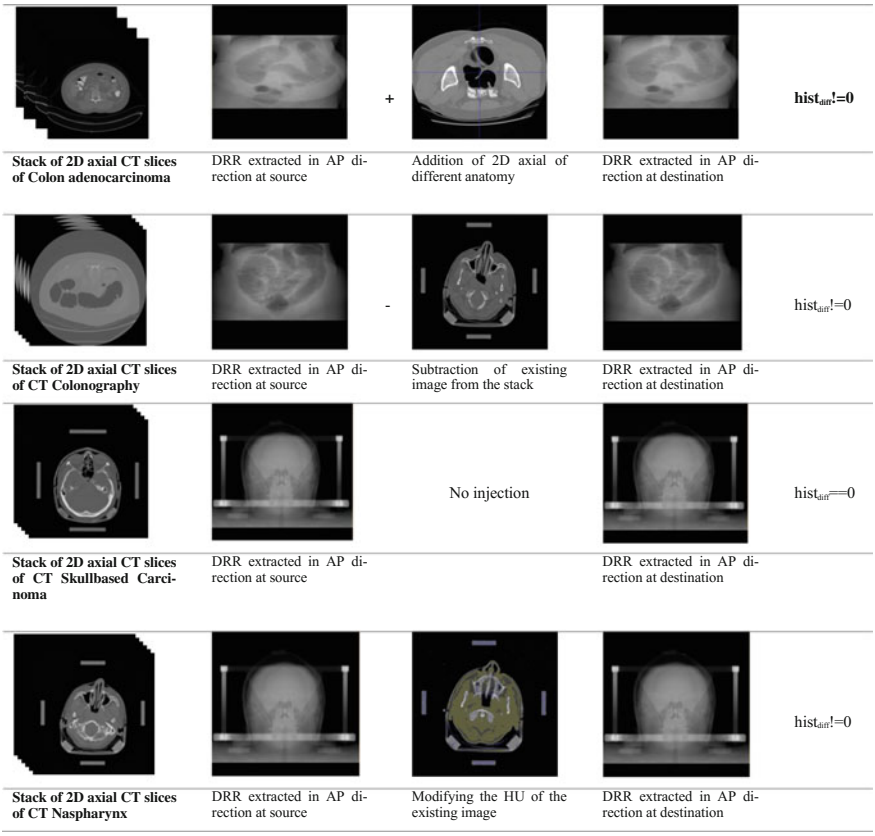


Fig. 4 The result of image injection techniques through addition and deletion of image from the dataset and the modification of pixel intensities in the image before transmitting to the destination. The difference of DRR of source dataset and the destination dataset is different in case of modification or else it is zero. Different anatomy datasets were considered during results analysis

insecure network. The DRR image is the ideal candidate to check any kind of injection technique as it is derived from the set of images. Any modification to the dataset pixels results in variation of pixel intensities of DRR image. And if such image comparison with original DRR fails, it infers that the data is modified by the adversary. The method was tested on twenty datasets with all adversarial act. This technique can be used to transmit radiology images in a hospital network or even to the outside world with due care on data integrity. The novelty of this hybrid approach is that combination of image processing and the cryptographic techniques have proved that patient dataset can be securely transmitted. More testing is pending in the real-time environment as well as with other imaging modalities like MRI and PET.

References

1. The DICOM Chapter 3 (2012) PS 3.3 [Online]. NEMA USA, Virginia, pp 390–1174. http://dicom.nema.org/standard.html. Accessed 19 June 2016
2. Kalender WA (2006) X-ray computed tomography. Phys Med Biol 51(13):29–43
3. Abdel-Nabi H, Al-Haj A (2017) Efficient joint encryption and data hiding algorithm medical images security. In: 8th international conference on information and communication systems (ICICS)
4. Al-Haj A, Abandah G, Hussein N (2015) Crypto based algorithms for secured medical image transmission. Inst Eng Technol (IET) J 9(6):365–373
5. Kobayashi L, Furuie S, Barreto P (2009) Providing integrity and authenticity in DICOM images: a novel approach. IEEE Trans Inf Technol Biomed 13(4):582–589
6. Natsheh QN, Li B, Gale AG (2016) Security of multi-frame DICOM images using X-OR encryption approach. In: International conference on medical imaging understanding and analysis, at Loughborough, UK
7. Puech W, Rodrigues M (2004) A new crypto-watermarking method for medical images safe transfer. In: 12th European signal processing conference
8. Blackledge J, Al-Rawi A, Tobin P (2014) Stegacryption of DICOM metadata. In: 25th IET Irish signals and systems conference 2014 and 2014 China-Ireland international conference on information and communications technologies (ISSC 2014/CIICT 2014)
9. Fathima Nasreen K, Chitra P (2016) A robust encryption and digital watermarking scheme for DICOM images using quaternion and DWT-SVD. In: Online international conference on green engineering and technologies (IC-GET), 2016
10. Bouslimia D, Coatrieuxa G, Rouxa C (2011) Joint encryption/watermarking algorithm for verifying the reliability of medical images: application to echo graphic images. In: Computer methods and programs in biomedicine in Elsevier, pp 45–47
11. www.cancerimagingarchive.net. National Cancer Institute (NCI) [Online]. https://public.cancerimagingarchive.net/ncia/login.jsf. Accessed 27 Feb 2016
12. Manjunath KN, Siddalingaswamy PC, GK Prabhu (2017) Measurement of smaller colon polyp in CT Colonography images using morphological image processing. Int J Comput Assist Radiol Surg (Springer) 1–11. http://dx.doi.org/10.1007/s11548-017-1615-4

A Review on Haze Removal Techniques

K. P. Senthilkumar and P. Sivakumar

Abstract Images captured in outer environment suffer from reduced scene visibility, reduced color contrast, and increased color fading. This may happen due to the presence of haze, fog, smoke, dust and noise present in the outer atmosphere. Haze is formed due to two basic processes, attenuation and air light. Attenuation diminishes the contrast of the image and air light makes the image to be whiter and hence the image captured is unclear. Haze removal or image dehazing is required in real-world weather conditions to obtain a fast and high-quality hazy free image which is used in various fields like satellite systems and aircraft systems. The intention of this review paper is to give a brief analysis on different haze removal techniques.

Keywords Haze removal · Image enhancement · Dark channel prior · Filtering Genetic algorithm

1 Introduction

The main goal in image processing is to analyze, understand, and obtain information from the image. But in certain cases, images taken in outer air atmosphere under abnormal conditions (Fog, haze, dust, and smoke) degrade the nature of pictures. So there is a reduced contrast and deteriorates in color of the image obtained. When the separation between the object and the camera is increased, the clarity of the image obtained is automatically reduced. The visibility of the image is reduced when it is having poor illumination.

K. P. Senthilkumar (✉)
Department of ECE, Kingston Engineering College, Vellore, India
e-mail: skkr8587@gmail.com

P. Sivakumar
Department of ECE, Karpagam College of Engineering, Coimbatore, India
e-mail: sivakumar.poruran@gmail.com

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_11

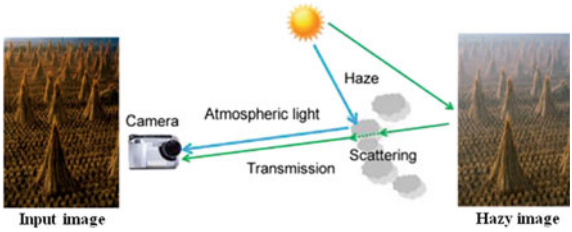


Fig. 1 Haze model. (Reproduced from Chen et al. [1])

1.1 Formation of Haze

Haze is an outer atmospheric phenomenon that partially covers the sky region and produces thin and very translucent effect. It is an aerosol having a radius of 2–10 μm and it is more brownish or bluish. Haze is produced due to two important phenomena attenuation and air light. The image captured under that situation got attenuated in the line of sight distance and reduces the color similarly air light effect adds the light from all other directions and it increases the whiteness in the image. The effect is shown in Fig. 1.

The presence of molecules count in the outer atmospheric region varies depending on the weather conditions. Haze, fog, and smoke come under the category of steady weather conditions, and snow and rain come under the dynamic weather conditions. In the haze environment, the illumination obtained in the image is varied because of considerable molecules count and the aerosols present in the outer environment. In haze environment, the images captured will have poor visibility and those images are considered as poor intensity and low contrast images. The content of the paper is organized in the following manner Sect. 2 gives the mathematical standard and haze removal methods; Sect. 3 elucidates various haze removal methods, Sect. 4 elucidates the study of various haze removal methods, Sect. 5 gives the various performance parameter calculation, Sect. 6 lists the various applications of haze removal methods, and Sect. 7 ends with the conclusion of the paper.

2 Mathematical Standard and Haze Removal Methodology

Haze model is universally used in several image processing, remote sensing, and different computer applications. This model is used for the image formation under very bad weather conditions, so image quality is very much reduced by the existence of haze, smoke, and fog.

Assume that the model of haze is linear, as per the definition of linearity, the position of the pixel in the image is changed. The mathematical model of hazy component is given by

$$X_C(P) = Z_C(P)t(P) + A_C(1 - t(p)) \tag{1}$$

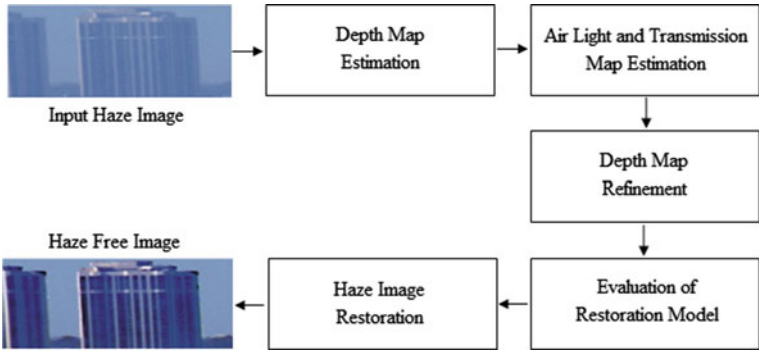


Fig. 2 Haze removal methodology

$$t(p) = e^{-\alpha d}(p) \tag{2}$$

where C is the color index of the channel (R, G, B), A_C indicates the global atmospheric light, Z_C indicates the image without haze, t is the medium transmission, $Z_C(P)t(P)$ gives the attenuation parameter, $A_C(1 - t(p))$ is the airlight, d(p) is the unknown depth information, and α is the scattering coefficient. The step-by-step haze removal from an image is explained above in Fig. 2. Air light and transmission map are evaluated in depth map estimation as the first step of haze removal methods. Depth estimation includes optical model, refined optical, restoration of perceptibility, preexistent dark channel, and preexistent color attenuation. Refinement of depth map is followed after depth estimation which is a technique used to reduce depth map errors, noise and undesired artifacts. Restoration of haze-free image is adopted using haze image restoration as the last step in the frames of reference and the mathematical standard for restoration is given by

$$Z_C(P) = X_C(P) + (1/t(p) - 1)(X_C(P) - A_C) \tag{3}$$

where 1/t(p) – 1 is amplification factor and $(X_C(P) - A_C)$ is the detail layer, and Z_C is the restored image. The objective of the method is to obtain the restored image Z from the haze image X. The restored image has improved contrast and it is free from haze and noise.

3 Haze Removal Techniques

Many haze removal methods were proposed to remove haze from single and multiple images. Haze removal methods have increased both neighborhood and total contrast of the scene and also it corrects the color distortion and provides depth information and visually pleasuring image. The hazy image and the dehaze images are shown in Fig. 3.

Fig. 3 Original input image and restored image. (Reproduced from Shen and Wu [2])

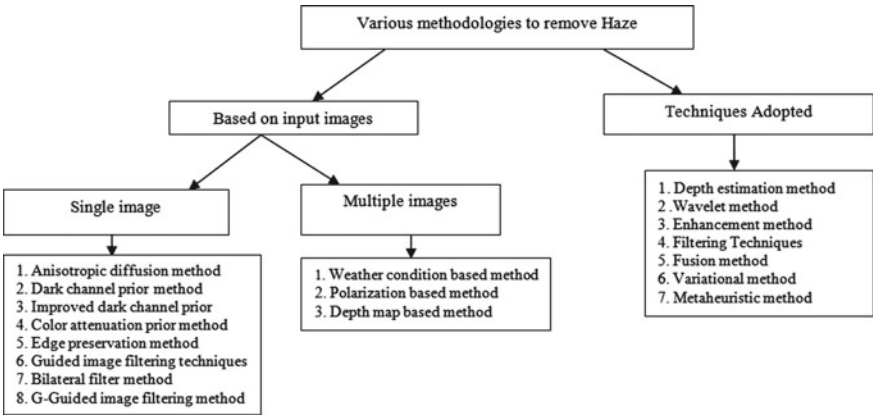


Fig. 4 Various haze removal methods

Dehazing techniques depend on the number of input images used and the methods are classified into (1) Multiple images dehazing and (2) Single image dehazing. Generally, removal of haze techniques are classified into various extended classes. Those methods are listed as (1) Depth estimation method (2) Wavelet method (3) Enhancement method (4) Filtering Techniques (5) Fusion method (6) Variational method, and (7) Meta-heuristic method. The various methods are shown in Fig. 4.

3.1 Multiple Images Based Haze Removal

In weather condition based method [3, 4], they used two or more scenes and they found the difference between the characteristics of the images but it does not give the better results. In polarization-based method [5], they used different scene images captured with various polarized filters but it has failed for dynamic images and images with dense haze. In depth based method [6], they used the 3D geometrical model and texture of the scene and also provided a good result, but it was not automatic. The comparison is shown in (Table 1).

Table 1 Multiple images based haze removal

Methods	Features	Drawbacks
Weather condition method	Used two or several images of the same scene under various climatic situations, also enhanced visibility was obtained	Impotent to handle the dynamic scene and does not give the result suddenly
Polarization method	Used several images taken with polarization filter captured in bad climatic situations and it has applied at any time instant	Unable to manage very dense hazy image and dynamic images
Depth model method	Used the 3D geometrical model and texture of the image and gave accurate results and scene depth	Not automatic

3.2 Single Image Based Haze Removal

Dark channel prior method (DCP) [7] was utilized to eliminate haze from images where some pixels in image have less intensity in any one color channel where the region of sky was not covered called as dark pixels, used to calculate the transmission map then they derived the refined transmission map to remove the haze effect. Here, transmission map was estimated accurately but it produced some halo effects in the output image. Modified dark channel prior method was the improved method of (DCP) which used bilateral filter for soft matting which smoothens the small-scale texture of the images. They utilized the DCP concept to calculate the atmospheric veil under the presence of sky region. The result obtained has greater efficiency, less execution time, and accurate results, but it produced halo artifacts in some regions and transmission map was not found accurately. In anisotropic diffusion method [8], the atmospheric light obtained by DCP method, improved contrast of the image was obtained without user interference. The air light was estimated accurately with contrast-enhanced image but visibility of the restored image was poor. Color attenuation prior method [9] was used to create an exclusive model of the scene depth information of the haze image, using supervised learning method. They obtained the recovered depth map which was used to recover the scene from the hazy image. The output obtained has high efficiency and good dehazing effect. Edge preserving decomposition method [10] was used to calculate the depth map for a hazy image which was based on two important principles modified color channel and simplified dark channel. The result obtained was accurate and was widely used in single image haze removal. The bilateral filter method [11] calculates the output of a pixel by its averaging of nearest pixel, it smoothens the haze image while conserving the boundary, and was widely used in noise reduction and image compression. The major drawback was that it produces gradient reversal artifacts in the output images.

The GIF [12] was derived from a regional linear model and generates the reference image which may be primary image or other different image and it preserves the edges of the image like bilateral filter which has O(N) computation time and has higher efficiency. The guided image filtering (GIF) and weighted guided image filtering does not preserve fine structure so globally guided image filtering was proposed [13] which was composed of a global structure transfer filter and a global edge preserving smoothing filter which was used to produce sharper images and visible fine structure. The comparison is shown in (Table 2).

Table 2 Single image based haze removal

Methods	Features	Drawbacks
Anisotropic diffusion method	It reduces haze without removing edges and lines from images and air light was estimated accurately	Transmission map was not estimated accurately and restored image visibility was poor
DCP method	Dark pixels was utilized to assess the depth map, also refined transmission map was estimated to remove the block effects	Halo effects were produced in the resultant image
Improved Dark channel prior method	Improved version of DCP method which used bilateral filter for matting. Estimation of air light was accurate and it has greater efficiency, less execution time	Halo effects were produced in some regions in resultant image and transmission map was not estimated accurately
Color attenuation prior method	A novel method for haze removal and powerful tool for scene depth estimation. It has high efficiency and good dehazing ratio	Average speed and large haze gradients
Edge preserving based method	The edges were preserved and accurate result was obtained	Average speed, halo artifacts, and large haze gradients
Bilateral filter method	Edges were preserved and air light was estimated accurately	Gradient reversal artifacts and poor speed
Guided image filter method	This filter has fast linear algorithm and computational complexity was independent of kernel size, also it preserved the edges	Halo artifacts problems and no fine structure preservation
Globally Guided image filter method	It is composed of a global structure transfer filter and global edge preserving smoothing filter, which preserves the fine structure of dehazed image	Color distortion problems

3.3 Techniques Adopted

In depth based estimation haze removal method, to evaluate the atmospheric veil, multi-scale tone method [14] was used to measure the nature and luminance of guidance image at different scale levels. But this method does not attain the true results for excessive hazy images. In wavelet method, improved wavelet transform [15] was used for efficient haze removal. In this method wavelet, transform was used in the initial stage to remove the haze and then Retinex algorithm was used to enhance the color performance and upgrade the color response in the restored image. In another method, fast wavelet transform technique [16] was introduced to improve the running time without any prior, and haze was completely removed and also sharpening of image was obtained simultaneously. In enhancement method, dark channel prior was used which utilized the soft matting method but it needs more memory and time and it was used for images which was very small only. In filtering method, the median filter and gamma correction [17] were used to construct a look-up table to determine the nature of restored images. This method has less calculation time and it enhanced the brightness of the image, but it does not preserve the corners in the restored image. The guided joint bilateral filter [18] was used to recover neighborhood smoothening as well as the preservation of edges, but this method also suffered from halo artifacts. The weighted guided image filtering method [19] used a familiar weight methodology to improve the filtering process and also it overcomes the problem of light line around the edges of the image. This method has a minimum calculation time when compared to other methods and also it increased the illumination of the restored image. An image fusion technique [20] was used in which it fused the output of guided image filtering, by a linear transformation method then an improved high boost filtering method is used to obtain the second input image which is used for the fusion process. This method improves the visual clarity of the images and it has increased running time efficiency. The variational method can overcome the drawbacks of physical assumption and over enhancement problem. Fang et al. [21] have designed a new method to remove haze from the images and also they removed the noise from an image, then base layer was also improved by using various techniques based on dark channel prior which removes segment artifacts problems. Galdran et al. [22] developed an energy utilization model which depends on hazy image in gray model, calculating an average value for restored image, but unwanted artifacts were found in closed regions. In the meta-heuristic method, they used genetic algorithm [23] for the haze removal to select the best parameters for producing the optimistic results. The genetic algorithm attains the best haze removal parameters but it does not guarantee the flawless solution.

4 Analysis of Various Haze Removal Methods

The analysis of various Haze removal methods is shown in Table 3.

Table 3 Analysis of various haze removal methods

S. no	Haze removal methods	Preservation of edges	Halo artifacts	Running time	Advantages	Applications
1	Variational dehazing	Yes	No	Moderate	Reduced over enhance-ment problem	Natural images
2	Dark channel prior	No	No	Poor	High-quality haze-free image	Outdoor images
3	Bilateral filtering method	Yes	No	Poor	Efficient noise sup-pression, no color drift	Underwater image enhance-ment
4	Fast image enhance-ment	No	No	Moderate	Zero color distortion	Intelligent transporta-tion system
5	Improved wavelet transform	Yes	No	Good	Efficient noise suppression	Natural images
6	Color lines based haze removal	No	No	Medium	Used for thin haze	Natural images
7	WGIF method	Yes	Yes	Excellent	Efficient noise suppression	Underwater images
8	Edge preserving	Yes	Yes	Moderate	Edges are preserved perfectly	Outdoor images
9	Improved dark channel prior	Yes	Yes	Good	Efficient noise suppression	Natural images
10	Genetic algorithm	Yes	Yes	Average	Optimistic results	Natural images
11	GGIF	Yes	Yes	Good	Edges are preserved perfectly	Underwater images

5 Performance Parameters

Performance parameters or metrics are used mainly in haze removal algorithms to check or analyze or reporting the quality of dehazed image. The important quality measures of an image are given by (1) Mean square error (MSE) (2) Peak signal-to-noise Ratio (PSNR).

5.1 Mean Square Error (MSE)

Mean square error is calculated between hazy image and the input image. It is calculated by using equation

$$MSE = 1/MN \quad \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left((C(x, y)) - \left(C^*(x, y) \right)^2 \right) \tag{4}$$

Here M and N are the input image apex and broadness, *C(x, y)* and $C^*(x , y)$ are restored image and original image.

Higher values of MSE indicate that low quality of image is obtained.

5.2 Peak Signal-to-Noise Ratio (PSNR)

Peak signal-to-noise ratio is calculated with respect to input image and hazy image and it is given by

$$PSNR = 10 log\left(255^2/MSE\right) \tag{5}$$

The higher value of PSNR shows that the image quality is very good.

6 Applications of Haze Removal Methods

Removal of haze methodology plays a vital role in many image processing areas. Some of the wide important areas in which the haze removal methods used are aircrafts, remote sensing, intelligent transportation systems, underwater image processing, Object detection, outdoor surveillance, railway systems, aerial imagery, image processing and video retrieval, computational photography/vision applications, video analysis and recognition, image classification, military and defense surveillance system, etc.

7 Conclusions

This paper presents various simple and faster haze removal methods which are used in several important applications. Various haze removal methods have been analyzed which gives a clear idea of how haze reduces the visibility of the images captured in outdoor environment. Haze removal from a single image is a very tedious work due to the calculation of scene depth map and also it requires certain prior knowledge. Also, it is very important to conserve the edges, illumination of the images, color characteristics, and appearance of the dehazed images. In future, various filtering methods have to be developed which gives optimal evaluation of depth map and to increase PSNR value with increased image quality and good running time.

References

1. Chen BH, Huang SC, Cheng FC (2016) A high efficiency & high speed gin intervention refinement filter for haze removal. IEEE J Disp Technol 12(7):753–759
2. Shen Y, Wu X (2016) Single image haze removal using Gaussian mixture model & sparse optimisation. In: IEEE conference on visual communication & image processing (VCIP), pp 1–4
3. Narasimhan SG, Nayar SK (2003) Contrast restoration of weather degraded images. IEEE Trans Pattern Anal Mach Learn 25(6):713–724
4. Narasimhan SG, Nayar SK (2000) Chromatic framework for vision in bad weather. In: Proceedings of IEEE conference on computer vision pattern recognition (CVPR), Hilton Head Island, SC, USA, pp 598–605
5. Shwartz S, Namer E, Schechner YY (2006) Blind haze separation In: Proceedings of IEEE conference on computer vision pattern recognition (CVPR). pp 1984–1991
6. Kopf J et al. (2008) Deep photo: model-based photograph enhancement and viewing. ACM Trans Graph 27(5):1–10
7. He K, Sun J, Tang X (2011) Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 33(12):2341–2353
8. Tripathi AK, Mukhopadyay S (2012) Single image fog removal using anisotropic diffusion. IET Image Proc 6(7):410–425
9. Qingsong Z, Jiaming MJ, Ling S (2015) A fast single image haze removal algorithm using color attenuation prior. IEEE Trans Image Process 24(11):3522–3533
10. Li Z, Zheng J (2015) Edge-preserving decomposition-based single image haze removal. IEEE Trans Image Process 24(12):5432–5441
11. Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. Proceedings of IEEE conference on computer vision pattern recognition, pp 839–846
12. He K, Sun J, Tang X (2013) Guided image filtering. IEEE Trans Pattern Anal Mach Intell 35(6):1397–1409
13. Zhengguo L, Senior Member, IEEE, Jinghong Z, Member, IEEE (2018) Single image dehazing using globally guided image filtering. IEEE Trans Image Process 27(1):442–450
14. Yuan-Kai W, Ching-Tang F (2015) Multi scale fusion of depth estimations for haze removal. IEEE Trans DSP 23(11):3245–3252
15. Rong Z, Jun WL (2014) Improved wavelet transform algorithm for single image dehazing. Optik-Int J Light Electron Opt 125(13):3064–3066
16. Wang Z, Feng Y (2014) A fast single image haze enhancement. Comput Electr Eng 40(3):785–795

17. Kumari A, Philip Joseph T, Sahoo SK (2014) Single image fog removal using gamma correction and Median filtering. IEEE Trans Image Process 12(1):44–50
18. Xiao C, Gan J (2012) Fast image dehazing using guided joint bilateral filter. Vis Comput 28(6):713–721
19. Li Z, Zheng J, Zhu Z, Yao W, Wu S (2015) Weighted guided image filtering. IEEE Trans Image Process 24(1):120–129
20. Riaz I, Yu T, Rehman Y, Shin H (2016) Single image dehazing via reliability guided fusion. J Vis Commun Image Represent 40:85–97
21. Fang F, Li F, Yang X, Shen C, Zhang G (2010) Single image dehazing and denoising with variational method. In: IEEE international conference on image analysis and signal processing (IASP), pp 219–222
22. Galdran A, Vazquez-Corral J, Pardo D, Bertalm (2015) Enhanced variational image dehazing. SIAM J Imag Sci 8(3):1519–1546
23. Guo F, Peng H, Tang J (2016) Genetic algorithm-based parameter selection approach to single Image defogging. Inf Process Lett 116(10):595–602

Secured Image Transmission in Medical Imaging Applications—A Survey

Pavithra Prabhu and K. N. Manjunath

Abstract Radiology Information Systems (RIS) involves the transmission of medical images between the PACS server and the workstation over the network for analysis. The images during transmission are vulnerable due to the intruders, and it is essential to secure the images. Because processing the adulterated images is unethical and can mislead the radio diagnosis. The combination of image processing and the cryptographic techniques can resolve this problem to some extent. Over a period, the technology has improved a lot, and several images securing methods are discussed by various researchers. This review paper is an effort in reviewing the existing secured image transmission technologies, their pros, and cons and the scope of the future work.

Keywords DICOM · Medical image exchanging · Watermarking techniques Joint encryption

1 Introduction

E-Healthcare system plays a major role in the modern world. It monitors the health condition and also provides facility to the doctors to provide appropriate medical treatments. As the technology of computing system is advanced, it is then lead to rapid enhancement of a better patient record management system and medical instrumentation. Medical imaging and radio diagnosis play a vital role in understating the anatomy and physiological aspects of human body on 2D cross-sectional images of 3D volume. Exchange of medical images over a public or private network is always vulnerable. Ensuring the safety of exchanged health records is made compulsory by

P. Prabhu · K. N. Manjunath (✉)
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India
e-mail: kn.manjunath@ieee.org

P. Prabhu
e-mail: pavithraprabhu104@gmail.com

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_12

the government in the system of Health Insurance Portability and Accountability Act (HIPAA). Secured implementation of an application requires three main characteristics to be considered which includes **Confidentiality**: a concept of ensuring that the data is not revealed to unauthorized people, **Integrity**: maintaining the consistency of the data, **Availability**: preventing unauthorized users and accessibility of data to the authentic user (also called as CIA triads of security) which ensures the safe transmission of data over the network. Confidentiality is necessary to prevent illegitimate access to the transmitted image and to preserve authenticity and integrity of the received image. At present, cryptography and digital watermarking are the main methods adopted by the medical research community.

In this paper, we first discuss on the brief introduction on medical imaging, then the challenges involved in the transmission of these images and finally the current state of the art methods in secure transmission and our review.

1.1 Medical Imaging

The DICOM standard is a technical framework in medical imaging applications for exchanging images between hardware/software from different vendors [1]. This is a common standard used for storing and transmission of DICOM images. Today, the radiology images, such as CT, MRI, PET, and Ultrasound modalities support the DICOM standard. After image acquisition, at the patient level, the 3D volumetric scan results in "n" number of images which are generally in hundreds to thousands. And the value of "n" depends on the important scanning parameter called Slice Thickness (ST). For example, an abdomen scan results in 3000 CT images and a head and neck scan 400 CT images, etc. These images are stored in PACS (Picture and Communication) server. The security of all those "n" number of images is to be considered. There might be an intrusion taking place while the transmission is in progress and the information obtained by the attacker may be used for illegal activities.

1.1.1 Challenges in Medical Image Exchanging

Cloud-based medical records and DICOM image services are evolving rapidly so that the tasks of a hospital's IT departments decrease, which allows them to focus on EMR adoption and clinical support systems. Main challenges exist in handling the images are how securely the images can be transmitted between the workstation and the PACS server and over the internet. Because adversarial effects are unpredictable, and nothing can be guessed about what type of attack happens on images.

2 Literature Survey

Many current works in the field of DICOM image security mainly concentrate on enhancing the image level security.

Bouslimi et al. [2] have discussed a joint encryption and watermarking technique to verify the reliability of the transmitted DICOM images based on the integration of RC4 algorithm which is a stream cipher method. There are two types of encryption algorithms, stream chipper encryption and block cipher encryption [3]. Stream ciphers are used when the data to be encrypted is not so large. When the large data is involved, they are divided into blocks and then fed to the encryption algorithms. They have used two substitutes for watermarking techniques which includes LSB substitution and Quantization Index Modulation (QIM). Through LSB, it is the simplest form of watermarking technique where it replaces LSB of each pixel by one bit of the data to be embedded. One must consider the gray intensities parity of the pixel to extract the data. In QIM, the components of image are quantified in accordance with the set of codebooks based quantizer to embed a message. Their proposed method showed that the decrypted image was less distorted ultrasound image. Alike to this method, the proposed method in [4] uses AES along with RC4 for encryption purpose. The output image was transparent to the DICOM standards.

Digital watermarking is used to identify the ownership of the delivered content. As shown in Fig. 1, a simple digital watermarking algorithm takes host image and watermark message along with the key as input to produce an embedded image as output.

Most of the standard encryption algorithm like AES and 3DES are computationally expensive to encrypt data of bigger size. A new lossless encryption method which uses the quaternation technique was implemented in [5]. This technique reduced the time taken for processing immensely. A counter mode of encryption and modular arithmetic operations was used to apply quaternion technique. In [6], the quaternion encryption and the algorithm for DICOM images were implemented using

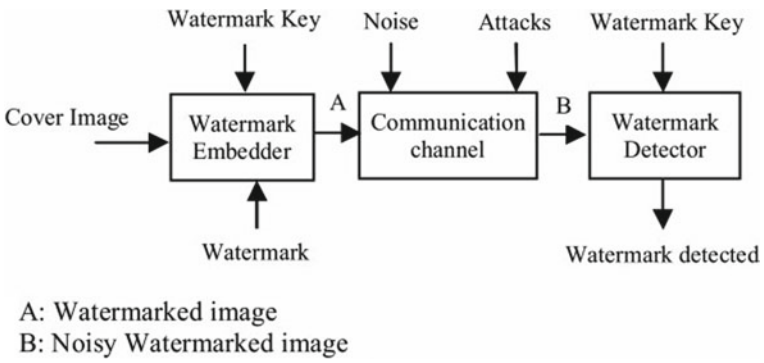


Fig. 1 A simple digital watermarking procedure

modified Feistel structure, but in [5], the algorithm was applied using counter mode of encryption. This satisfied the security necessities which included more randomness in the cipher with the use of additional input matrices such as initialization matrix and counter matrix. For digital signature, patient's signature and data were used as watermark images and then embedded in DICOM image using Discrete Wavelet Transform and Singular Value Decomposition. The computation speed of the encryption makes it easier to apply to a large number of images because of this quaternion-based encryption.

Various approaches of watermarking have been implemented to offer the security services which are necessary for radiology applications. The methods are mainly classified into reversible methods and irreversible methods. The reversible methods can reconstruct the original image without much distortion after extracting the watermark. Abdel-Nabi et al. [7] have proposed a joint encryption and data hiding algorithm which is a combination of standard encryption algorithm and reversible watermarking. It simultaneously embeds two watermarks where one is in spatial domain and another in encrypted domain. The algorithm uses reversible data hiding technique with the help of histogram shifting method so that the output image is distortion free. The first watermark is embedded by considering most of the pixels and encrypted using standard encryption algorithm. The remaining part is left unencrypted and is used to embed the second watermark. The algorithm implemented here was said to be separable since the decryption and watermarking extraction are independent of each other. From this method, it can be assured that the authenticity and integrity can be verified since it is retaining the original encrypted image.

Both Discrete Cosine Transform (DCT) and DWT have their disadvantages when it is used for watermarking techniques. Meheto et al. [8] proposed a combined DCT-DWT approach where it compensated the drawbacks of each method. It was an adaptive technique which identified the coefficients and less HSV areas were chose to hide the data. The image is divided into Region of Interest (ROI) and Region of Noninterest (RONI). Mantos and Maglogiannis have implemented a method based on ROI reversible steganography scheme for images [9]. The private data of the patient which is extracted from the DICOM header and the hashing values calculated are embedded in the region of interest. The recovery data is inserted in RONI area which is used to recover the ROI. The results of this experiment proved that the visual quality of original image and the encrypted image is good in terms of PSNR.

A new blind high payload capacity watermarking algorithm in the special domain which preserves ROI was proposed by Kundu and Das [10]. The scheme includes a selection of polygonal ROI and applying the SHA-256 hash algorithm. Later steps include watermark generation and embedding. The main advantage of this proposed scheme was that it gave remarkably good results with high payload capacity.

Storage can be big overhead when it comes to processing a huge volume of data. To reduce this overhead, a reversible crypto watermarking technique was discussed in [11]. This methodology consisted of stream cipher encryption of the image using a private key, and the secret key ciphering is done by public–private key cryptosystem ELGamal. The ELGamal encrypted secret key is used to watermark DCT domain using the spread spectrum coding algorithm. The proposed watermarking algorithm

in the transform to DCT domain is spread spectrum technique. It uses the index sorting method to obtain the first *n* high-frequency coefficient so that the position to insert the watermark can be found. The watermark is applied over the region where the accumulating frequency of the bins is negligible and cannot be detected. The watermark was not placed in regions which are not significant.

To ensure the robustness against attacks, the secret key is embedded in the transform domain. If capacity is a concern, then pixel manipulations based on prediction errors could be an be effective media for transmitting. The empirical results had the analysis of the reversible crypto watermarking method, histogram analysis, and entropy requirements.

Messages modified using DCT coefficient does not maintain any dependencies. To overcome this problem, the author Liao et al. [12] proposed a method where the dependencies are preserved at the identical location in adjacent DCT blocks.

Another technique based on reversible watermarking was proposed by Sarani and Amudha [13] where the aim was to provide patient authentication and information secrecy. In the methodology, the image was divided into groups that every group will be containing four pixels which are represented in singular value. Before finding the singular value, the LSB pixel of the group is changed by using the flipping function.

Crypto-based algorithms were discussed by Kobayashi et al. [14]. These methods provided CIA assurance in radiology applications. The author has proposed a scheme which takes advantage of the hierarchy of data arrangement in DICOM standard. Pixel data confidentiality is provided through encryption. But the keys that are used to encrypt are stored in the header without encryption, and the security provided for the pixels may not be guaranteed. With the help of digital signatures which uses internally generated keys, authenticity and integrity were provided. But the proposed method failed to provide CIA of the header data.

Al-Haj et al. [15] in their work had proposed an algorithm which uses internally generated symmetric keys and hash values. The algorithm provided by them was aimed to overcome the limitations set by the DICOM standard and Kobayashi's scheme [14]. The limitation information is discussed in [15]. This algorithm uses Advanced Encryption Standard-Galois Counter Mode (AES-GCM), the elliptic curve digital signature algorithm, and the whirlpool hash function. AES-GCM takes three parameters as input which includes the data which is to be encrypted, 256-bit key value and a 256-bit initial vector and performs universal hashing over a binary Galois field [16]. Since AES-GCM is faster than other National Institute of Standards and Technology (NIST) algorithms, it can be easily used with large data. Whirlpool hash function is stronger than normal hash functions like SHA-256, SHA-1, SHA-224, MD-5, and SHA-384 [17]. The author selected whirlpool because there were no reported attacks on it. The third algorithm which they had included is Elliptic Curve Digital Signature Algorithm (ECDSA) that produces shorter signature than DSA which is then stored in the DICOM header. The first algorithm was AEC-GCM, ECDSA, and the whirlpool hash function, whereas the second algorithm consisted of AEC-GCM and ECDSA only. The PSNR values achieved by the proposed algorithm were low. This proved that the two images are not correlated, and thus the concept of confidentiality has been achieved.

As discussed above in DICOM images, the header data is subjected to confidentiality but not the pixel data. The authors Natsheh and Gale [18] came up with an encryption approach for multi-frame DICOM images. The method included an XOR-based AES encryption algorithm. The first encrypted image is used as a key to the rest of the images in the frame. This resulted in less encryption–decryption time along with ensuring confidentiality.

Another crypto watermarking technique for DICOM image transmission was implemented by Puech and Rodrigues [19] with a blend of public, private keys and secret key ciphering, and watermarking. It uses RSA algorithm which is a public key cryptosystem. In RSA, if the key size is chosen carefully, it will be a good method to share the secret key. The encryption procedure along with a secret key is applied to the image data. To encrypt the secret key, an encryption method based on public–private keys was used. Later, this secret key is embedded in the encrypted image. They generated the public and private keys with RSA algorithm. Data is encrypted in the form of streams of pixels. Since the method is tested on the JPEG images, it gives better results. But when huge data is involved RSA key generation might not be a good approach.

To overcome the issue regarding confidentiality, in [20], the author has provided a *stegacryption* method which is involved using steganographic methods for hiding data in an image and encrypting the data. The proposed method encrypts the private data and then embeds it into the DICOM image in an undetectable way. This algorithm maintains the private data involved in a DICOM image even when the image is converted into a standard image file format.

Laiphrakpam and Khumanthem [21] proposed an improvised version of ElGamal cryptosystem and elliptic curve algorithm. For the elliptic curve analog ElGamal encryption, the encoding of the plain text is carried out first before encryption. The author proposed a method where this separation calculation was removed. This paper reduced data expansion in DICOM images and enhanced the execution speed.

Through the steganography process, it is possible to hide patients secrete information in the DICOM image which is being transmitted. Ulutas et al. [22] proposed a work that was based on Shamir's secrete sharing algorithm where the image was shared among n clinicians and to reveal the message, k of them must be gathered. This method provided confidentiality, authenticity, and hiding of patient's information.

3 Comparison Analysis

The comparison of all these techniques is shown in Table 1 along with different objectives of information security tested. Most of the methods have checked the authenticity and confidentiality of the data, but only few have been tested on DICOM images. And all these methods were tested on single image instead of on entire patient dataset. Practically, in radiology, we need methods to transmit the entire dataset at once securely which guarantees the secured transmission in the presence of the adversaries over an insecure network.

Table 1 Analysis of various methods discussed over CIA

Authors	Methodology overview	Achieved confidentiality	Achieved authenticity	Tested for DICOM images
Dalel et al.	RC4 + LSB substitution + QIM	Yes	Yes	No
Saganya et al.	AES + RC4 + QIM	Yes	Yes	Yes
Fathima Nasreen and Chitra	Quaternation technique + Countermode of encryption + DW T-SVD	Yes	Yes	Yes
Dzwon kowski et al.	Quaternation + Modified Feistel structure	Yes	Yes	Yes
Hiba et al.	Reversible watermarking technique	Yes	Yes	No
Petros et al.	ROI Reversible Steganography scheme	Yes	No	Yes
Anna Babu et al.	Elgamal_Discrete Cosine transform function	Yes	Yes	No
Kobayashi et al.	Hashing techniques	No	Yes	Yes
Ali Al-Haj et al.	AES-Galois + Whirlpool + ECDSA	Yes	Yes	Yes
Puech et al.	Public Key and Secret key ciphering (RSA) + Watermarking	Yes	Yes	No
Blackleedge et al.	Stegacryption method	Yes	No	Yes

None of these papers have discussed different adversarial techniques like image deletion from the dataset, image addition to a dataset and modification of the image pixel intensities.

3.1 Application Context

The author in [2] implemented the methodology on Echo graphic images. Though this methodology induced information loss, it was able to achieve good diagnosis. In [4] the methodology guaranteed priori and posteriori protection of images which was performed on ultrasound images and PET images. The author in [5] used medical images which are in the DICOM format were considered along with the signature

of the patient and the data. In [6] a modified DCOM network to identify the location where encryption can be performed to achieve high security and better computation speed is being implemented.

To provide access privilege to the end user, the images like MRI, CT, Ultrasound, and X-Ray images were considered by the author in [7]. The main application was implemented by considering the steganography concepts. The proposed method was applied for DICOM images over the Internet and storing of those data over the cloud infrastructure. The work proposed in [14] was developed in a java application. This method was an approach in improving the DICOM standards to enhance the security in class of integrity and authenticity. In paper [15], the author focused on providing all the three CIA triads for the DICOM image. In [19] by considering safety transmission of images with the help of both public and private keys.

4 Conclusion and Future Scope

Since the evolving of medical image distribution in the healthcare system, the importance of confidentiality, integrity, and authenticity aspects has become a necessity. Over a period, many watermarking algorithms were developed and tested by several authors which had some drawbacks and some advantages over one another. This paper discussed the importance of secure medical image transmission system and problems associated with it. Finally, we reviewed several medical watermarking and cryptographic methods with explanations and comparisons drawn between them. As observed from the papers, single image encryption techniques are available in abundance. But when it comes to transmission of the entire patient dataset, there is a need to improve security of all the images transmitted and number of literature discussing this are too limited. This can be considered as the new area to explore the sustainable solutions against eavesdrop, attacks, and violation of data integrity over an insecure network.

References

1. The DICOM chapter 3 (2012). PS 3.3. [Online] Virginia: NEMA USA, pp 390–1174. http://dicom.nema.org/standard.html [Accessed 19 June. 2016]
2. Bouslimi D, Coatrieux G, Roux C (2011) Joint encryption/watermarking algorithm for verifying the reliability of medical images: application to echo graphic images. Comput Methods Prog Biomed Elsevier 45–47
3. Stallings W (2010) Cryptography and network security, Pearson, 5th ed pp 91–430
4. Suganya G, Amudha K (2014) Medical image integrity control using joint watermarking techniques. In: International conference on green computing communication and electrical engineering (ICGCCEE)
5. Fathima Nasreen K, Chitra P (2016) A robust encryption and digital watermarking scheme for DICOM images using quaternation and DWT-SVD. In: International conference on green engineering and technologies (IC-GET)

6. Dzwonkowski M, Papaj M, Rykaczewski R (2015) A new quaternion based encryption method for DICOM images. IEEE Trans Image Process 24:11
7. Abdel-Nabi H, Al-Haj A (2017) Efficient joint encryption and data hiding algorithm medical images security. In: 8th international conference on information and communication systems (ICICS)
8. Mehto A, Mehra N (2015) Adaptive lossless medical image watermarking algorithm based on DCT and DWT. In: International conference on Information and Privacy, Nagpur India
9. Mantos PL, Maglogiannis I (2016) Sensitive patient data hiding using a ROI Reversible steganography scheme for DICOM images. J Med Syst. https://doi.org/10.1007/s10916-016-0514-5
10. Kundu MK, Das S (2010) Lossless ROI medical image watermarking technique with enhanced security and high pay load embedding capacity. In: International conference on pattern recognisation
11. Babu A, Ayyappan S (2015) A reversible crypto-watermarking system for secure medical image transmission. In: IEEE INDICON
12. Liao X et al (2017) Medical JPEG image steganography based on preserving inter-block dependencies. Comput Electr Eng. http://dx.doi.org/10.1016/j.compeleceng.2017.08.020
13. Sarani N, Amudha K (2015) A security technique based on watermarking and encryption for medical images. In: IEEE sponsored conference ICIIECS
14. Kobayashi L, Furuie S, Barreto P (2009) Providing integrity and authenticity in DICOM images: a novel approach. IEEE Trans Inf Technol Biomed 13(4):582–589
15. Al-Haj A, Abandah G, Hussein N (2015) Crypto based algorithms for secured medical image transmission. Instit Eng Technol (IET) J 9(6):365–373
16. Dworki M (2007) Recommendation for block cipher modes of operation: Galois/countermode (GCM) and GMAC (NIST Special Publication, 800–38D)
17. Barreto P, Rijmen V (2003) The WHIRLPOOL hashing function. http://planeta.terra.com.br/informatica/paulobarreto/whirlpool.zip
18. Natsheh N, Li B, Gale AG (2016) Security of multi-frame DICOM images using X-OR encryption approach. In: International conference on medical imaging understanding and analysis
19. Puech W, Rodrigues M (2004) A new crypto-watermarking method for medical images safe transfer. In: 12th European signal processing conference
20. Blackledge J, Al-Rawi A, Tobin P (2014) Stegacryption of DICOM metadata. In: 25th IET irish signals & systems conference 2014 and 2014 China-Ireland international conference on information and communications technologies (ISSC 2014/CIICT 2014)
21. Laiphrakpam DS, Khumanthem MS (2017) Medical image encryption based on improved Elgamal encryption technique. Optik 147(2017):88–102
22. Ulutas M, Ulutas G, Nabiyev VV (2011) Medical image security and EPR using Shamir's secrete sharing scheme. J Syst Softw 84 (2011):341–353

Evolution of Methods for NGS Short Read Alignment and Analysis of the NGS Sequences for Medical Applications

J. A. M. Rexie and Kumudha Raimond

Abstract In medical and genomic research, the Next Generation Sequencing (NGS) has a major role. Presently, NGS data are produced at the rate of 10 TB a day and challenge the storage and data processing capacities. These huge datasets are being used by a wide sort of applications such as customized cancer healing and precision medicine. NGS technologies offer prospects for understanding unidentified species and complex syndrome. To utilize genomic data for such applications, the genomic data in the form of short reads produced by NGS initially has to be assembled into whole genome sequence. And then, the sequences have to be compared for similarity and variation discovery which will be useful for analyzing and arriving at health-related solutions. In this paper, the fundamental methods for short read alignment such as assembly-based and alignment-based methods are discussed. Followed by which, the different ways to compare the sequences to check the alignment for similarity/dissimilarity discovery are discussed. This comparative analysis report can be utilized for health-related medical decisions.

Keywords NGS · Short read alignment · Multiple sequence alignment
Pairwise sequence alignment

1 Introduction

The human genome is encoded as Deoxyribonucleic Acid (DNA) particles which are in the form of double-stranded structure. It looks like twisted double threads consisting of two correlative chains of nuleotides [1]. Both the chains are a progression of nucleotides, spoke to as the alphabets A, T, G, and C, where A is complement to T and G to C. The DNA carries the characteristics from parents to their children. The sample DNA is collected from hair, spit, blood or skin of the species for

J. A. M. Rexie (✉) · K. Raimond
Department of Computer Sciences Technology, Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: rexiejoseph@gmail.com

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_13

which DNA is to be sequenced. The genetic material is extracted by the biological lab experiments. As a result of such experiments, a string of around 3.2 billion base pairs (the combinations of A, T, G, and C) will be generated. Approximately, 99.5% of the reference sequence of human will be similar to all human beings. The balance 0.5% only differs the characteristics of individuals. It may seem to be very less with respect to the reference genome. But still, even this 0.5% represents millions of nucleotides, which in turn gives the clue that there may be these many amount of different characteristics possible.

The main technique to arrange DNA was created by Sanger in 1975, and another strategy was proposed by Maxam and Gilbert in 1977, which were named the "chain-end strategy" and the "concoction sequencing technique", individually. Being not so much mind-boggling but rather more versatile in contrast with the compound sequencing technique, the chain-end strategy, in the end, was utilized broadly for original sequencing. DNA libraries with various embed sizes can be effectively created, which ends up being valuable as entire genome gathering profits by the reads with blended embed sizes. In any case, the low throughput and high cost of the original sequencing prompted a basic move in the procedure, taking us to Second Generation Sequencing (SGS) [2] called as Next Generation Sequencing (NGS).

The NGS strategies, for example, Illumina/Solexa sequencing methods, commonly generates 50–200 million 32–100 bp reads/run [3]. Aligning these short reads to a genome sequence as huge as human raises a grand opportunity to develop optimized (with respect to time) sequence assembly programs. Since the sequence generated will be utilized for health-related solutions, it is mandatory to generate the accurate sequence. By considering the accuracy as a challenge, various strategies have been devised to assemble the short reads delivered by NGS technique.

Fundamental techniques for assembling the NGS data can be categorized as two principle ideas such as alignment-based and assembly-based sequencing [4]. The alignment-based technique needs a reference sequence as a guideline, and figures out the location to fix each short read in order to match with the reference genome. The limitation of this approach is that it is possible only for the species for which reference sequence is available. However, the issue is solved through Assembly-based sequencing. This approach reconstructs the original DNA sequence by aligning and merging fragments from the DNA short reads.

The main issue with the sequence assembly is that some of the reads may be repeated more than once due to duplication during short reads generation. This fundamental problem has to be addressed in reading alignment/assembly approach. Most genomes contain a considerable volume of recurring fragments. Such duplicate fragments may get aligned to more than one region in the reference genome [4], which in turn lead to blunder in the aligned resulting sequence. The short read which is exactly getting mapped with any other short reads is identified as the duplicate and can be skipped [5]. Once the data is aligned and corrected, it is ready for further analysis and variation discovery.

The scope of research in NGS is the enhancement of alignment algorithms, error-correction techniques, and analysis of the sequence for further applications such as

genetic disease risk, genetic fingerprinting, genetic ancestry test, genetic compatibility test, genetic paternity test, and personalized medicine [6].

In this paper, Sect. 2 covers the genome assembly algorithms such as alignment-based and assembly-based approaches. Then the pairwise and multiple sequence alignment algorithms for comparing and analyzing the sequences are discussed in Sect. 3 followed by conclusion in Sect. 4.

2 Next Generation Sequencing Alignment Algorithms

2.1 Alignment-Based Approach

Hash-Based Aligner

In most of the alignment-based approaches, hash-based alignment strategy is commonly used for aligning the sequences in line with the reference sequence. The general idea of the method is described here as in [3].

For understanding the method, consider the following notations used to define the algorithm: Let,

Σ	Alphabet set
'#'	End marker. (Also, '#' is not part of alphabet set and is considered as lexicographically the least)
$S = s_0 s_1 s_2 \ldots s_{l-1}$	String of length l having always # as the last symbol (i.e., s_{l-1} is '#')
$S[m] = s_m$	mth symbol of S for $m = 0, 1 \ldots l - 1$
$S[m, n] = s_m \ldots s_n$	Substring
$S_m = S[m, l - 1]$	Suffix of S
SA(m)	Suffix array (SA) of S (the start position of the mth smallest suffix)
|S|	Length of the string S

The Burrows–Wheeler Transform (BWT) of S is defined as follows:

$$BT(m) = \begin{cases} \#, & SA(m) = 0 \\ S[SA(m) - 1], & Otherwise \end{cases} \quad (1)$$

In applications, the suffix array will be constructed first and then BWT will be generated instead of using the Eq. 1. The example given in Fig. 1 better demonstrates the calculation of BWT.

Let X be a substring of S. Since all suffix strings having X as prefix are arranged next to each other, the position where X occurs in S will be in an interval in the SA. This leads to defining the following:

0	science$
1	cience$s
2	ience$sc
3	ence$sci
4	nce$scie
5	ce$scien
6	e$scienc
7	$science

Index

S = science$

0	7	$scienc	e
1	5	ce$scie	n
2	1	cience$	s
3	6	e$scien	c
4	3	ence$sc	i
5	2	ience$s	c
6	4	nce$sci	e
7	0	science	$

m *SA(m)* *BT(m)*

SA = (7, 5, 1, 6, 3, 2, 4, 0)

BT = enscice$

Fig. 1 Computing SA and BT for S = science$

$$\overline{I}(X) = \text{minimum}\{i : X \text{ is the prefix of } S_{SA(i)}\} \tag{2}$$

$$\underline{I}(X) = \text{maximum}\{i : X \text{ is the prefix of } S_{SA(i)}\} \tag{3}$$

In specific, $\underline{I}(X) = 0$ and $\overline{I}(X) = 1 - 1$, if X is an empty string. The SA interval of X is defined as the interval $\left[\underline{I}(X), \overline{I}(X)\right]$. The set of positions of all occurrences of X in S is $\left\{SA(i):\underline{I}(X) \leq i \leq \overline{I}(X)\right\}$. Knowing the intervals in suffix array we can get the positions. Hence, the problem of sequence alignment is transformed to search for the SA intervals for the substrings of S that match the query.

Seed-and-Extend Based Aligner

Burrows–Wheeler Alignment—Maximal Exact Match (BWA—MEM) is the popular technique for which seed-and-extend strategy, as depicted in Fig. 2, is the base. For each of the short reads provided by the NGS technique, seed areas are identified, which is exactly aligned with the substring of the reference sequence [7]. It is followed by extension of the seed on both the sides of the seed by allowing insertions/deletions, i.e., inexact matches.

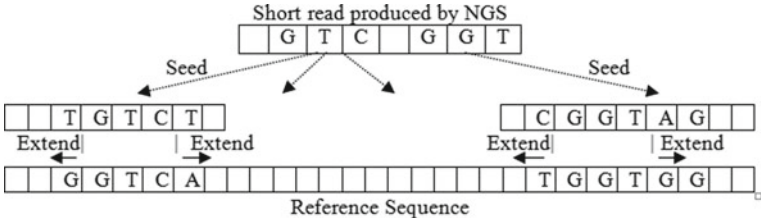


Fig. 2 Seed-and-extend strategy

2.2 Assembly-Based Approach

Overlap-Layout Assembler

Most assembly-based approach algorithms apply overlap-layout method to reproduce the genome from a huge amount of short reads. Greedy technique is used to merge the reads having overlap to form a longer sequence called contig. An issue may arise because of the ambiguities caused by the recurring fragments that are present in genome reads. To overcome such issue, longer reads can be used to connect the recurring fragments, consequently enabling us to precisely line the contigs together. By considering the reads as vertices and the count of overlapping reads as edges, a graph representing the relation among the short reads is shown in Fig. 3a. Further, by applying Hamiltonian algorithm (as in Fig. 3b) which will visit every vertex only once, the sequence can be assembled [8].

K-mer Graph-Based Assembler

In this method, initially, a de-Bruijn-graph is constructed. K-mers are represented as vertices in de-Bruijn-graph which is a directed graph. The K-mers that partly cover the substring of other K-mers are connected through edges. For the construction of such a graph, first each short read is read through linearly by a sliding window of size K, and then the read subsequences of size K are represented as nodes. Followed by which, Euler path algorithm is applied (as shown in Fig. 4) to rebuild the whole sequence by visiting every edge not more than once [9].

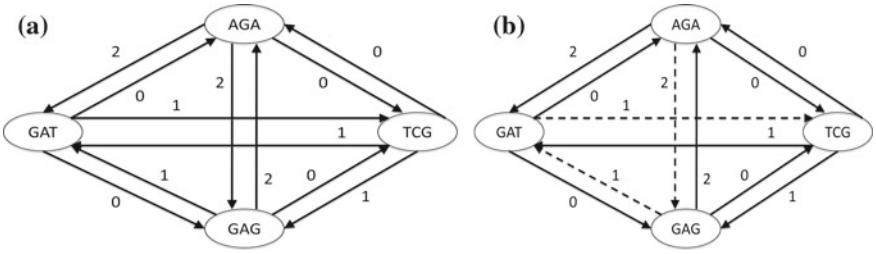


Fig. 3 **a** Let S = {GAT, TCG, GAG, AGA}. Edge cost represents size of overlapping subsequence. **b** Hamiltonian path tracing every vertex just once PATH = AGAGATCG

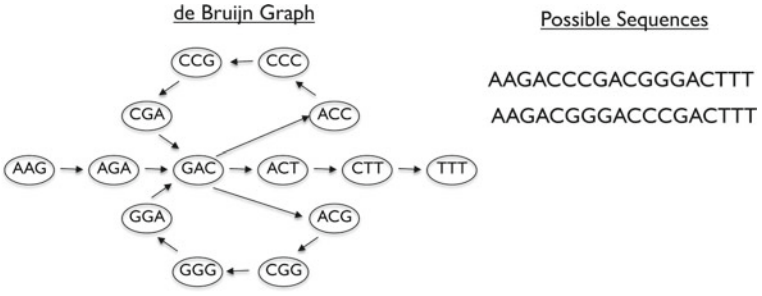


Fig. 4 K-mer graph (k = 3)

3 Sequence Alignment for Analysis

There are two ways to do the analysis through comparison of DNA sequences. One is by alignment of the sequences, and the other is without alignment [10].

3.1 Sequence Alignment Algorithms

Based on the research needs and resources in hand, a variety of sequence alignment problems are existing [11]. Two types of sequence alignment solutions for analysis, i.e., variation discovery are used in general, known as pairwise sequence alignment and multiple sequence alignment.

A. **Pairwise Sequence Alignment**

Certainly, pairwise sequence alignment focuses on evaluating a pair of sequences to figure out the degree of similarity between them or identifying the equivalent portion of sequences present in both the sequences. It can be any of two types, such as global alignment and local alignment. The following example well explains the basic principles of such alignments.

Global alignment	G	A	–	–	–	–	T	A	–	A	A	T	G	A	–	–	–	–	T
Reference sequence	G	A	C	C	G	A	T	A	C	A	A	T	G	A	T	A	C	G	T
Local alignment	–	–	–	–	G	A	T	A	–	A	A	T	G	A	T	–	–	–	–

Primarily based on these methods, pairwise alignment tries to evaluate whether or not two sequences are sharing common fragments of sequence, using the condition stated as parameters. The most used basic method for such alignment is dynamic programming technique. A table of m × n is constructed to compute the optimal score of all possible sequential combinations of the two sequences being compared. By tracing back the table, similar parts of the sequences can be found [12].

B. **Multiple Sequence Alignment**

A multiple sequence alignment commonly avoids the query whether or not the sequences are similar, rather specializes in the quasi-surest answer of aligning sequences that have been shown to be related by way of other tactics, in fact often via pairwise alignment [13]. It is obvious from its name that it takes into consideration more than one sequence and offers insights into the structural, and from time to time practical, relationships of genomes, genes and proteins.

3.2 Alignment-Free Sequence Analysis

The weak point of alignment-based approach is that the correctness of the alignment falls down quickly when the sequence uniqueness goes down below a critical value [14]. Alignment-based approach tries to do the alignment accurately with respect to the reference sequence. This will be yielding good results for the sequences which are having upper levels of homology. But, while analyzing the sequences of cancer genome with the normal one, it limits the correctness of the prediction [15].

In general, the occurrences of strings of K characters in the given sequences will be counted in alignment-free analysis. The string of length K is named as K-mers. In these methods, K used to be constant. But, in some cases, the number of characters in the string considered will be varying [16]. The method in [17] uses the strategy of "shortest unique substrings", i.e., it checks for the occurrence of shortest substrings which should not be in sequence more than once. Hence, the unique regions can be identified directly using the short reads rather the aligned complete sequence.

4 Conclusion

The improvements in NGS technologies for generating DNA sequence of any species in less time compared to the first generation sequencing methods and the huge amount of data generated every day has directed to the discovery of many algorithms. Even then, still there is a need for more analysis in the field of medical for health-related personalized-solutions. This paper has given the fundamental concepts on sequence alignment and analysis of the sequences. This can lead to further research to discover enhanced computational solutions which can improve the methods available for sequence alignment and analysis of the sequences.

References

1. Ayday E, De Cristofaro E, Hubaux J-P, Tsudik G (2015) Whole genome sequencing: revolutionary medicine or privacy nightmare? Comput Publ IEEE Comput Soc 48(2)
2. Hengyun L, Giordano F, Ning Z (2016) Oxford nanopore MinION sequencing and genome assembly. Genom Proteom Bioinf 14:265–279
3. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bio Inf 25(14):1754–1760
4. Chelsea J-T, Ju ZZ, Wang W (2017) Efficient approach to correct read alignment for pseudogene abundance estimates. IEEE/ACM Trans Comput Biol Bioinf 14(3)
5. Xu H, Luo X, Qian J, Pang X, Song J, Qian G et al (2012) FastUniq: a fast *De Novo* duplicates removal tool for paired short reads. PLoS ONE 7(12)
6. Roy A, Diao Y, Mauceli E, Shen Y, Wu BL (2012) Massive genomic data processing and deep analysis. Proc VLDB Endow 5(10)

7. Houtgast EJ, Sima V-M, Bertels K, Al-Ars Z (2016) GPU-accelerated BWA-MEM genomic mapping algorithm using adaptive load balancing. In: Hannig F et al (ed), ARCS 2016, LNCS 9637, pp 130–142
8. Chena C-C, Ghaffarib N, Qiana X, Yoona B-J (2017) Article optimal hybrid sequencing and assembly: feasibility conditions for accurate genome reconstruction and cost minimization strategy. Comput Biol Chem 69:153–163
9. Bresler G, Bresler M, Tse D (2013) Optimal assembly for high throughput shotgun sequencing. Bioinformatics 14(Suppl 5):S18
10. Haubold B, Reed FA, Pfaffelhuber P (2011) Alignment-free estimation of nucleotide diversity. Bioinformatics 27(4):449–455
11. Baichooa S, Ouzounisb CA (2017) Computational complexity of algorithms for sequence comparison, short-read assembly and genome alignment. BioSystems 72–85, 156–157
12. Haque W, Aravind A, Reddy B Pairwise sequence alignment algorithms—a survey. In: ISTA '09 Proceedings of the 2009 conference on Information Science, Technology and Applications, pp 96–103
13. Kieran Boyce A, Sievers F, Higgins DG (2014) Simple chained guide trees give high-quality protein multiple sequence alignments. Proc Nat Acad Sci United States Amer 111(29):10556–10561
14. Zielezinski A, Vinga S, Almeida J, Karlowski WM (2017) Alignment-free sequence comparison: benefits, applications, and tools. Genome Biol 18:186
15. Li Y, Heavican TB, Vellichirammal NN, Iqbal J, Guda C (2017) ChimeRScope: a novel alignment-free algorithm for fusion transcript prediction using paired-end RNA-Seq data. Nucleic Acids Res 45(13)
16. Leimeister C-A, Morgenstern B (2014) kmacs: the k-mismatch average common substring approach to alignment-free sequence comparison. Bioinformatics 30(14):2000–2008
17. Haubold B, Pierstorff N, Möller F, Wiehe T (2005) Genome comparison without alignment using shortest unique substrings. BMC Bioinf 6:123

Caries Detection in Non-standardized Periapical Dental X-Rays

D. Osterloh and Serestina Viriri

Abstract Dental caries are currently one of the most prevalent diseases in the modern world. Early detection and diagnosis of the disease is the best treatment available to dental healthcare professionals and is crucial in preventing advanced stages of decay. This paper presents an effective model for caries detection across a variety of non-uniform X-rays using individual tooth segmentation, boundary detection and caries detection through image analysis techniques. The tooth segmentation is implemented using integral projection and an analytical division algorithm. The boundary detection is implemented through the use of top and bottom hat transformations and active contours. Finally, the caries detection was achieved through the use of blob detection and cluster analysis on suspected carious regions. The cluster analysis generates its results relative to the image being analyzed and as such, forms the unsupervised evaluation approach of this paper. The viability of this unsupervised learning model, and its relative effectiveness of accurately diagnosing dental caries when compared to current systems is indicated by the results detailed in this paper, with the proposed model achieving a 96% correct diagnostic.

Keywords Dental X-rays · Tooth segmentation · Boundary detection · Caries identification

1 Introduction

Despite advances in oral healthcare, dental caries remain the most widespread of oral diseases, with approximately 36% of the world's population showing signs of the infection [1]. This has led to many attempts to improve the detection rate of caries

D. Osterloh · S. Viriri (✉)
School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Durban, South Africa
e-mail: viriris@ukzn.ac.za

D. Osterloh
e-mail: 209501289@ukzn.ac.za

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_14

in order to prevent more serious oral diseases from developing. Traditionally, dental X-rays have been used by oral healthcare professionals to assess unobservable areas of the tooth and make a diagnosis through observation [2]. Newer advancements in computer vision have led to the development of computer-aided diagnosis systems in order to assist in the identification and diagnosis process. Unfortunately these systems have a high false positive rate at identifying caries and as such have not been usable as standalone systems [3].

The goal of this paper is to propose a caries detection model to assist in the treatment of dental caries. The proposed model aims to rectify the shortcomings of existing models and provide more accurate results by implementing a new approach to the diagnostic algorithm.

Several factors have led to the unfavourable identification results with respect to caries diagnosis. First, dental X-rays are noisy and low in contrast due to the low dosage rates in the capture process [4]. These low dosage rates can also affect the visibility of caries due to the X-rays not fully penetrating the teeth. There is no workaround for this as the low dosage rates ensure the health of the patient [5], thus image enhancement techniques must be utilized to assist in computer vision. Second, the majority of segmentation research focuses on tooth segmentation for the purposes of human identification. As a result, features required for caries detection are lost in favour of preserving crown shape in order to match teeth. Finally, current caries detection algorithms use a supervised learning model as the basis of their comparative model. Suspected caries regions are compared against a set of classifiers which are obtained from a learning set where the presence of caries is known. If there are similarities between the test image and the classifiers, the algorithm provides a positive caries diagnosis.

There have been varying degrees of research into optimizing the each of the specific aspects of radiograph processing. Ahmad et al. [6] tested the effects of four image enhancement techniques, namely, adaptive histogram equalization (AHE), contrast adaptive histogram equalization (CLAHE), median adaptive histogram equalization (MAHE) and sharp contrast adaptive histogram equalization (SCLAHE) in an attempt to determine which provided better results in terms of improving X-ray quality. Further research by Bharathi et al. [7] looked at the effectiveness of median, finite impulse response (FIR) and Gaussian filters in reducing noise levels.

Research with regards to tooth segmentation alternates between the use of integral projection or active contours. Nomir and Abdel-Mottaleb [8] proposed a method adapted from the works of Hu et al. [9]. A mask of the initial image was obtained by performing an iterative and adaptive threshold. Integral projection was performed on this mask based on the assumption that most, if not all, of the non-teeth related pixels have been removed. This method was used again by Nomir et al. [10] for human identification. Lin et al. [11] also used an adaptation of the method presented in Ref. [8] for use in human identification. Jain and Chen [12] and Frejlichowski and Wanat [13] further developed this method to incorporate a probability model.

Segmentation through active contours was used by Zhou and Abdel-Mottaleb [14] as well as Oliveira [15]. Rad et al. [16] compiled an evaluation of these various segmentation methods.

Not much research has been done with respect to caries identification itself. Solanki et al. [17] used an unsupervised learning approach where the shape contour of each tooth was analyzed. Oprea et al. [18] proposed that a binary threshold be applied on a high contrast image. A subsequent rule check was performed to determine if any black pixel groups occurred within the tooth or along its boundary and flagged these as caries. Oliveira [15] made use of a supervised learning approach and developed a set of classifiers for caries detection. Finally, Zhang et al. [19] used a blob detection method to isolate potential dental caries for 3D rendering and assessment.

In order to achieve the goal of an improved caries detection model using an unsupervised learning approach, this paper presents a unique diagnostic model. This model is comprised of both adapted algorithms from previous research and novel algorithms which reduce the inaccuracies inherent to current methods.

2 Segmentation

The segmentation of the X-rays into individual teeth was achieved through a three-stage process. This process consists of pre-processing and image enhancement, adaptive and iterative thresholding and separation line selection using a novel algorithm.

2.1 Tooth Separation

Tooth separation was handled in two parts. The potential separation lines are initially generated through integral projection, which determines the gap regions between identified teeth. Following this, an evaluation algorithm determines the line of best fit.

Integral Projection Integral projection is able to analyze pixel intensities across an image and detect regions of darker pixels. As such, it provided the best solution for the detection of gaps between teeth, where the spaces between two adjacent teeth are easily identifiable from the thresholded mask obtained in the previous stage. Areas where clusters of black pixels were present between pairs of adjacent clusters of white pixels were identified as valleys.

Line Selection Separation lines were calculated using the gap clusters as training points for a linear regression model. Two variations of the simple linear regression algorithm were used. The first algorithm was the standard formula defined as follows:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{1}$$

$$= \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_j}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right)^2}, \tag{2}$$

$$= \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2}, \tag{3}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}. \tag{4}$$

where n denotes the number of points, β denotes the gradient of the slope and α denotes the y-intercept.

The second formula was a weighted linear regression model which proved effective in generating a correct separating line in cases where cluster distribution was favoured in one direction. In cases where there was an equal distribution of points around the median then the simple linear regression model was used. If the distribution of points was greater or less than the median then the value of n in the above equation was calculated to be half the total number of points.

In order to determine the best separation lines, the algorithm proposed by Frejlichowski and Wanat [13] was adapted to work on periapical X-Rays. The original algorithm determined separation lines based on the nature of panoramic X-rays, where all teeth are in view. It uses the uniform nature of teeth sizing to determine spacing across the entire row of teeth. Due to the nature of the X-rays being analyzed, where the number and types of teeth present in each X-Ray varied across each image, the adapted algorithm was required in order to achieve correct results. By combining the rotation algorithms used in [8] and [12] with an altered probability model derived from [13], the new segmentation algorithm was developed which incorporated both rotational and probabilistic functions.

For segmentation lines where the number of intersection points is equal to, or greater than, a previously determined optimal line, a new set of acceptance criteria were introduced. Based on the probability formula implemented in [12], vertical lines have a higher probability of generating successful segmentation results. The weighting system judged potential line candidates by using slope gradient and intersection point percentage relative to the total separation line. The probability of a line being the best segmentation line was determined by

$$P = |nW_T + \frac{W_I}{IP_{deep}} - J| \tag{5}$$

where P was the probability of the line being correct, n was the number of already segmented teeth, WT was the assumed width of the previous teeth, WI was the width of the image, IP is the number of integral projection points representing gaps between the teeth and J was the projected point of the segmentation line. The desire of the algorithm was to minimize the value of P where P actually represents the probability of a line being incorrect. Lines which fall on the projected segmentation

value have a P rating of 0 meaning there is close to 0 probability of it being incorrect. As segmentation lines move away from the projected point the value of P increases resulting in unfavourable selection chances.

To accommodate for acceptance of separation lines for impacted or extremely close adjacent teeth, a second algorithm was used. If more than 60% of the separation line intersected with teeth pixels then the line was discarded and the gap was regarded as a space between molar roots.

3 Caries Detection

Caries detection was handled in two stages. Potential regions of interest were first identified using an edge detector which highlighted all locations where dark spots, and by extension possible caries, were present. Once a region of interest had been defined a novel algorithm was applied to the area in question in order to assess the validity of the caries flag.

3.1 Blob Detection

After testing several methods, a blob detector was implemented for the detection of potential caries in the demarcated search space. Regions of possible decay appeared substantially darker when compared to the surrounding tooth matter, due to the high contrast of the image from the top and bottom hat transformation performed during the boundary detection phase. Blob detection algorithms were able to capitilize on this, owing to their ability to locate local maxima. The blob detection model proposed by Lindeberg [20] was implemented as it was not affected by scaling issues which arose from the varying sizes of the teeth being processed. The model used a Laplacian of the Gaussian approach to detect darker regions, which was defined as a convolution kernel of the form

$$LoG = \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} e^{-\frac{x^2=y^2}{2\sigma^2}} \tag{6}$$

where σ was the width of the kernel. This was approximated to a 5×5 kernel for the purposes of implementation defined as

$$LoG = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & -16 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

The use of a four-connected kernel resulted in some loss of definition around the edges of the caries clusters which negatively impacted the diagnostic method, therefore the eight-connected kernel was implemented.

3.2 Caries Analysis

Region of Interest Generation To achieve the goal of diagnosing whether dental caries were present with a non-supervised assessment model, image analysis techniques were implemented in order to assess the regions of interest using standard dentistry techniques. The depth of the search region was already known relative to the tooth, falling between 10 and 15% of the overall width. Teeth were approximated to fall between 7.5 and 9.0 mm in width as defined by Chu [21]. Due to the images being periapical X-rays and not panoramic, the exact tooth being analyzed was unknown as the X-Rays were taken of varying locations. In order to approximate the width, the following formula was proposed:

$$W = Tmax - \frac{T_{variance}(P_{max} - P_{calculated})}{P_{variance}} \tag{7}$$

where W was the estimated width, T was the width of the tooth and P was the percentage depth of the search space, determined to be 10–15%. $T_{variance}$ was obtained by calculating the difference of the maximum and minimum tooth width values and was determined to be 1.5. The value for $P_{variance}$ was calculated to be 5 following the same process. This formula was derived using the probability that teeth which required smaller analysis regions represented the narrower spectrum of teeth, whereas teeth with wider search regions represented the wider spectrum.

4 Experimental Results

4.1 Segmentation Results

The success rate of the segmentation method was evaluated based on its ability to correctly separate teeth in the upper and lower jaw regions individually as well as the combined results of both regions. This provided both specific results as to whether the algorithm performed better on a particular jaw region, as well as a holistic view as to how well it performed on average when looking at both jaw regions. The dental X-ray (periapical) dataset was provided by the Dental Clinic, Health Center at the Universiti Teknologi Malaysia (UTM).

Teeth were considered correctly separated if the separation line did not cause partial separation or division of the teeth. Teeth which were already partial as a result

Table 1 Region specific segmentation results comparison

	Upper jaw (%)	Lower jaw (%)
Oliveira [15]	72	72
Nomir and Abdel-Mottaleb [8]	84	81
Proposed approach	85	90

Table 2 Overall segmentation results comparison

	Accuracy (%)	Implementation
Nomir and Abdel-Mottaleb [8]	82.5	Thresholding
Said et al. [22]	83	Thresholding
Shah et al. [23]	58.1	Active contour
Phong-Dinh et al. [16]	77.23	Thresholding and integral projection
Oliveira [15]	71.91	Active contour without edge
Lai and Lin [24]	83	Region growing
Proposed approach	87.5	Thresholding and integral projection

of being at the edge of the X-ray were considered correctly separated if no further partiality was caused. Teeth which were not correctly segmented were either caused as a result of extremely poor contrast in the original image, where the enhancement techniques could not establish a distinction between teeth and non-teeth structures, or due to impacted teeth.

A comparison of the results on a jaw specific basis are presented in Table 1. The results obtained by Oliveira and Nomir and Abdel-Mottaleb are used as a comparison, due to the similarity of the implemented methods used to achieve dental segmentation.

As can be seen, with a combination of the adapted and novel algorithms discussed in this paper, the segmentation results improved over existing methods. Table 2 provides a comparison of the proposed method to other implementations of the segmentation process, as described in Ref. [16].

These results indicate that the method proposed in this paper offers a noticeable improvement on existing models. Furthermore, it indicates the diagnostic algorithm received the greatest quantity of correctly segmented images for evaluation.

4.2 Caries Detection Results

A collection of ground truth data was used to evaluate the success rate of the detection method. The data contained markers for the location of identified caries, as well as the locations of false positive regions. The false positive regions were defined as

Table 3 Caries identification results comparison

	Correctly categorized (%)	False positives (%)	Missed caries (%)
Tracy [25]	94	–	6
Dykstra [26]	60	20	20
Oliveira [15]	98	–	2
Valizadeh et al. [27]	90	–	10
Proposed approach	96	2	2

locations along the boundary of each tooth where caries were incorrectly identified. This occurred due to a misinterpretation of the region, either due to the contrast of the X-ray, or because a partial set of caries identifiers were present which led to the algorithm interpreting the results as a caries location.

To determine whether these rates fall within acceptable limits, a comparison was done against the different diagnostic methods available. These comprised of caries detection performed by dentists using the Logicon Caries Detector system, as discussed by Tracy et al. [25], unassisted caries diagnosis by dentists, as discussed by Dykstra [26], and caries detection performed by a supervised learning model, using the method proposed by Oliveira [15]. The results of this comparison can be seen in Table 3.

5 Conclusion

In this paper, an unsupervised learning model for caries detection was presented. The proposed model is implemented using a segmentation method to separate the X-rays into individual teeth, a boundary detection method to determine the edges of the teeth for caries analysis and finally a diagnostic algorithm that assesses the boundary using image analysis techniques. Both the proposed segmentation method and caries detection algorithm obtained favourable results when compared to similar models due to the novel approaches described in this paper. As such, the caries detection model outlined in this paper provides a viable alternative to existing models for use in caries detection.

References

1. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V et al (2013) Years lived with disability (ylds) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the global burden of disease study 2010. The Lancet 380(9859):2163–2196

2. Booshehry MZ, Fasihinia H, Khalesi M, Gholami L (2011) Dental caries diagnostic methods. DJH 2(1)
3. Amaechi BT (2009) Emerging technologies for diagnosis of dental caries: the road so far. J Appl Phys 105(10):102047
4. Noor NM, Khalid NEA, Ali MH, Numpang ADA (2010) Fish bone impaction using adaptive histogram equalization (ahe). In: (2010) Second international conference on computer research and development. IEEE, pp 163–167
5. Sakata M, Ogawa K (2009) Noise reduction and contrast enhancement for small-dose x-ray images in wavelet domain. In: (2009) IEEE nuclear science symposium conference record (NSS/MIC). IEEE, pp 2924–2929
6. Ahmad SA, Taib MN, Khalid NEA, Taib H (2012) An analysis of image enhancement techniques for dental x-ray image interpretation. Int J Mach Learn Comput 2(3):292–297
7. Bharathi KK, Muruganand S, Periasamy A (2014) Digital image processing based noise reduction analysis of digital dental xray image using matlab. J NanoSci NanoTechnol 2(1):198–203
8. Nomir O, Abdel-Mottaleb M (2005) A system for human identification from x-ray dental radiographs. Pattern Recognit 38(8):1295–1305
9. Hu S, Hoffman EA, Reinhardt JM (2001) Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images. IEEE Trans Med Imaging 20(6):490–498
10. Nomir O, Abdel-Mottaleb M (2007) Human identification from dental x-ray images based on the shape and appearance of the teeth. IEEE Trans Inf Forens Sec 2(2):188–197
11. Lin P-L, Lai Y-H, Huang P-W (2012) Dental biometrics: human identification based on teeth and dental works in bitewing radiographs. Pattern Recognit 45(3):934–946
12. Jain AK, Chen H (2004) Matching of dental x-ray images for human identification. Pattern Recognit 37(7):1519–1532
13. Frejlichowski D, Wanat R (2011) Automatic segmentation of digital orthopantomograms for forensic human identification. In: International conference on image analysis and processing. Springer, pp 294–302
14. Zhou J, Abdel-Mottaleb M (2005) A content-based system for human identification based on bitewing dental x-ray images. Pattern Recognit 38(11):2132–2142
15. Oliveira J (2009) Caries detection in panoramic dental X-ray images
16. Rad AE, Mohd MS Rahim, Rehman A, Altameem A, Saba T (2013) Evaluation of current dental radiographs segmentation approaches in computer-aided applications. IETE Tech Rev 30(3):210–222
17. Solanki A, Jain K, Desai N (2013) Isef based identification of rct/filling in dental caries of decayed tooth. Int J Image Process (IJIP) 7(2):149–162
18. Oprea S, Marinescu C, Lita I, Jurianu M, Visan DA, Cioc IB (2008) Image processing techniques used for dental x-ray image analysis. In: (2008) 31st international spring seminar on electronics technology. IEEE, pp 125–129
19. Zhang H, Boyles MJ, Ruan G, Li H, Shen H, Ando M (2013) Xsede-enabled high-throughput lesion activity assessment. In: Proceedings of the conference on extreme science and engineering discovery environment: gateway to discovery. ACM, p 10
20. Lindeberg T (1998) Feature detection with automatic scale selection. Int J Comput Vis 30(2):79–116
21. Chu SJ (2007) Range and mean distribution frequency of individual tooth width of the maxillary anterior dentition. Pract Proc Aesthet Dent 19(4):209
22. Said EH, Nassar DEM, Fahmy G, Ammar HH (2006) Teeth segmentation in digitized dental x-ray films using mathematical morphology. IEEE Trans Inf Forensics Secur 1(2):178–189
23. Shah S, Abaza A, Ross A, Ammar H (2006) Automatic tooth segmentation using active contour without edges. In: (2006) Biometrics symposium: special session on research at the biometric consortium conference. IEEE, pp 1–6
24. Lai Y, Lin P (2008) Effective segmentation for dental x-ray images using texture-based fuzzy inference system. In: International conference on advanced concepts for intelligent vision systems. Springer, pp 936–947

25. Tracy KD, Dykstra BA, Gakenheimer DC, Scheetz JP, Lacina S, Scarfe WC, Farman AG (2010) Utility and effectiveness of computer-aided diagnosis of dental caries. Gen Dent 59(2):136–144
26. Dykstra B (2008) Interproximal caries detection: how good are we? Dent Today 27(4):144–146
27. Valizadeh S, Goodini M, Ehsani S, Mohseni H, Azimi F, Bakhshandeh H (2015) Designing of a computer software for detection of approximal caries in posterior teeth. Iranian J Radiol 12(4)

Segmentation of Type II Diabetic Patient's Retinal Blood Vessel to Diagnose Diabetic Retinopathy

T. Jemima Jebaseeli, C. Anand Deva Durai and J. Dinesh Peter

Abstract Diabetic Retinopathy is one of the ophthalmic reasons for visual deficiency. The favored fixate of consideration is on the estimation of deviation in the breadth of the retinal veins and the new vessel development. To witness the progressions, segmentation has to be made primarily. A framework to improve the quality of the segmentation result over pathological retinal images is proposed. The proposed method uses adaptive histogram equalizer for preprocessing, pulse coupled neural Network model for automatic feature vector generation and extraction of the retinal blood vessels. The test result represents that the proposed method is enhanced than other retinal competitive methods. The evaluation of the proposed approach is executed over standard public DRIVE, STARE, REVIEW, HRF, and DRIONS fundus image datasets. The proposed technique improves the segmentation results in terms of sensitivity, specificity, and accuracy.

Keywords Medical imaging · Retinal blood vessel · Diabetic Retinopathy Fundus image · Feature extraction

1 Introduction

Diabetic Retinopathy (DR) is an ailment that prevails among Type II diabetic patients. If the disease is not predicted early, individuals may lose their vision. At the advanced stage, it prompts to separation of the retina from the eye. DR makes the external retinal layer thick. The lipids and proteins inside the retinal layers are produced a tiny blot of lesions, exudates, microaneurysms and cotton wool spots in the eye. There are diverse segmentation techniques recommended by several researchers. Azzopardi et al. [1] designed a B-COSFIRE filter. The B-COSFIRE operation is impacted by

T. Jemima Jebaseeli (✉) · J. Dinesh Peter
CSE, Karunya University, Coimbatore, India
e-mail: jemima_jeba@karunya.edu

C. Anand Deva Durai
College of Computer Science, King Khalid University, Abha, Saudi Arabia

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_15

the estimations of its parameters. Geetha Ramani et al. [2] presented the Principle Component Analysis (PCA) strategy for producing the element vector; K-means clustering is performed on this result to assemble the pixels cluster as either vessel or non-vessel. The classification of vessel cluster enhanced the accuracy but diminished the sensitivity measure at times. Mookiah et al. [3] have proposed the framework with 13 considerable features for the Probabilistic Neural Network (PNN) for retinal blood vessel segmentation and SVM [4, 5] to elite the best classifier. While everyone these techniques work only on the fundus images with no pigmentation [5, 6]; till there are issues in detaching the vascular tree with no discontinuities in distinguishing the small blood vessels. A few techniques are prescribed to enhance the parameters of PCNN to decrease the iteration [7] with a specific end goal to accomplish virtuous segmentation results. There is no method to consequently find the vital parameters of PCNN to achieve obvious image segmentation.

The proposed model segments the blood vessels on pathological images and to uncover the blood vessel at their cross-sections. The novelty of the proposed method is depicted as follows:

- PCNN's threshold limit is chosen iteratively by the seed point vessel pixel of the source image so that the image is segmented in one-time firing process and no particular rule is required for the iteration stop condition.
- The bidirectional searching procedure utilizes sufficient information from the image while segmentation. The global parameters are registered for each associated image. Subsequently, the system can standardize the coupled weights of the neurons to carry its recurrence.
- Better execution and expeditious performance compared to PCNN based segmentation algorithms which command the quantity of cycles and image entropy as iteration stop condition.

2 Materials

The fundus images are used for diagnosing Diabetic Retinopathy. The fundus camera captures the fundus images at a different field of views. The color fundus images are utilized for validation of the proposed approach. There are various researchers used the publically available databases: DRIVE, STARE, REVIEW, HRF, and DRIONS [8] for experimentation. We have used totally 201 fundus images from these databases.

3 Methods

The proposed method uses Adaptive Histogram Equalizer (AHE) for removing the illumination of images captured by various fundus cameras at different FoVs. The preprocessed images are segmented using unsupervised Pulse Coupled Neural

Network. The novelty of the proposed method is to optimize the parameters of PCNN using Genetic Algorithm. The optimized fitness value generates the feature vectors automatically for training the network to recognize the blood vessels. The tiny blood vessels are connected with the nearest neighbors by redefining the synaptic weight in the proposed method.

3.1 Segmentation Process

The proposed segmentation technique is shown in Fig. 1. The blood vessels are segmented in two phases. In the first phase, the Adaptive Histogram Equalizer (AHE) removes the illumination and improves the contrast of the retinal image. In the second phase, the PCNN model produces the feature vectors to locate the retinal blood vessels based on generated feature vectors. Finally, morphological processing enhances the edges of the segmented retinal blood vessels.

3.1.1 Preprocessing

The fundus image of the retina has RGB channel in the color spectrum. In which red and blue channel have more illumination compared to the green channel. Also, more intensity was found for blood vessels compared to the background. Hence, the green channel is considered for enhancement. The Adaptive Histogram Equalizer (AHE) is used for preprocessing the green channel of the retinal image.

$$EI = \left(\sum_{pixel' \varepsilon R(pixel)} \frac{G\left(N - N'\right)^{l}}{L^2} \right) * mi \tag{1}$$

where EI is the Enhanced image; the neighborhood pixel is $pixel'$; $pixel' \varepsilon R(pixel)$ is the window of neighborhood pixel with length L; $mi = 255$ is the maximum intensity of image; l is to control the contrast enhancement level of vessel pixels. The method inverts the intensity of the image and controls the illumination.

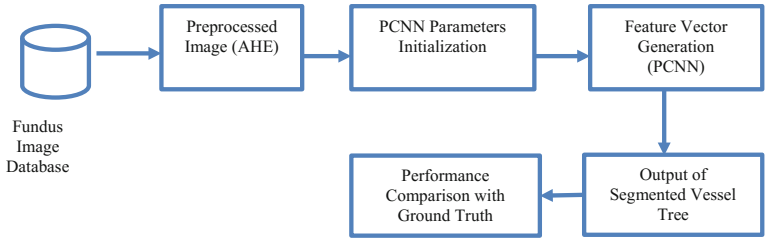


Fig. 1 The proposed architecture retinal blood vessel segmentation

3.1.2 Segmentation of Retinal Blood Vessels

PCNN is an unsupervised model and its parameters have to be set in advance to automatically generate the feature vectors. In the proposed model, the PCNN parameter template will be generated only one time, but it is suitable for different kinds of fundus database images. Based on the experimentation, the parameters are optimally tuned. The PCNN model receives stimulus from feeding and linking fields. The signal is generated until it exceeds the dynamic threshold value to produce the feature vectors. Depending on the iteration as well as the parameters, the dynamic output will be generated. The output signal insists the system to identify the edges of the blood vessels. The proposed PCNN model operates on diverse fundus imaging modalities since it is generic to its nature. Compared with the conventional PCNN models the proposed PCNN model has the following novelties:

- The scheming of the pulse threshold to fire the neuron is computed dynamically.
- The strength of the active neuron is redefined by the linking weight.
- The iterations minimize the computation time for segmenting the image.

The fitness function of the PCNN model is optimized through Genetic Algorithm by comparing the source image $Image_{i,j}$ with denoised image $Denoise\ Image_{i,j}$ at ith and jth iteration. The Mean Square Error of the reference image I_{MSE} is given as

$$I_{MSE} = 10\log\left(\frac{\left\|Image_{i,j} - DenoiseImage_{i,j}\right\|^2}{ImageSize_{i,j}}\right) \tag{2}$$

The fitness value of ith and jth iteration is H_{ij} and modified in the proposed approach as

$$H_{ij} = 10\log\left(\frac{Image\max^2}{I_{MSE}}\right) \tag{3}$$

The input retinal image with a grayscale value H_{ij} is taken as 2D matrix. The initial state of the neuron is set to 0. In the first iteration, internal activity of the neuron U_{ij} is equal to the input neurons H_{ij}. The neurons threshold attenuates from the initial value. The process continues until neurons threshold attenuated to smaller or equal to the corresponding U_{ij}. When the neuron is fired, output pulse $Y = 1$. Then, the dynamic threshold of E_{ij} increases and pulse output will be stopped. Then E_{ij} begin to attenuate and this attenuation continues until E_{ij} is smaller or equal to the corresponding U_{ij}. Hence the pulse output is generated again. The synaptic weight value of W_{ijkl} is defined in the proposed approach as, the reciprocal of Euclidean between (i, j) and (k, l) of the surrounding linking neurons.

$$W_{ijkl} = \frac{1}{\sqrt{(i-k)^2 + (k-l)^2}} \tag{4}$$

The linking neuron $L_{ij}(n)$ is to connect the blood vessel pixel with the neighborhood pixel is defined as

$$\mathrm{L_{ij}(n) = e^{-\alpha L}\, L_{ij}(n-1) + V_L \sum_{kl} W_{ijkl} = \frac{1}{\sqrt{(i-k)^2 + (k-l)^2}} Y_{kl}(n-1)} \quad (5)$$

where V_L is the voltage potential of linking neuron.

The active neuron U_{ij} is identified as

$$\mathrm{U_{ij}(n) = H_{ij}(n)\left(1 + \beta\, L_{ij}(n)\right)} \quad (6)$$

where β is the linking coefficient constant. The pulse generator produces the neuron Y_{ij} as output and its binary value to resolve the neuron status.

$$Y_{ij}(n) = \begin{cases} U_{ij}(n) - H_{ij}(n) & U_{ij}(n) > E_{ij}(n-1) \\ 0 & Otherwise \end{cases} \quad (7)$$

E_{ij} is the active threshold value of the network determines the threshold potential to fire the neuron.

$$E_{ij}(n) = \begin{cases} e^{-\alpha E} E_{ij}(n-1) & Y_{ij}(n) = 0 \\ V_E & Otherwise \end{cases} \quad (8)$$

where V_E is the voltage potential of the dynamic threshold.

The feature vector generated by PCNN is

$$G(n) = \sum_{ij} Y_{ij}(n) \quad (9)$$

The generated feature vector distinguishes the blood vessels from non-blood vessels.

4 Performance Analysis

The parameters utilized for evaluating the performance of the proposed methods are sensitivity, specificity, and accuracy. The proposed technique can separate all the small vessels with no deviation. The method is delicate to classify the non-vessel pixels. Subsequently, it is capable to categorize the non-vessel pixels from the real vessel pixels. The acquired performance measures of the proposed method are given in Table 1 illustrate that the proposed PCNN model is having improved sensitivity, specificity, and accuracy compared with the competitive methods. The proposed system is instigated using MATLAB R2010a and tested over DRIVE (40 images), STARE (20 images), REVIEW (16 images), HRF (15 images), and DRIONS (110 images) fundus image databases. But, many of the methods do not consider

Table 1 Segmentation performance of the proposed PCNN model

Datasets	Sensitivity (%)	Specificity (%)	Accuracy (%)	FPR
DRIVE	68.03	99.43	98.01	0.004
STARE	69.54	99.20	99.08	0.010
REVIEW	67.22	97.99	99.35	0.020
HRF	72.20	99.15	98.77	0.023
DRIONS	70.36	99.02	99.01	0.015
Average	69.47	98.95	98.84	0.014

large datasets. The focus of this method is to segment the retinal blood vessels automatically without training data. The segmented vessel map of the proposed method is shown in Fig. 2. The performance of the proposed method is matched with experts annotated ground truth images. The system achieved an average value of 69.47% sensitivity, 98.95%, specificity, and 98.84% accuracy.

5 Results and Discussions

The proposed system is implemented using MATLAB R2010a and tested over DRIVE (40 images), STARE (20 images), REVIEW (16 images), HRF (15 images), and DRIONS (110 images) fundus image databases. The segmented vessel map of the proposed method is shown in Fig. 2. The depigmentation found in the image at the vessel points are removed automatically by the PCNN model and enable the system to track all the tiny vessels unseen by this depigmentation. The obtained vascular tree structure of the input images are evaluated with the ground truth blood vessel image map and it indicates that more accuracy is found in the proposed retinal vessel segmentation method. The system achieved average value of 69.47% sensitivity, 98.95% specificity, and 98.84% accuracy.

In the literature, the PCNN parameters of 2D Otsu [9] take $V_\theta = 10{,}000$. Compared to 2D Otsu method, the proposed PCNN model's parameters take $V_\theta = 200$, also it takes only four iteration count for segmenting the image. The Otsu method has 0.0280 FPR whereas the proposed method has 0.014 FPR. CNN fails to capture some fine vessels around the optic disc [10–12]. The accuracy of random forest depends on classifier and their correlations. Each classifier has low correlations and unfit for classification and regression on images with tremendous noises [11, 13].

6 Conclusions

The limitation of the supervised method is that it needs labeled images to train the system. The supervised method is not suitable for clinical image analysis to predict the severity of Diabetic Retinopathy. The proposed unsupervised retinal blood vessel

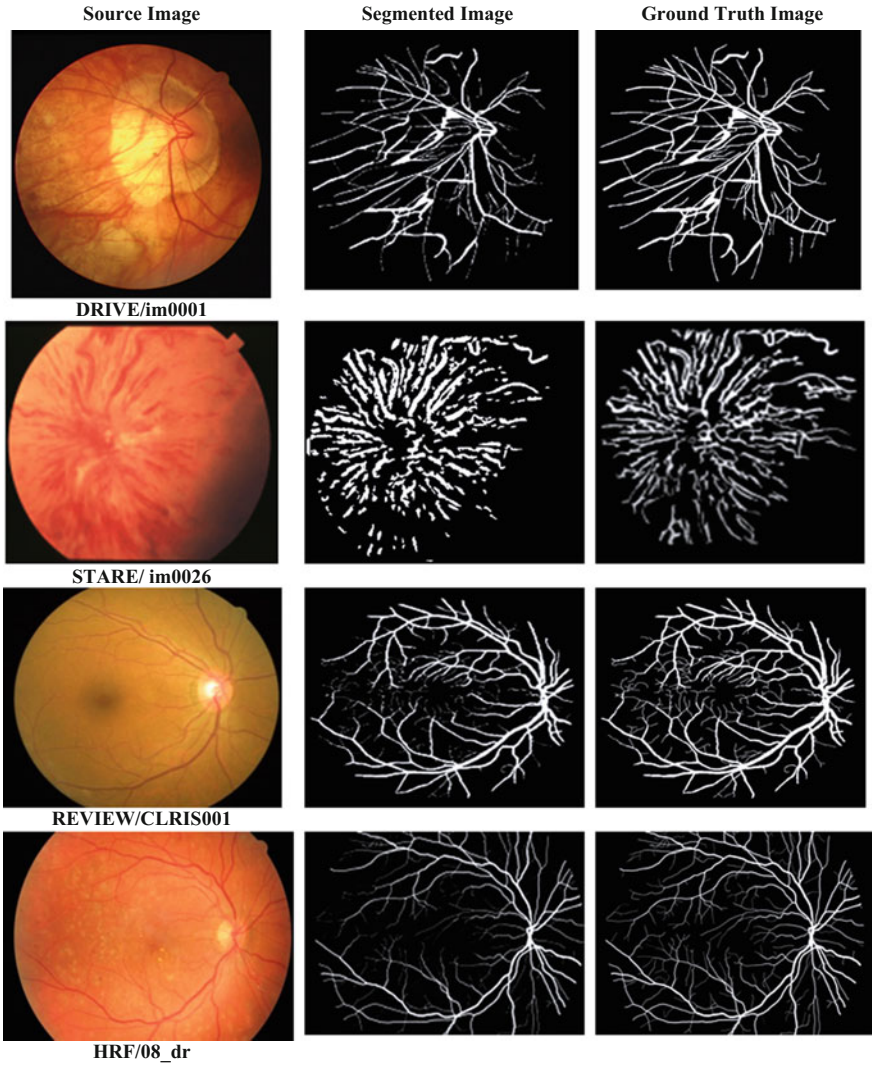


Fig. 2 Extracted vasculature structure of the proposed method of different databases

segmentation technique is a suitable investigative tool for early treatment. Also, the method eases the mission of the ophthalmologists in the diagnosis of the patients by considering the retinal blood vessels. An extensive evaluation was performed over all the fundus images of DRIVE, STARE, REVIEW, HRF, and DRIONS Databases. The end results prove that the proposed method detects all the blood vessels accurately.

References

1. Azzopardi G, Strisciuglio N, Vento M, Petkov N (2015) Trainable COSFIRE filters for vessel delineation with application to retinal images. Med Image Anal 19:46–57
2. Geetha Ramani R, Balasubramanian L (2016) Retinal blood vessel segmentation employing image processing and data mining techniques for computerized retinal image analysis. Biocybern Biomed Eng 36(1):102–118
3. Mookiah MRK, Acharya UR, Martis RJ, Chua CK, Lim CM, Ng EYK, Laude A (2013) Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: a hybrid feature extraction approach. Knowl Syst 39:9–22
4. Vega R, Sanchez-Ante G, Falcon-Morales LE, Sossa H, Guevara E (2015) Retinal vessel extraction using lattice neural networks with dendritic processing. Comput Biol Med 58:20–30
5. Rashmi Panda NB, Puhan GP (2016) New binary hausdorff symmetry measure based seeded region growing for retinal vessel segmentation. Biocybern Biomed Eng 36:119–129
6. Alipour SHM, Rabbani H, Akhlaghi M (2014) A new combined method based on curvelet transform and morphological operators for automatic detection of foveal avascular zone. SIViP 8:205–222. https://doi.org/10.1007/s11760-013-0530-6
7. Jemima Jebaseeli T, Sujitha Juliet D, Anand Devaduri C (2016) Segmentation of retinal blood vessels using pulse coupled neural network to delineate diabetic retinopathy, Springer, digital connectivity—social impact volume 679 of the series, Communications in computer and information science, pp 268–285
8. DRIVE: Digital Retinal Image for Vessel Extraction, http://www.isi.uu.nl/Research/Databases/DRIVE, STARE: http://cecas.clemson.edu/~ahoover/stare/, REVIEW: http://reviewdb.lincoln.ac.uk/Image%20Datasets/Review.aspx, HRF: https://www5.cs.fau.de/research/data/fundus-images/, DRIONS: http://www.ia.uned.es/~ejcarmona/DRIONS-DB.html
9. Yao C, Chen H (2009) Automated retinal blood vessels segmentation based on simplified PCNN and fast 2D-Otsu algorithm. J Cent South Univ Technol 16:0640–0646
10. Dashtbozorg B, Mendonça AM, Campilho A An automatic graph-based approach for artery/vein classification in retinal images. IEEE Trans Image Process 23(3), March 2014
11. Prasanna P, Jain S, Bhagatt N, Madabhushi A (2013) decision support system for detection of diabetic retinopathy using smartphones. In: 7th international conference on pervasive computing technologies for healthcare and workshops. https://doi.org/10.4108/icst.pervasivehealth.2013252093
12. Fraz MM, Welikal RA, Rudnicka AR, Owen CG, Strachan DP, Barman SA (2015) QUARTZ: quantitative analysis of retinal vessel topology and size—an automated system for quantification of retinal vessels morphology. Expert Syst Appl 42:7221–7234
13. Bourouis A, Feham M, Hossain MA, Zhang L (2014) An intelligent mobile based decision support system for retinal disease diagnosis. Decis Support Syst 59:341–350

A Novel Corner Elimination Method for the Compression of Wireless Capsule Endoscopic Videos

Caren Babu and D. Abraham Chandy

Abstract The paper presents a lossless compression algorithm relevant for hardware implementation in wireless capsule endoscopy. The video generated by the wireless capsule endoscopy is huge and in order to satisfy the transmission requirements, the data need to be compressed. However, due to the power and memory limitations, the traditional compression algorithms are not appropriate. Therefore, this paper proposes a method to address these limitations and presents a design suitable for wireless capsule endoscopy. A corner elimination algorithm is introduced to improve the compression ratio by an average of 3. The resulting algorithm produces an average compression ratio of 80.23 and an average PSNR of 31.7 dB.

1 Introduction

The recent advancement in medical field along with the growth of telemedicine is quite promising for the future wireless capsule endoscopy (WCE). WCE is a non-invasive method used to investigate the gastrointestinal (GI) tract by swallowing a capsule which travels through the body and captures images/videos in real time [1]. Then these images are transmitted to outside and displayed for diagnosis by a gastroenterologists. The capsule requires an average of 12–16 hours inside the GI tract for inspection. The transmission of this huge amount of data is a complex task in terms of bandwidth and cost. One way to take care of this problem is to have an efficient compression algorithm capable of reducing the data. However, it is very much important to keep the quality of the image intact as this would be a greater concern for the diagnosis of such images.

C. Babu (✉) · D. Abraham Chandy
Department of Electronics and Communication, Karunya Institute of Technology and Sciences, Coimbatore 641114, India
e-mail: carenbabu1@gmail.com

D. Abraham Chandy
e-mail: abrahamdchandy@gmail.com

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_16

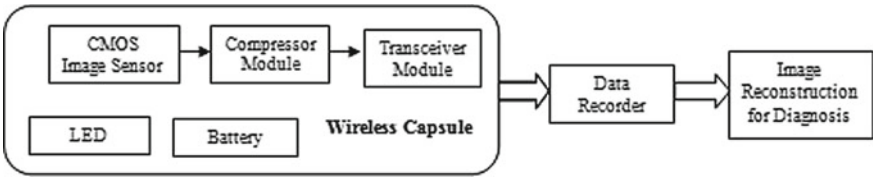


Fig. 1 Representation of a WCE system

The capsule consists of a CMOS image sensor, light emitting diode, compression module and a battery as shown in Fig. 1 [2]. The image data is captured using the sensor and is applied to the compression module. The LED is provided as a light source for the operation of the capsule. It is battery operated and hence the power requirements should be minimized while designing the modules inside the capsule. Therefore, WCE does not promote complex algorithms, high computations, increased memory requirements which results in constrains on hardware implementation. A minimum frame rate needed is 10 frames per second (fps) for diagnosis [3] and this is possible to achieve only by applying compression to the input data.

1.1 Motivation

The increased demands and growth of WCE and associated medical image processing really inspired us to develop a novel compression algorithm suitable for WCE images/videos. The implementation inside the capsule has put forward several constrains on the design of the compression algorithm. There are a few works focusing on developing compression algorithm for WCE. However, they mainly focused on Bayer CFA images taken directly from the sensor. Considering the existing literature and application requirements, our proposed algorithm is suitable for RGB image taking into consideration the properties of such images. The proposed approach is an efficient scheme in providing good compression ratio without discarding the quality of the image.

1.2 Paper Organization

The Sect. 2 gives an overview of the literatures on WCE compression. The method and material details are explained in Sect. 3 where the details of the database used and our proposed compression approach is explained step by step. The Sect. 4 explains the evaluation criteria used and the performance of proposed algorithm is discussed in Sect. 5. We conclude our work in Sect. 6.

2 Literature Review

There are a few works reported for developing a compression algorithm for WCE. It can be broadly classified as lossy and lossless based on the reconstructed image quality. The reconstructed image is not replica of the original image in lossy compression techniques. Discrete Cosine Transform (DCT) based algorithms are widely used [4–6], with certain modifications to reduce the associated computation. Though this can achieve a good compression performance, the reconstructed image quality is low. Lossless compression can overcome this disadvantage and is preferred for WCE compression. The lossless compression ensures that we can reconstruct the image exactly similar to the original one. The works done in lossless compression can be classified into two based on the type of input image, namely, Bayer Color Filter Array (CFA) image and RGB image.

The Bayer CFA image [7–9] is taken directly from the sensor output. The correlation between the neighboring pixels is less in this format and some transformation is applied so that the redundancy in the neighboring pixels can be utilized for compression. Though this yields good results, it does not consideration the properties of endoscopic images. Only a few works reported in the literature that works on RGB image. These works have used the endoscopic images available from the database [10]. Color transformation is applied prior to compression in certain works as chrominance based representation aids in compression [11]. It is reported to have developed a compression algorithm based on neighboring pixel correlation. It is observed throughout the literature that only a compression system that satisfies the design criteria discussed in our previous work [12] is acceptable.

3 Method and Material

3.1 Database Details

Based on the study and analysis of endoscopic images, we propose a compression algorithm suitable for RGB endoscopic images from the database [13]. So far, no works have been reported using this database. All the 147 images available are used in our study. The database is open access and contains high-resolution images taken from GI tract. Figure 2 shows the sample images.

3.2 Proposed System

Simplified block diagram of the proposed system is shown in Fig. 3. At first, the RGB input image is transformed to YUV color space. This is followed by an elimination module which removes the uninterested corner areas of the image. Thus, the image

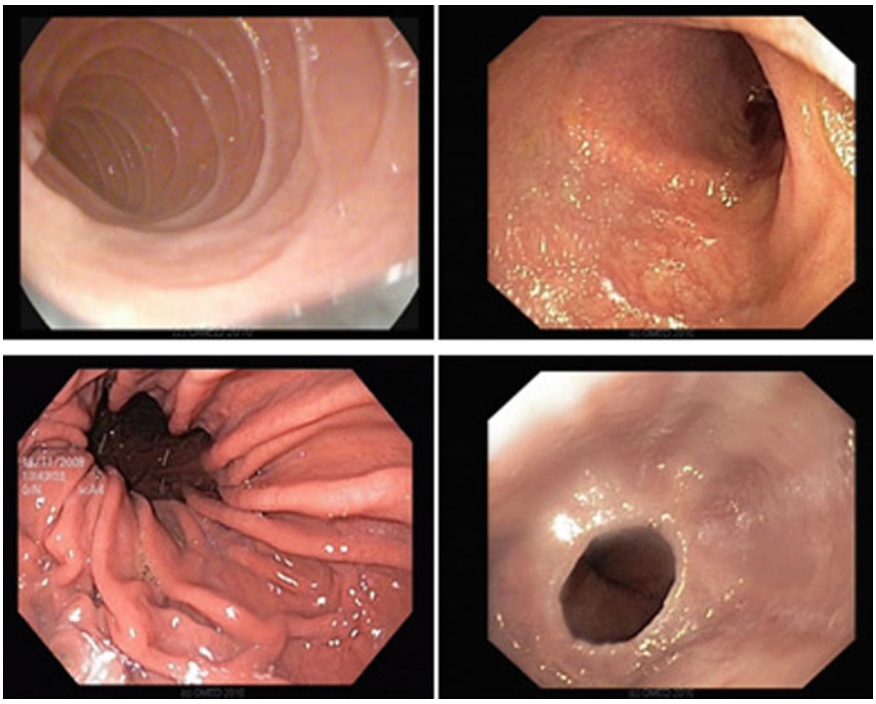


Fig. 2 Sample images of GI tract taken from [13]

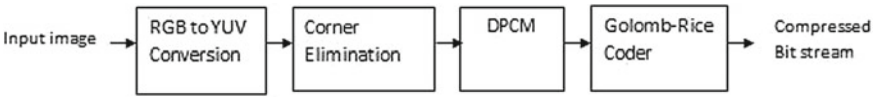


Fig. 3 Block diagram of the proposed system

compressor inside the capsule will not consider the corner pixels during compression. Due to the small variations between the neighboring pixels, differential pulse code modulation (DPCM) is a good choice as an encoder. Then, an entropy coder based on Golomb Rice coding is used to convert it into compressed bit stream. The different stages of proposed algorithm are discussed briefly below.

3.2.1 Color Transformation

The RGB input images of GI tract exhibits the dominance of red color compared to green and blue components. It is also noted that the blue component is absent in certain images. This is based on the study of the histogram of such images and calculation of entropy of each component as discussed in our previous work [12]. It

clearly indicates that the chrominance based representation yields good compression results. Therefore, RGB input image is transformed to YUV color space. In YUV, the luminance is stored in Y component, U and V represent the chrominance components and its relationship with RGB is given by the Eqs (1)–(3). The RGB to YUV conversion is theoretically lossless. At the receiver, it is possible to reconstruct RGB components from the YUV using the reverse equations [5].

$$Y = ((66R + 129G + 25B + 128)/256) + 16 \tag{1}$$

$$U = ((-38R - 74G + 122B + 128)/256) + 128 \tag{2}$$

$$V = ((122R - 94G + 18B + 128)/256) + 128 \tag{3}$$

3.2.2 Corner Elimination Algorithm

The image sensor is encased in the capsule and due to this corner areas are distorted. Here, we propose a corner elimination algorithm to eliminate this. Let (R_i, C_j) represent the position of any pixel in a given image. R_i represents the rows and C_j represent columns of a image, where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, N$ for a $N \times N$ input image. Now a circle of radius N/2 is drawn with center $(R_{N/2}, C_{N/2})$ in such a way that it eliminates the corner pixels that are blanked out. The corner elimination algorithm will select only the pixel values that fall inside the circle as shown in Fig. 4. The other regions which are outside the circle are eliminated and hence will not be taken for next stage.

The algorithm works as follows. Consider a N × N input endoscopic image. The pixels are taken from this image one after the other in a raster-scan fashion. The corner elimination algorithm will first check for the position of that pixel. Based on its

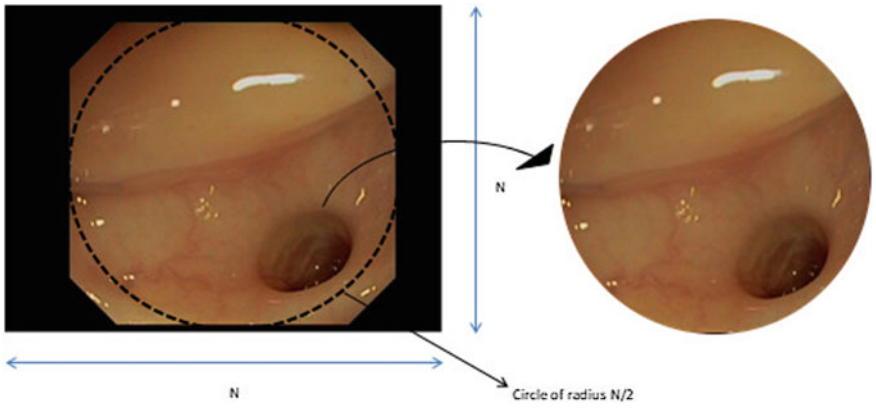


Fig. 4 Extraction of pixel region by applying corner elimination algorithm

position value, it will be either directed to the compression module or eliminated. The compression algorithm will then process only these selected pixels. This helps the algorithm to process reduced number of pixels and hence achieve higher compression performance.

3.2.3 DPCM

The study of endoscopic images clearly indicates that there are no sharp edges in the image. The component values are changing in a gradual pattern in a row. Therefore, it is advisable to transmit only difference between the pixels so that data to be transmitted is reduced. DPCM is a lossless encoder with reduced computation and is best suitable option here. The first pixel in a row is transmitted as such. Instead of transmitting the second pixel, the difference between first and second pixel is transmitted and so on. Let $Z(R_1, C_j)$ denote the pixel value of the element at first row and first column when j takes value 1 whereas $Z(R_1, C_{j+1})$ represents pixel value at first row and second column. Then the equation to compute DPCM is given as

$$dZ = Z(R_1, C_j) - Z(R_1, C_{j+1}) \tag{4}$$

This is done for all pixels in a row followed by next rows in the image. The computation is performed for all the three color planes namely Y, U and V. It is observed that applying DPCM is very effective in reducing the data with minimum visible degradation to the images.

3.2.4 Entropy Coding

The next step is to find a suitable entropy coder that can code the difference with fewer number of bits as possible. The dY, dU and dV values follow a two-sided geometric distribution and therefore, Golomb coder will give optimum code length. Considering the hardware constrains on WCE, Golomb Rice coder, a simple version of Golomb code is used in our work.

The difference value can be either positive, negative or zero. The Golomb Rice coder will work with only the positive values of input. Therefore, a mapping is performed where, positive values mapped to even integer and negative values to odd integer as given in Eq. (5). This is followed by a truncation to 255 which is expressed using 8 bits in binary. After that Golomb Rice coding algorithm explained in [14] is applied on the data.

$$dZ_{map} = \begin{cases} 2dZ, & if\ dZ \geq 0 \\ 2dZ - 1, & \text{otherwise.} \end{cases} \tag{5}$$

4 Evaluation Criteria

Considering the WCE application, we propose two parameters to evaluate the efficiency of our proposed compression scheme. The quality of the reconstructed image is analyzed using the peak signal to noise ratio (PSNR) expressed in dB. The PSNR is given by the Eq. (6) where MSE is the mean square error of the reconstructed image [11].

$$PSNR_{overall} = 20log_{10}(255^2/MSE) \ , \tag{6}$$

The compression rate (CR) is calculated using the number of bits required to represent the original image to that of the compressed bit stream using the expression given below. CR is expressed in percentage (%).

$$CR = \left[1 - \frac{Totalbitsaftercompression}{Totalbitsbeforecompression}\right] \times 100 \tag{7}$$

5 Performance Assessment

Images from the database (details given in Sect. 3) are taken directly and YUV color transformation is applied. This is followed by DPCM based prediction and Golomb Rice coding. The simulation results obtained by our proposed method is better than method in [15] as shown in the Table 1. In addition, by introducing the corner elimination algorithm discussed in Sect. 3.2.2, we could get an average improvement of 3% in CR values. However, the PSNR value remains unchanged. This clearly indicates that introducing corner elimination will assure an improvement in the compression performance.

The results obtained for WCE using the proposed compression algorithm is quiet promising to be implemented inside the capsule. A PSNR value above 30 dB indicates that it is possible to reconstruct the image with good quality capable of medical diagnosis. When compared with the other works implemented on WCE, the proposed system could produce competitive CR and PSNR values.

Table 1 Simulation results

Proposed algorithm	Avg. CR (%)	Avg. PSNR (dB)
Khan et. al [15]	76.24	27.13
Proposed algorithm without corner elimination	77.56	31.7
Proposed algorithm with corner elimination	80.23	31.7

6 Conclusion

In this paper, we present the design of a novel compression algorithm with an efficient corner elimination technique suitable for WCE. The input images from the database are first converted to YUV color space. The low complex DPCM algorithm is applied so to obtain the difference between neighboring pixels and encoded using Golomb Rice coding. A corner elimination algorithm is proposed to improve the results in terms of CR. The proposed system provides a CR of 80.23% and PSNR of 31.7 dB which is acceptable for medical diagnosis. Compared to other works targeted to WCE, the proposed algorithm gives fairly good results and is promising for future endoscopy.

References

1. Slawinski PR, Obstein KL, Valdastri P (2015) Capsule endoscopy of the future: whats on the horizon? World J Gastroenterol 21(37) 10528–10541
2. Angermann Q, Histace A, Romain O, Dray X, Pinna A, Granado B (2015) Smart video Capsule for early diagnosis of colorectal cancer: toward embedded image analysis. In: Computational Intelligence in Digital and Network Designs and Applications. Springer, New York, NY, USA, pp 325–350
3. Mylonaki M, Fritscher-Ravens A, Swain P (2003) Wireless capsule endoscopy: a comparison with push enteroscopy in patients with gastroscopy and colonoscopy negative gastrointestinal bleeding. Gut 52(8):1122–1126
4. Turcza P, Duplaga M (2007) Low-power image compression for wireless capsule endoscopy. In: Proceedings of 2007 IEEE international workshop on imaging systems and techniques, pp 1–4
5. Wahid K, Ko S, Teng D (2008) Efficient hardware implementation of an image compressor for wireless capsule endoscopy applications. In: Proceedings of IJCNN, pp 2761–2769
6. Peterson HA, Peng H, Morgan JH, Pennebaker WB (1991) Quantization of color image components in the DCT domain. In: Proceedings of SPIE human vision, visual processing, and digital display II, vol 1453, pp. 210–222
7. Chen X, Zhang X, Zhang L, Qi N, Jiang H, Wang Z (2008) A wireless capsule endoscopic system with a low-power controlling and processing ASIC. In: Proceedings of IEEE Asian solid-state circuits conferences, pp 321–324, Nov 2008
8. Xie X, Li G, Chen X, Li X, Wang Z (Nov. 2006) A Low-Power Digital IC Design Inside the Wireless Endoscopic Capsule, IEEE Journal of Solid State Circuits 41(11):2390–2400
9. Chen SL, Liu TY, Shen CW, Tuan MC (2016) VLSI implementation of a cost-efficient near-lossless CFA image compressor for wireless capsule endoscopy. IEEE Access 4:10235–10245
10. Gastro lab Available online: http://www.gastrolab.net
11. Khan TH, Wahid KA (2011) Lossless and low power image compressor for wireless capsule endoscopy. VLSI Des 2011:1–12
12. Babu C, Chandy DA (2017) DPCM based compressor for capsule endoscopic videos. In: ICSPC'17, pp 89–93, July 2017. https://doi.org/10.1109/CSPC.2017.8305813
13. WCE Clinical Endoscopy Atlas Available online: http://www.endoatlas.org/
14. Gallager RG, van Voorhis DC (1975) Optimal source codes for geometrically distributed integer alphabets. IEEE Trans Inf Theory 21(2):228–230
15. Khan TH, Wahid KA (2014) Design of a lossless image compression system for video capsule endoscopy and its performance in in-vivo trials. Sensors 14(11):20779–20799

Prediction of Two Year Survival Among Patients of Non-small Cell Lung Cancer

Yash Dagli, Saumya Choksi and Sudipta Roy

Abstract Lung cancer ranks as second most prevalent type of cancer. Still predictions for survival of lung cancer patients are not accurate. In this research, we try to create a prediction model, with the help of machine learning to accurately predict the survival of non-small cell lung cancer patients (NSCLC). Clinical data of 559 patients was taken for training and testing of models. We have developed multilevel perceptron model for survival prediction. Other models developed during this study were compared to measure performance of our model. Attributes that are found to be useful as biomarkers for prediction of survival analysis of NSCLC have also been computed and ranked accordingly for increase in accuracy of prediction model by implementing feature selection method. The final model included T stage, N stage, Modality, World Health Organization Performance status, Cumulative Total Tumor dose, tumor load, Overall treatment time as the variables. Two year survival was chosen as the prediction outcome. Neural Network was found as the best prediction model with area under Curve (AUC) of 0.75. By far to our knowledge Multilevel Neural Network is found to be the best model for predicting two-year survival among patients of non-small cell lung cancer.

Keywords Multilevel neural network · Non-small cell lung cancer
Machine learning · Feature selection · ReliefF · Survival prediction

Y. Dagli (✉) · S. Choksi
Department of Computer Science and Engineering, U.V. Patel College of Engineering, Ganpat University, Kherva, Mehsana 384012, India
e-mail: yashdagli98@gmail.com

S. Choksi
e-mail: saumyachoksi@gmail.com

S. Roy
Department of Computer Science and Engineering, Institute of Computer Technology, Ganpat University, Kherva, Mehsana 384012, India
e-mail: sudiptaroy01@yahoo.com

S. Roy
Department of Computer Science and Engineering, Calcutta University Technology Campus, JD-2, Sector-III, Salt Lake City, Kolkata 700098, India

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_17

1 Introduction

Lung cancer is second most prevalent type of cancer. With 2,34,030 new cases being recorded alone in United States as per 2018 [1]. Non-small cell lung cancer shares 80–85% from the reported cases [1]. One of the critical concerns for a patient diagnosed with cancer is the survival expectancy [2]. At present 5 year survival expectation of lung cancer patients is 15%, stage I has 5 year survival expectation of 70%, 45% in stage II and 10–30% for stage III, while that for IIIB and IV are respectively 10% and < 5% [1].

Though there have been notable advancements in technology still survival prediction uses basic methods. Currently TNM staging is used for survival prediction, although its primary use was to suggest operability scheme that can be performed on patient [3]. Where T stage describes the extent and size of main (primary) tumor, N stage refers to nearby (regional) lymph nodes that have cancer, M stage refers to whether the cancer has metastasized. Other predictive models such as Naïve Bayes model, Random Forest, Support Vector Machine have been proposed previously for survival prediction [3, 4] still a desirable accuracy is not achieved. Feature selection in previously developed models such as, Naïve Bayes model was based on expert opinion [3], while Random Forest model implemented feature selection by using ReliefF algorithm [4].

This study aims to improve the 2-year survival prediction for patients suffering from NSCLC. Features which seem to best predict the outcome have been identified by implementing ReliefF and Recursive Feature Elimination algorithms, upon these selected features prediction models have been developed using Logistic regression and Neural Network. The results show that Neural Network is the best predictor model for our problem with the input variables (features) selected by ReliefF algorithm. Further the features have also been ranked according to their share in predicting the outcome by use of Recursive Feature Elimination.

2 Methods

2.1 Data

Clinical data of five hundred and fifty-nine patients that have underwent treatment of chemotherapy or radiotherapy has been taken [3]. This data is available publicly at [5]. The data includes 41 features. Two-year survival is taken as the study endpoint. The data needed to go through a data cleaning step due to presence of noise (missing values). Data cleaning is a crucial step in machine learning process [6]. Mean binning method has been applied for data cleaning (i.e. to remove missing values), but new values (missing values) derived by data cleaning are guesses made based on the mean value of data, thus these values are not entirely reliable. It is necessary for data to be consistent while applying machine learning algorithms to get better results [7],

Table 1 Statistics of data used in this study i.e. of Maastro Clinic

	Count	Mean	Standard deviation	Min	Max
intake_who	239	1.108787	0.807454	0	4
T_stage	239	1.698745	1.092968	0	3
N_stage	239	1.556485	1.150425	0	3
Modality	239	1.259414	0.593793	0	2
TotalTumorDose	239	65.09414	11.63597	3.42	95.52
FEV	239	77.53305	21.64322	29	138
Tumorload	239	112.8200	126.3878	0.9813	775.01
Ott	239	31.05857	8.139844	1	96
TwoYearSurvival	239	0.410042	0.492873	0	1

so the missing values have been removed. Finally after performing the data cleaning process final dataset contains 239 patient values. Data taken for model development has the statistics shown in Table 1.

2.2 Model Development

Systematic procedure is followed for model development, as shown in Fig. 1. The steps include a Data cleaning step where the data is transformed according to the suitability of the algorithm. The next step includes Feature selection, as data includes many features it is necessary to know which feature best predicts the output.

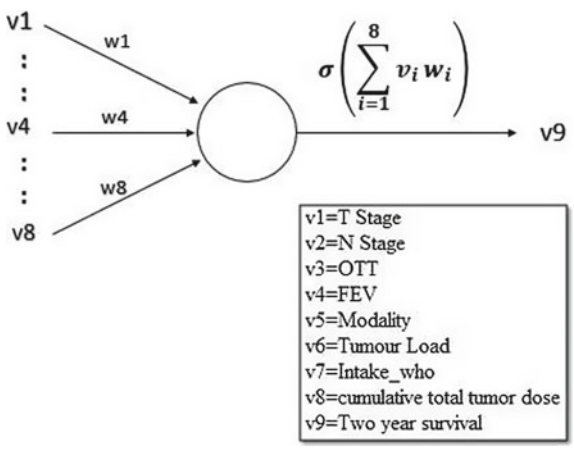


Fig. 1 Linear classifier model network structure

These features selected play as input to the model. Next the model is developed and trained upon these features. Finally, the model is evaluated against the training dataset. TensorFlow is used for model development.

2.2.1 Feature Selection

During the Feature Selection process a subset of features from all features are selected. It is used to make the data interpretable and insightful while reducing the dimensionality [8]. Feature selection different has methods such as filter methods, wrapper methods, embedded scheme [9]. We have implemented Recursive Feature elimination and ReliefF as our feature selectors.

ReliefF
It works by adjusting weights of features, by comparing between class and within class distance from neighbouring samples [8]. Following is a pseudo code for implementation of ReliefF [10].

1. *Set all weights W[A]:=0.0;*
2. *For i:=1 to m do begin*
3. *Randomly select an instance* R_i
4. *Reclassify* R_i *instances and build a* Z_i *dataset.*
5. *Calculate* $k(Z_i)$ *as a k percentage of the minority class in* $Z_{i.}$
6. *Find* $k(Z_i)$ *nearest hit Hj and* $k(Z_i)$ *near miss* $M_{j.}$
7. *For A:=1 to a do*
8. $W[A] = W[A] - \sum_{j=1}^{k(Z_i)} \frac{Diff(A,R_i,H_j)}{[k(Z_i)*m]} + \frac{Diff(A,R_i,M_j)}{[k(Z_i)*m]}$

where W[A] denotes the weight of attributes, m is the nearest neighbours selected, R_i is a randomly selected instance from Z, Z is a set on instances, Hj and M_j are number of hits and miss ($k(Z_i)$). Diff function is used to calculate the difference between attribute A and instance R_i. ReliefF algorithm selected intake_who, T_stage, N_stage, Modality, CumultativeTotalTumorDose, FEV, tumorload, ott as our features.

Recursive Feature Elimination
Recursive Feature elimination is a type of wrapper method. It creates a linear SVM (support vector machine) model for each iteration, where SVM is a classification algorithm. For each iteration it works by removing the feature has the least significance, this process is repeated until all features in dataset are exhausted [11]. After all features have been removed features are ranked according to when they are removed, the later the feature is removed the higher the rank it attains. Ranking of intake_who, T_stage, N_stage, Modality, CumultativeTotalTumorDose, FEV, tumorload, ott was found to be in the order of 4, 6, 8, 2, 5, 7, 1, 3 respectively, with tumorload ranked as first.

2.2.2 Perceptron Network

The first model we have created is a linear classifier. This is simplest type of neural network, with a single-layer perceptron. It uses the equation Eq. 1 for passing values from input layer to the next layer, where w_i represents the weights, x_i the input attributes and b is the bias term [12].

$$z = \sum_{i=1}^{n} w_i x_i + b \tag{1}$$

In this network the first layer is the input units layer and second layer is both the output layer as well as hidden layer. We have taken features selected by reliefF as input units. Learning rate was set to 0.2 as it is said to be optimal. Activation function as sigmoid function and cross entropy as cost function, as it learns fast if the error is large as compared to quadratic cost function. Sigmoid function has the equation shown in Eq. 2, it was taken as cost function (i.e. z) as it maps the input in range of 0–1 which helps in classifying it to an output class. Network structure of Linear Classifier used for development of our model is shown in Fig. 2. Further it is possible to derive the derivative of logistic regression used for calculation of gradient descent, the derivative is shown in Eq. 3. Cross Entropy cost function uses equation shown in Eq. 4, which allows for faster learning, because of the fact that larger the difference the faster it learns, in Eq. 4 p is the value derived from Eq. 2 and y is the true value.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{2}$$

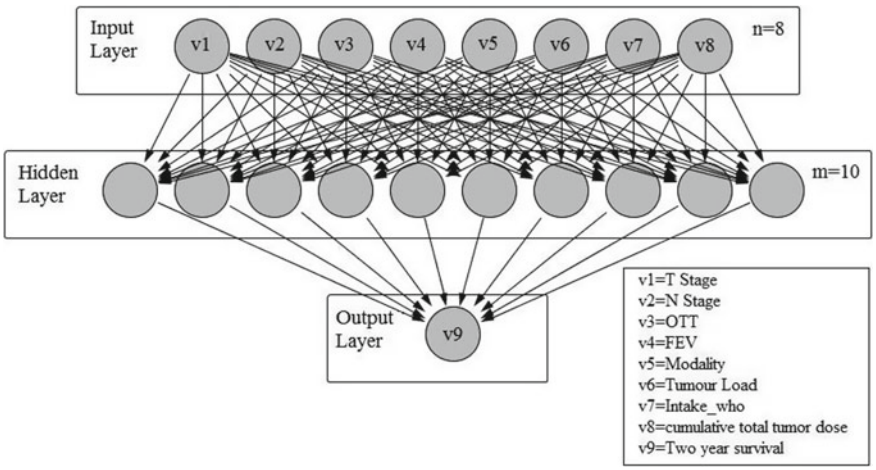


Fig. 2 Network structure of neural network with single hidden layer of 10 neurons, abbreviations: n = number of input nodes, m = number of output nodes

$$g'(logistic) = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) \tag{3}$$

$$cross\ entropy = -(ylog(p) + (1-y)log(1-p)) \tag{4}$$

An improvement in the linear classifier is a single layer neural network with n neurons. Our implementation or Neural Network has 1 hidden layer with 10 neurons. It has a learning rate of 0.05, with rectified linear unit (ReLU) as activation function and softmax cross entropy as cost function. ReLU is based on the equation shown in Eq. 5.

$$\sigma(z) = \max(0, z) \tag{5}$$

It returns either a max value or 0 according to the input, i.e. 0 for negative values and the input value for positive value.

$$p_j = \frac{e^{a_i}}{\sum_{k=1}^{n} e_k^a} \tag{6}$$

$$\frac{\partial p_j}{\partial a_j} = \begin{cases} p_i(1-p_j)\ if\ i = j \\ -p_j * p_i\ if\ i \neq j \end{cases} \tag{7}$$

$$H(y, p) = -\sum_i y_i \log(p_i) \tag{8}$$

Softmax Cross Entropy is used as the cost function. Softmax is based on the equation shown in Eq. 6, it takes a N-dimensional vector and transforms it into a vector of real number in range (0, 1), where a is the value of nth vector. Derivate of softmax is shown in Eq. 7. Cross Entropy finds the distance between what the model predicted as the true value and the real true value, it is defined in Eq. 8. Structure of our Neural Network is shown in Fig. 2.

3 Results

A train test split of 90–10% is taken. Min-Max normalization is applied for data transformation. Confusion Matrix of each model has been computed for assessment of the model. It uses four variables for assessment of model. True positives (TP) depicts the positive tuples that were correctly labelled as positive by the classifier. True negative (TN) represents the negative tuples that were correctly labelled as negative by the classifier. 3. False Positives (FP) are the negative tuples that were classified as positive by classifier. 4. False Negative (FP) are the positive tuples that were falsely classified as negative.

Comparison with previously developed models for two year survival suggests Neural Network outperforming other models. SVM model has AUC of 0.59 [13],

Table 2 Confusion matrix for logistic regression abbreviations: TN = True Negative, TP = True Positive, FN = False Negative, FP = False Positive

n = 239	Predicted: NO	Predicted: YES	Total
Actual: NO	TN = 115	FP = 26	141
Actual: YES	FN = 52	TP = 46	98
Total	167	72	239

Table 3 Confusion matrix for Linear Classifier

n = 239	Predicted: NO	Predicted: YES	Total
Actual: NO	TN = 110	FP = 31	141
Actual: YES	FN = 47	TP = 51	98
Total	157	82	239

Bayesian network of Jayasurya et al. has AUC of 0.56 [14], Bayesian network of Arthur Jochems et al. has AUC of 0.66 [3]. Function corresponding for plotting a logit function is shown in Eq. 9. Output class is then decided by applying Eq. 9 to input variables and then Eq. 10 to the output of Eq. 9 get the final prediction. The t in the Eq. 10 represents the output of Eq. 9. The features selected by Feature Selection algorithm are taken as input variables in our model. This model was created for comparison purpose.

Plane created to separate the classes has a intercept of −1.7257 and coefficients for intake_who is 0.15505293, T stage is −0.23676378, N stage is 0.41042416, Modality is −0.00192756, TotalTumorDose is 0.01071944, tumorload is −0.00377834 and that of ott is 0.0210605 respectively. Accuracy of 0.6736 is achieved for this model.

$$\beta + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n \tag{9}$$

$$P_+ = \frac{e^t}{1 + e^t} \tag{10}$$

Confusion Matrix for logistic regression is shown in Table 2. Table 2 shows that from total of 239 values represented as n, logistic regression identified 46 (TP) true positive out of 98 positive values (i.e. The patient survived, total actual yes), whereas 115 (TN) true negative from 141 negative values (i.e. The patient died, total actual no). While the False Negative (FN) and False Positive (FP) were found to be 52 out of 98 and 26 out of 141 respectively.

Confusion matrix for single perceptron neural network model is shown in Tabel 3. We tried to remove the features that ranked the lowest in the feature selection, expecting to get a better result, but the same accuracy was reached proving that the model assigned a negligible weight to these features. This proves that Linear Classifier weighs the features as per their importance in predicting the output.

Table 3 shows that from total of 239 values represented as n, logistic regression identified 51 (TP) true positive out of 98 positive values (i.e. The patient survived, total actual yes), whereas 110 (TN) true negative from 141 negative values (i.e. The

patient died, total actual no). While the False Negative (FN) and False Positive (FP) were found to be 47 out of 98 and 31 out of 141 respectively. Comparatively Single Perceptron Neural Network is found to be better than logistic regression.

Confusion Matrix of multilayer neural network model is shown in Table 4. Different choices of artificial neurons as well as for number of hidden layers are taken. ReLU was used as the activation function with learning rate of 0.2 and one hidden layer with ten neurons. Increasing both the number of neurons and hidden layer for our model deteriorated the outcome. This proved that increasing the hidden layers or the number of neurons does not necessarily increase the outcome.

Table 2 shows that from total of 239 values represented as n, logistic regression identified 67 (TP) true positive out of 98 positive values (i.e. The patient survived, total actual yes), whereas 116 (TN) true negative from 141 negative values (i.e. The patient died, total actual no). While the False Negative (FN) and False Positive (FP) were found to be 31 out of 98 and 25 out of 141 respectively, it is found that this model performed better than any previously developed model.

Table 5 shows comparison of models developed in this study along with their measures derived from confusion matrix, where Accuracy shows how frequently is the classifier predicts correctly, Misclassification rate depicts how frequently it is false, True Positive Rate shows when the output is actually 1, how many times does the model classify it as 1, False Positive Rate shows when the output is actually 0, how frequently does the model classify it as 1, Specificity is when the output is actually 0, how frequently does it predict 0, Precision **is** when it predicts 1, how often is it right. Accuracy of Multilayer Neural Network is 0.7656 which is greater than other models.

Table 4 Confusion matrix for multilayer neural network

n = 239	Predicted: NO	Predicted: YES	Total
Actual: NO	TN = 116	FP = 25	141
Actual: YES	FN = 31	TP = 67	98
Total	147	92	239

Table 5 Comparison of models developed in this study

Total = 239	Logistic regression	Single perceptron neural network	Multilayer neural network
Area under the ROC curve (AUC)	0.642	0.650	0.753
95% confidence interval	0.578 to 0.703	0.586 to 0.711	0.693 to 0.806
Accuracy	0.6736	0.6736	0.7656
Misclassification rate	0.3263	0.3263	0.2343
True positive rate	0.4693	0.5204	0.6836
False positive rate	0.1843	0.2198	0.1773
Specificity	0.8156	0.7801	0.8226
Precision	0.6388	0.6219	0.7282

4 Conclusion

This main objective of this method was to develop an accurate prediction model for two year survival prediction of patients who have suffered from non-small cell lung cancer. We propose a model to predict the two-year survival by use relief with Multilayer Neural Network. ReliefF should be used for feature selection while neural network should be adopted to develop the prediction model for two-year survival prediction of NSCLC patients. This study solves problems relating to the prediction of two-year survival of non-small cell lung cancer. Using neural network to predict two year survival of non-small cell lung cancer is a novel approach that has not been worked on which provides better result compared to previously developed models.

References

1. Siegel RL, Miller KD, Jemal A (2018) Cancer statistics. CA Cancer J Clin 68(1):7–30
2. Clément-Duchêne C, Carnin C, Guillemin F, Martinet Y (2010) How accurate are physicians in the prediction of patient survival in advanced lung cancer? Oncol Express 15(7):782–789
3. Jochems Arthur et al (2017) Developing and Validating a Survival Prediction Model for NSCLC Patients Through Distributed Learning Across 3 Countries. Int J Radiat Oncol *Biol* Phys 99(2):344–352
4. Mei X (2017) Predicting five-year overall survival in patients with non-small cell lung cancer by reliefF algorithm and random forests. In: 2017 IEEE 2nd advanced information technology, electronic and automation control conference (IAEAC)
5. https://www.cancerdata.org/publication/developing-and-validating-survival-prediction-model-nsclc-patients-through-distributed. Accessed Jan 2018
6. Devi S et al (2015) Study of data cleaning and comparision of data cleaning tools. Int J Comput Sci Mob Comput 4(3):360–370
7. Li X, Shi Y, Li J, Zhang P (2007) Data mining consulting improve data quality. Data Sci J 6
8. Wang Z et al (2016) Application of ReliefF algorithm to selecting feature sets for classification of high resolution remote sensing image. In: 2016 IEEE international geoscience and remote sensing symposium (IGARSS)
9. Ladha L (2011) Feature selection methods and algorithms. Int J Comput Sci Eng 3
10. Bolón-Canedo V, Sánchez-Maroño N, Alonso-Betanzos A (2013) A review of feature selection methods on synthetic data. Knowl Inf Syst 34
11. Lin G et al (2012) A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. J Chromatogr B Anal. Technol Biomed Life Sci
12. Yuan LCJ, GX, HCH (2012) Recent advances of large-scale linear classification. In: Proceedings of the IEEE
13. Dehing-Oberije C et al (2009) Development and external validation of prognostic model for 2-year survival of non-small-cell lung cancer patients treated with chemotherapy. Int J Radiat Oncol Biol Phys 74:355–362
14. Jayasurya K et al (2010) Comparison of Bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. Med Phys 37:1401–1407

Prediction of Chronic Kidney Diseases Using Deep Artificial Neural Network Technique

Himanshu Kriplani, Bhumi Patel and Sudipta Roy

Abstract The progression of the chronic kidney disease and methodologies to diagnose chronic kidney disease is a challenging problem which can reduce the cost of treatment. We studied 224 records of chronic kidney disease available on the UCI machine learning repository named chronic kidney diseases dating back to 2015. Our proposed method is based on deep neural network which predicts the presence or absence of chronic kidney disease with an accuracy of 97%. Compared to other available algorithms, the model we built shows better results which is implemented using the cross-validation technique to keep the model safe from overfitting. This automatic chronic kidney disease treatment helps reduce the kidney damage progression, but for this chronic kidney disease detection at initial stage is necessary.

Keywords Chronic kidney diseases · Deep neural network · Random forest Support vector machine · Gradient descent · Optimization

H. Kriplani (✉) · B. Patel
Department of Computer Science and Engineering, UV Patel College of Engineering, Ganpat University, Mehsana 384012, Gujarat, India
e-mail: kriplani.himanshu96@gmail.com

B. Patel
e-mail: patelbhumi0112@gmail.com

S. Roy
Department of Computer Science and Engineering, Institute of Computer Technology, Ganpat University, Mehsana 384012, Gujarat, India
e-mail: sudiptaroy01@yahoo.com

S. Roy
Department of Computer Science & Engineering, Calcutta University Technology Campus, Kolkata 700098, India

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_18

1 Introduction

Chronic kidney disease (chronic kidney disease) [1] is a disease in which the kidneys gets damaged or deep neural network to filter blood and so the waste fluids that body produces remain inside, which further causes other health problems. Blood helps the body organs to function better, therefore, it is very important to have it clean and pure; if our kidneys will not work it becomes a major concern. This damage happens over many years. More the damage, less the kidneys function and therefore makes the body unhealthy. It is becoming a major threat in the developing and undeveloped countries. Its main cause for occurrence is diseases like diabetes, high blood-pressure. Other risking circumstances causing chronic kidney disease include heart disease, obesity, and a family history of chronic kidney disease. Its medications which are dialysis or kidney transplant are very costly and so we need an early detection. In the United States (US) [2], about 117,000 patients developed end-stage renal disease (ESRD) requiring dialysis, while more than 663,000 prevalent patients were on dialysis in 2013. 5.6% of the total medical budget was spent for ERDS in 2012 which is about $28 billion. In India, CKD is widespread among 800 per million populations and ESRD is 150–200 per million populations.

Hybrid Modified cuckoo search-neural network in chronic kidney disease classification [3] had given a higher accuracy but they have not defined the data preprocessing steps. A modified SVM [4] was used to increase the higher accuracy. They [4] have stated data processing and classification of support vector machine and used a random forest instead. The data preprocessing technique used by the paper above is different from what we have used. The paper above has also not stated how the value of accuracy has been calculated in it. We have taken most prominent factors with bias (parameters) as given in [5] that are: Blood-Pressure, Serum Creatine, Pack Cell Volume, Hypertension Factors, and Anemia Factors. They [5] have used K-nearest neighbor and we have used deep artificial neural network which can be more robust for a huge amount of data. Performance of KNN would degrade within crease in size of dataset and also it sometimes gets biased for some attributes. Therefore, many algorithms like naive Bayes, Support Vector Machine (SVM), and Artificial Neural Network have contributed in its recognition. From this we are going to focus on deep neural network. A deep neural network is a computational model based functions provided by the biological neural networks. It consists of nodes which in the biological analogy represent neurons, connected by arc. It corresponds to dendrites and synapses. Each arc associated with a weight while at each node.

The aim of our study is to use various algorithms on the dataset of chronic kidney disease and explore them to find outperforming algorithm for detection [3]. The rest of the paper is organized as follows, Sect. 2 describes the theoretical background for classifiers used in this study for chronic kidney disease, Sect. 3 describes experimental studies and Sect. 4 concludes the paper.

2 Methodology

2.1 Data Set

This study makes use of the dataset from the UCI Machine Learning Repository named Chronic Kidney Disease uploaded in 2015 [5]. This dataset has been collected from the Apollo hospital (Tamilnadu) nearly 2 months of period and has 25 attributes, 11 numeric and 14 nominal. The attributes and its description of those attributes which our study consists are mentioned in appendix. Total 400 instances of the dataset in which 224 of them are used for the training to prediction algorithms, out of which 105 has label chronic kidney disease (chronic kidney disease) and 119 has label non chronic kidney disease (nonchronic kidney disease).

2.2 Deep Neural Network

The deep learning model built using artificial neural networks [4] are considered very robust. The network consists of interconnected neurons which are mapped using mathematical functions to each other. To use a NN we have to first train the model which is finding of the proper weights involved in the mathematical function for mapping. In the second phase we test our model with already classified dataset to find the accuracy of our dataset.

Neurons have a basis in Biology. This neuron in deep learning is termed as perceptron but both are used interchangeably. A single neuron takes inputs with weights associated with each weight. We use activation function for the output of the neuron.

$$a = f\left(\sum_{i=0}^{N} WiXi\right) \tag{1}$$

Here, f is the activation function [6]. W and X are weights and Inputs. Activation function can be Gaussian, Hyperbolic, logistic or even linear function. Activation function is capable of estimating complex nonlinear relationships making neural network flexible. There are many activation functions available.

Sigmoid function is an activation function having a output in the range of [0, 1]; whereas tanh function has an output range of [−1, 1]. Relu function is a type of activation function returning the max(0, x). Relu and tanh tend to have best performance. There are many more activations functions.

$$f(x) = \frac{1}{1 + e^{-(x)}}$$

A neural network as shown in Fig. 1 is build using number of layers. These layers are built of number of neurons. Each neural network is divided into three sections.

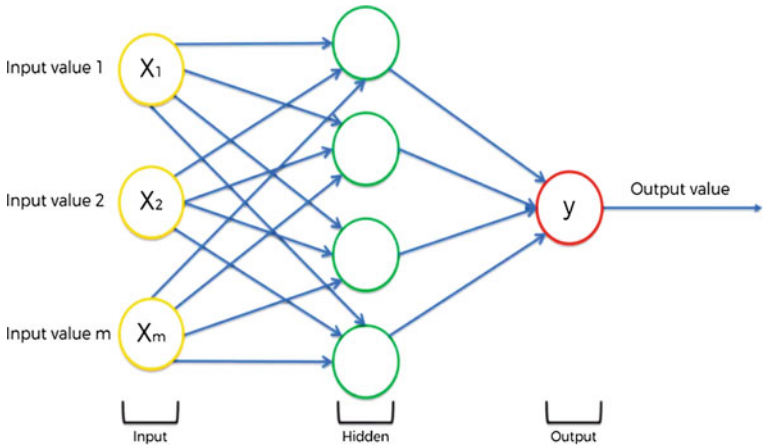


Fig. 1 Structure of deep neural network

First is the input layer, second is the output layer and between these two layers there are hidden layers. The activation function activates the neuron and which in result passes information to next layers.

Input layer is set of neurons which take the actual input data for processing. The size that is the number of neurons can be decided in accordance with the input data. There can be number of hidden layer. Overfitting [7] means that the model we build only works well for our dataset. The model will not result with the same accuracy for the other dataset which is distinct from our dataset when the model is overfitting. It occurs when the model shows high variance and low bias. Overfitting can be precluded by using techniques such as validation or cross-validation. The model can also underfit if the number of neurons and the number of hidden layers as well are properly selected. Underfitting models deep neural network to capture the underlying trend in the data. It shows low variance and high bias.

The output layers contain the number of neurons equal to the number of classes we want as output. We use relu and tanh for input and hidden layer as this keeps back propagation more fruitful. The output layer can have any activation functions but sigmoid is the most used activation function for output layer. There is a cost involved in calculations is the difference between the real and predicted values. Thus we need to reduce the cost of the model that is increase the accuracy of the predictions. There are many functions available to calculate the cost of the model. To decrease the cost we use various optimization techniques explained as under.

The computation from input layer up to output layer is known as forward propagation. After the computation it is possible that estimated output is far away from actual output resulting high error. To cope up with this problem we go back and change the weights to get the least error possible. Such a process is known as backward propagation. A round of forward and backward propagation is termed as epoch. This process is repeated to a set of weights which results into highest accuracy.

To train our model we have used three layers. For input and the hidden layer, "Relu" is used as activation function. Relu outputs in 0 or 1. For the output layer, Sigmoid function is used which outputs into range of [0, 1]. Units are the number of the perceptron in a layer. Initializer is the definition of how the weights should be initialized. Input dimension is only defined in the first layer which shows how much parameters will be taken as input.

2.3 Optimization

Gradient descent [8] is one of the known techniques for cost reduction for epochs. Here as shown in the figure cost function is plotted and to find the minimum of the cost we study the slope of the cost function at that point as shown in Fig. 2. If the slope is negative we go downhill and if it is positive we go up hill. This way we find the least cost values.

This technique works when the cost function drawn is in convex shape. But for cost functions which are not convex we use stochastic gradient descent [9] which finds the local minima of the cost function stochastically as shown in Fig. 3. Finally we sum up the result to find the global minima of the cost function.

We have used adam (adaptive moment estimation) which computes independent adaptive learning rates or the weight which needs to be adapted for different parameters. At the end we implemented K-fold cross-validation [10]. It partitions the dataset randomly into K sub-equal parts. Of the K equal parts, one of it is used for testing the model and remaining k − 1 to train the model.

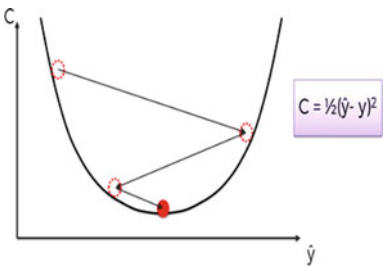


Fig. 2 Convex cost function graph

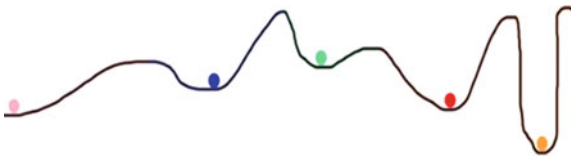


Fig. 3 Nonlinear cost function

3 Result and Discussion

The model created using the 18 parameters and after implementing the concepts explained above we get an accuracy of 97.76%, F1 measure of 97.6% as seen in Table 1. We implemented other algorithms on the dataset and found out following True positive rate (TPR), false positive rate (FPR), Precision, Recall and F measure [11, 12, and 13].

The value of True positive rate is average of all true positive values in each epoch. Similarly the value of false positive rate is calculated for each algorithm.

Precision shows how many percentages of tuples that the classifier labeled as positive are actually positive that is how many records were correctly classified into chronic kidney disease by the model we built. Recall is the percentage of positive tuples did the classifier label as positive that is the percentage of records classified as CKD. Harmonic mean of precision value and recall value is known as F measure.

Artificial neural network build by us has the highest precision, recall, and true positive value compared to other implementations. Naïve Bayes has an equivalent precision. The neural network identifies 95.2% presence of chronic kidney disease, i.e., the true positive rate of the model. With more and more data, the model can be trained and may result in greater accuracy with faster results without any human intervention. The comparison is as shown in Table 1.

Table 1 Accuracy and performance by class

		True positive	False positive	Precision	Recall	F-measure
Chronic kidney disease	Naïve Bayes	0.952	0	1	0.952	0.976
	Deep neural network	0.952	0	1	0.952	0.976
	Logistic	0.943	0	1	0.943	0.971
	Random forest	0.952	0	1	0.952	0.976
	Adaboost	0.962	0	1	0.962	0.981
	SVM	0.962	0	1	0.962	0.981
Nonchronic kidney disease	Naïve Bayes	1	0.048	0.96	1	0.979
	Deep neural network	1	0.048	0.96	1	0.979
	Logistic	1	0.057	0.952	1	0.975
	Random forest	1	0.048	0.96	1	0.979
	Adaboost	1	0.038	0.967	1	0.983
	SVM	1	0.038	0.967	1	0.983

Table 2 Kappa statistics, roc and rmse

	Root mean squared error	Accuracy (%)	Kappa statistics	ROC
Naïve Bayes	0.1407	97.7679	0.9551	1
Deep neural network	0.1214	97.7679	0.9551	1
Logistic	0.1582	97.321	0.946	0.99
Random forest	0.0753	99.1071	0.9821	1
Adaboost	0.1131	98.2143	0.9641	0.99
SVM	0.1336	98.2143	0.9641	0.98

Fig. 4 Loss function plot

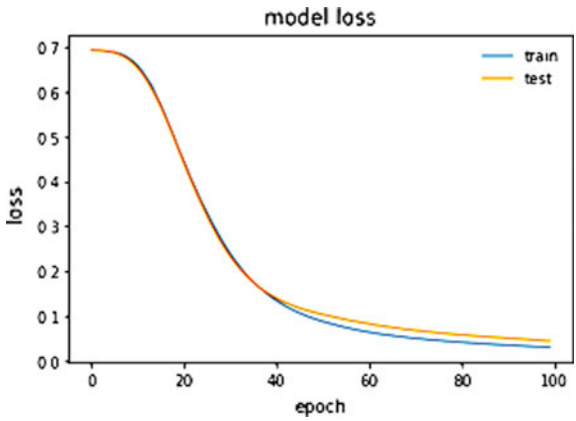


Table 2 shows the information about the accuracy of all algorithms with kappa statistics, roc, and root mean square error. The graph shown in Fig. 4 shows the decrease in loss of the model with gradual increase in the number of epochs. The loss function of the training set tends to zero and for the testing set tends to a bit greater than zero. The root mean squared error deal with the real magnitude of the data, thus it gives a magnitude regardless of negative or positive values of error which is observed in prediction of the model. Kappa statistics shows the real percent raters with a score of 0.95 which is near perfect agreement that is the prediction of the chronic kidney disease. Kappa statistics shows the real percent raters with a score of 0.95 which is near perfect agreement that is the prediction of the chronic kidney disease. Receiver operating characteristics curve is the area under the curve of the plot is false positive rate v/s true positive rate. We get almost full area covered under curve thus its value is 1 for our neural network model.

The confusion matrixes of all implemented algorithms [14] are given in Table 3. The deep neural network has 119 true positive predictions and 100 true negative predictions. The test set used for the prediction was distinct from that of the training dataset. This table describes that 119 rows of data were correctly classified into a not chronic kidney disease and 100 where classified correctly into a 100 chronic kidney

Table 3 Confusion matrix for the comparable method

	NOT chronic kidney disease	Chronic kidney disease
Naïve Bayes	119(TP)	0(FN)
	5(FP)	100(TN)
Deep neural network	119(TP)	0(FN)
	5(FP)	100(TN)
Logistic	119(TP)	0(FN)
	6(FP)	99(TN)
Random forest	119(TP)	0(FN)
	2(FP)	103(TN)
Adaboost	119(TP)	0(FN)
	4(FP)	101(TN)
SVM	119(TP)	0(FN)
	4(FP)	101(TN)

diseases. Thus out of 224 record 219 (true negative and true positive) are correctly classified in our neural network.

In Naïve Bayes the confusion matrix is similar to that of the model we built but the root mean squared error of Naïve Bayes is greater than our deep learning model. No occurrences of a particular attribute values and the class label results into probability estimate equal to zero. According to Naive Bayes conditional independence assumption, posteriori probability will be affected as we get zero by multiplying all the probabilities. Logistic regression has false positive higher than that of ours and also it has not covered 100% of the area under the ROC curve and it requires large sample size to achieve stable results. The random forest has performed well with lesser false negative rate while AdaBoost has covered only 99.2% area under the ROC curve. But this AdaBoost is sensitive to noisy data, outliers and susceptible to the overfitting problem than proposed algorithms. Support vector machine has performed well but suffers from the determination of the parameters for a given value of the regularization and kernel parameters and choice of kernel. Here SVM moves the problem of overfitting from optimizing the parameters to model selection.

4 Conclusion

We trained a neural network model to predict the presence of chronic kidney disease. It uses a list of parameters. If this model can be well trained using a varied range of parameters, it may result in more accurate predictions. Clinics and hospitals can use this for faster and digitized methodology for prediction of chronic kidney disease. Of all the other models compared, deep neural network is the best one. The results

would be more promising with increase in dataset. Thus it has outperformed other classifiers and is able to detect the chronic kidney disease more efficiently.

References

1. Webster et al (2017) Chronic kidney disease. The Lancet 389(10075):1238–1252
2. https://www.kidney.org/news/newsroom/factsheets/End-Stage-Renal-Disease-in-the-US. Accessed Jan 2018
3. Chatterjee et al (2017) Hybrid modified cuckoo search-neural network in chronic kidney disease classification. In: 14th international conference on engineering of modern electric systems (EMES), pp 164–167
4. Ahmad M et al (2017) Diagnostic decision support system of chronic-kidney-disease using SVM. In: 2017 second international conference on informatics and computing (ICIC). IEEE
5. Roy Sudipta et al (2016) Brain tumor classification using adaptive neuro-fuzzy inference system from MRI. Int J Bio-Sci Bio-Technol SERSC 8(3):203–218
6. Wang SC (2003) Artificial neural network. In: interdisciplinary computing in java programming. The Springer international series in engineering and computer science, vol 743. Springer, Boston, MA
7. Dua D, Karra Taniskidou E (2017) UCI machine learning repository (http://archive.ics.uci.edu/ml). Irvine, CA: University of California, School of Information and Computer Science
8. Zhang C, Woodland PC (2015) Parameterised sigmoid and reLU hidden activation functions for DNN acoustic modelling. In: INTERSPEECH-2015, pp 3224–3228
9. Domingos P (2000) Bayesian averaging of classifiers and the overfitting problem. In: Proceedings of the international conference on machine learning (ICML), pp 223–230
10. Narendra K, Parthasarathy K (1991) Gradient methods for the optimization of dynamical systems using neural networks. IEEE Trans Neural Netw 2(2):252
11. Roy S et al (2017) Heterogeneity of human brain tumor with lesion identification, localization, and analysis from MRI. Inf Med Unlocked. https://doi.org/10.1016/j.imu.2018.02.006
12. Patel B et al Computerized skin-cancer lesion identification using combination of clustering and entropy. In: IEEE 2017 international conference on big data analytics and computational intelligence, proceedings IEEE, pp 89–94, 23–25 March 2017. Andhra Pradesh, India
13. Roy S et al (2017) Artifacts and skull stripping: an application towards the preprocessing for brain abnormalities detection from MRI. Int J Control Autom SERSC 10(4):147–160
14. Mathews et al (1993) A stochastic gradient adaptive filter with gradient adaptive step size. IEEE Trans Signal Process 41:2075–2087

Monitoring Acute Lymphoblastic Leukemia Therapy with Stacked Denoising Autoencoders

Jakob Scheithe, Roxane Licandro, Paolo Rota, Michael Reiter, Markus Diem and Martin Kampel

Abstract For acute lymphoblastic leukemia treatment monitoring, the ratio of cancerous blood cells, called Minimal Residual Disease (MRD), is in practice assessed manually by experts. Using flow cytometry, single cells are classified as cancerous or healthy, based on a number of measured parameters. This task allows application of machine learning techniques, such as Stacked Denoising Autoencoders (DSAE). Seven different models' performance in assessing MRD was evaluated. Higher model complexity does not guarantee better performance. For all models, a high number of false positives biases the predicted MRD value. Therefore, cost-sensitive learning is proposed as a way of improving classification performance.

1 Introduction

In childhood Acute Lymphoblastic Leukemia (ALL), a cancer that affects the generation pathway of a specific white blood cell type, the therapy response is assessed by examining these cells one-by-one using Flow CytoMetry (FCM). This technique enables the representation of each measured cell in a blood sample by parameters (also called features) describing physical and biochemical properties which are obtained by optical measurements. *Minimal Residual Disease (MRD)* is used to assess a patient's response to treatment and consequently, to guide and adapt the medication intensity [1]. It is defined as the ratio between blasts (cancer cells) and

J. Scheithe · R. Licandro (✉) · M. Reiter · M. Diem · M. Kampel
Computer Vision Lab, Institute of Visual Computing and Human-Centered Technology, TU Wien, Favoritenstrasse 9/193-1, 1040 Vienna, Austria
e-mail: licandro@cvl.tuwien.ac.at

R. Licandro
Computational Imaging Research Lab, Department Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Lazarettgasse 14, 1090 Vienna, Austria

P. Rota
Central Research Lab, Pattern Analysis and Computer Vision, Istituto Italiano di Tecnologia, Via Morego, 30, 16163 Genova, Italy

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_19

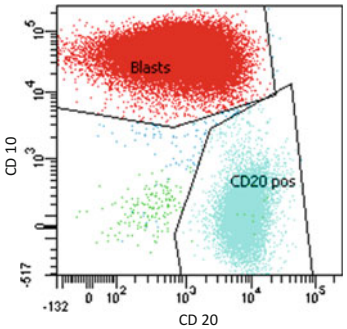


Fig. 1 FCM scatter plot of the features CD 10 and CD 20. Blast populations are illustrated in red and the drawn polygons represent gates of interest in the manual gating procedure

all cells observed within a FCM measurement [2]. Operators identify blast cells by displaying pairs of parameters (features) for the whole sample in scatter plots and by manually drawing borders to separate (gate) in a defined gating hierarchy healthy from cancerous cells. In Fig. 1 an example for a scatter plot is visualized. Red dots mark the identified blast cells and other colors the non-blast cells. This procedure is strongly dependent on the operator's expertise and is highly subjective. The challenges in MRD assessment arise from the detection of small cancerous cell populations in late therapy stages, which compose around 0.1% of the amount of all cells observed. Additional challenges lie in the limited amount of blood available in the test tube, as well as treatment-dependent and age-related variances. Artificial Networks, specifically with deep architectures, gained recently attention in the medical and machine learning community, according to the ability to abstract data on a higher level and to improve prediction and classification tasks [3]. In the imaging domain, Convolution Neural Networks (CNN) are a powerful technique for image segmentation [4, 5], object recognition [6, 7], or for predicting semantic descriptions from medical images [8]. Within this work, the architecture named Stacked layout of Autoencoders (SAE) [9] is used for the classification of FCM data. We used SAE instead of CNN for their easier adaptability to different data types. SAE are trained to encode the input in a way to learn important structures among features [10] and consequently to be able to solve the binary (Blast, Non-Blast) classification problem for FCM measured blood cells in ALL [11].

The contribution of the work is three fold: (1) we provide an automatic MRD assessment pipeline for FCM measurements of ALL samples using SAE, (2) we created three distinct Stacked Auto Encoder architectures with different levels of complexity, network shape and parameter choices and (3) compared and evaluated the presented techniques against a simple Fully Connected (FC) network substantially.

This paper is organized in the following way: First, the created and evaluated SAE architectures for MRD assessment are introduced. Subsequently, the evaluation setup and obtained results are summarized. This work concludes with a discussion and outlook for future work.

2 Methodology

An SAE is an artificial neural network that consists of two parts and is trained in two distinct steps. The first part is a stack of *AutoEncoders (AE)*, that are trained in an unsupervised manner. The second part is an ordinary *MultiLayer Perceptron (MLP)*. The composite network, consisting of the stack of AEs and the MLP is trained in the second, supervised phase. The FCM data used as input, consists of up to 300,000 measured cells per patient, with nine parameters (dimensions of input layer) assessed per cell (cf. Sect. 3). An AE is a three-layer, fully connected Artificial Neural Network (ANN) with an input layer, followed by a so called *encoder* layer and a *decoder* layer. The decoder has the same size as the AE's input dimension. The loss function is the mean squared error between input and decoder output and the activation function the *rectified linear unit (ReLU)*

$$f(x)_j = \max(x_j, 0). \tag{1}$$

Thus, the AE effectively tries to learn the identity function on all observed training data [12]. The series of AEs is trained one-by-one using backpropagation: After the i-th AE is trained, its decoder layer is omitted and the encoder activation is fed as the input to the $i + 1$-th AE, which is trained thereat. This way, the input is handed down through this *stacked autoencoder (SAE)* and the activation of the last AE's encoder layer is called the activation of the stack. Usually, the encoders are smaller than the decoders, such that the dimension of the input is reduced step by step, possibly resulting in the extraction of useful features. This architecture (however with *cross entropy*, Eq. 2, as a loss function) was proposed in [13]. In the second—supervised—phase, the trained SAE's activation is passed to an MLP with one hidden layer and output dimension k equal to the number of classes in the classification task. In our case, we have $k = 2$ classes: "cancerous" and "healthy". The combined network consisting of the SAE and the MLP is an MLP itself. After SAE pre-training, the composite network is trained via backpropagation to approximate the class labels. See Fig. 2 for an illustration of the unsupervised and supervised training process.

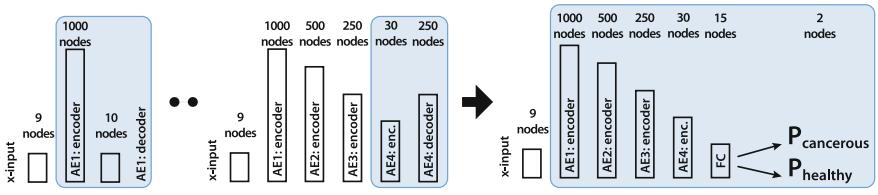


Fig. 2 Unsupervised (first and second drawing) and supervised (right drawing) training. The units, whose weights are being changed are marked by a blue box

The loss function is *softmax cross entropy*, that, for binary classification (where $y_1 = 1 - y_2$ and $\hat{y}_1 = 1 - \hat{y}_2$), takes the following form, see [14, Eq. 5.21].

$$L(\hat{\mathbf{y}}, \mathbf{y}) = -y_1 \ln \hat{y}_1 + (1 - y_1) \ln(1 - \hat{y}_1). \tag{2}$$

For the output layer, the *softmax* activation function [14, Eq. 5.150] is used

$$\sigma(\mathbf{a})_j = \frac{e^{a_j}}{\sum_{k=1}^{2} e^{a_k}}, \quad j = 1, 2. \tag{3}$$

Softmax activation is a common choice for classification tasks, because the outputs sum up to 1 and allow a probabilistic interpretation

$$\sum_j \sigma(\mathbf{a})_j = 1, \qquad \sigma(\mathbf{net}(\mathbf{x}))_j \equiv P(\text{“x belongs to class j”}). \tag{4}$$

Here, $\mathbf{a} = \mathbf{net}(\mathbf{x})$ is the net activation for data point $\mathbf{x}$ of the composite network. In all other layers throughout the network, ReLU activation is used. Once the network is trained, the softmax activations $\hat{\mathbf{y}} \in [0, 1]^k$ can be interpreted as predicted likelihoods for an unseen data point $\mathbf{x}$ to belong to the classes $1, \ldots, k$. The index of $\max(\hat{y}_i)_{i=1,\ldots,k}$ is called the predicted class for $\mathbf{x}$. In the following, we often use the term SAE to refer to the combination of SAE and MLP. As in every (deep) ANN and in particular because the AEs approximate the identity function, overfitting can prevent good generalization to unseen data [14, 1.3]. Two methods were tested to address this:

Weight decay promotes sparsity in the model parameters by penalizing large weights. During unsupervised training, the norm of each AE's encoder and decoder weight vectors is added to the loss function, multiplied by a factor β. Thus, minimizing the loss function leads to a decline of the weights, such that they tend to be close to zero. In this case the L_2-Norm is used as in [12]. See Fig. 3 for a visualization of the impact of this regularization term.

In a *denoising stacked autoencoder* (DSAE), noise is added to the input of the encoder layer. Still, the decoder output is compared to the uncorrupted input, when calculating the mean square error loss. This way, the data is augmented by

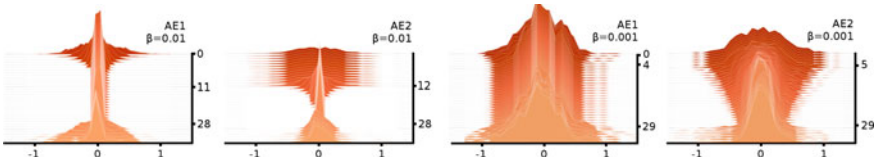


Fig. 3 Histograms of weight distributions changing throughout the training process (from back to front) for an SAE with two AE elements. **Left pair**: Strong L_2-regularization ($\beta = 0.01$) during unsupervised phase, steps 0 – 11 (AE1) and 12 – 28 (AE2). **Right pair**: Moderate regularization ($\beta = 0.001$), visible during steps 0 – 4 (AE1) and 5 – 29 (AE2)

Table 1 Evaluated network architectures

Denotation	No. of AEs	Encoder sizes	MLP size
Full-size (D) SAE	4	[1000, 500, 250, 30] nodes	15 nodes
Mid-size (D) SAE	3	[500, 100, 20] nodes	15 nodes
Small (D) SAE	2	[50, 15] nodes	15 nodes
MLP only	–	–	15 nodes

incorporating randomness. Rather than learning an identity mapping specifically for the presented training data, we expect each AE to extract more significant features, that are still detectable in noisy data. Two different kinds of noise were used as in [15]: *Masking noise*, whereat a certain fraction $1 - \alpha$ of the input elements is randomly chosen and set to zero and *Gaussian noise*. For the latter, a normally distributed random number $\varepsilon \sim \mathcal{N}(0, \sigma)$ is added to each input element.

Four different network shapes with different levels of complexity were evaluated. They are defined in Table 1. For the full-size network, masking noise, Gaussian noise and L_2-regularization were tested as well as an unregularized model. These are denoted as "DSAE (mask)", "DSAE (gauss)", "Sparse SAE" and "SAE (no reg.)". An automatized approach to optimize hyperparameters was used. In *hyperopt* [16], the search space is defined by giving distributions, from which the parameters are sampled. A real valued function of the parameters describing the model performance (e.g., prediction accuracy) defines the optimization task. When searching for minima of this function, hyperopt uses a *Tree of Parzen Estimators (TPE)* as an adaptive exploration strategy. The TPE algorithm starts by randomly sampling a number of parameter combinations. For each parameter x, one *Gaussian Mixture Model (GMM)* $l(x)$ is fitted to the outcome of the trial runs with the best parameter choices and one GMM $g(x)$ to the rest of the trials. The parameters for the next iteration are chosen by maximizing $l(x)/g(x)$, the probability of the next iteration to belong to the first (well performing) group and not to the second one [17]. For details, see [18].

3 Results

The data pool was annotated FCM data from day 15 of the ALL treatment, consisting of blood samples from 200 patients. Treatment followed the AIEOP-BFM 2009 protocol.[1] On average, each sample consists of roughly 240.000 flow cytometry measurements (single cells) with manually annotated classifications. For training, since the relative number of cancerous cells was low (0.85–11%), balanced sets were generated by randomly omitting events from the prevailing class in each sample. For testing unbalanced data was used.

[1] For the conducted randomized clinical trial AIEOP-BFM 2009, approximately 2000 ALL patients between age 1–18 years in 20 countries in and outside Europe were observed per year (see http://www.bfm-international.org/trials.php [assessed 2018-04-14]).

Table 2 Average classification performance during tenfold cross-validation. The best value for each measure is printed bold

Method	Precision	Recall	F1-score
DSAE (mask)	**0.9166**	0.9726	**0.9419**
DSAE (gauss)	0.9009	**0.9896**	0.9385
Sparse SAE	0.8974	0.9721	0.9296
SAE (no reg.)	0.9110	0.9696	0.9373
Midsize SAE	0.9014	0.9640	0.9281
Small SAE	0.9032	0.9735	0.9346
MLP only	0.9076	0.9695	0.9342

For each cell, FCM measurement yields nine parameters, consisting of two optical (FSC-A, SSC-A) and seven fluorescence based parameters (CD20, CD10, CD45, CD34, SYTO41, CD19, CD38). Each parameter defines one dimension in the nine-dimensional data space. Spillover compensation via a correction matrix provides statistical independence of the data, such that partial overlapping of fluorescence spectra of different fluorochromes can be resolved. Finally, parameter values are normalized to the range between 0 and 1. The dataset used in this work was generated in collaboration with experienced clinicians from the Children's Cancer Research Institute in Vienna. All participants' guardians (parents) and patients were informed about the aim of the study and gave their written, informed consent prior to inclusion. The capacity of the models was assessed based on average precision, recall and F1-score during tenfold cross-validation.[2] The results are presented in Table 2.

For all tested models, recall was higher than precision. More than 96% of the cancer cells were identified correctly, while almost 10% of the predicted cancer cells were false positives. This might be related to the balanced training and very unbalanced test data [19, 8.6.5]. Since the amount of cancer cells in the test data is low (0.85–11%), a low total number of false positives occurs (e.g., in the set with true MRD of 0.85%: 10 k out of 3.8 m false positives, or 0.26% false positive rate), such that these cases of type I error have low impact on the overall cost function during training. Still, precision is low (in this example 75%) and the predicted MRD (1.1%) considerably higher that the actual MRD (0.85%). In other words, the error in predicted MRD, which is *the* critical quantity in assessing treatment success, is not adequately reflected in the cost function, guiding the model in a suboptimal direction during training. While recall was similar for higher and lower MRD in the test data, precision varied between 75% for the set with 0.85% MRD and 97.5% for the set with 11% MRD. The denoising SAE using masking noise achieved the highest F1-score and precision. Surprisingly however, the margin to the networks with few units is small: The one-layer MLP with only 15 hidden nodes (182 trainable parameters) achieved an F1-score only 0.0077 lower than the best model. Overall, the differences between the tested models were small. We conclude, that high network

[2] See [19, 8.5.1] for a definition of the performance metrics.

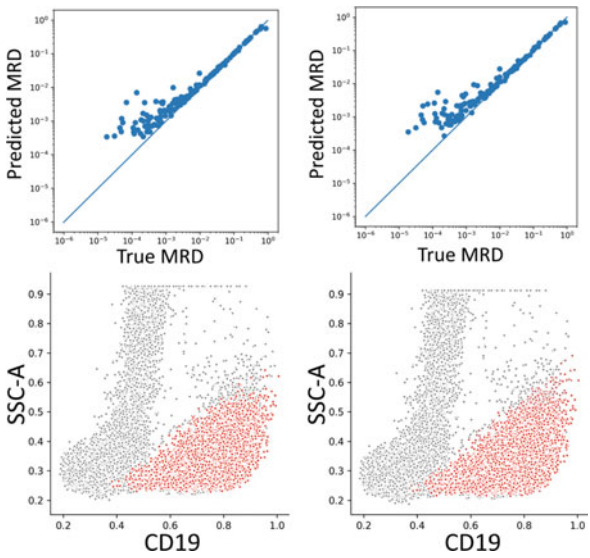


Fig. 4 **First row**: True versus predicted MRD on logarithmic scale for DSAE (mask) to the left and MLP only to the right. Each dot represents one of the 200 patients. **Second row**: Scatter plots as used for manual gating with predicted cancer cells in red and predicted healthy cells in gray. Again, results for DSAE (mask) to the left, MLP only to the right

complexity is not necessary for this task, but other means are needed, especially addressing the bias towards false positive classification. Beyond the common performance metrics, determining MRD is the critical task in assessing ALL treatment response. MRD is defined as the ratio of cancerous and all cells observed, $\text{MRD} = \frac{N_{\text{cancerous}}}{N_{\text{allCells}}}$. Again, comparison of the best ranking model and the smallest network in Fig. 4 shows no substantial distinction.

4 Conclusion

For current MRD assessment in ALL patients via FCM measurement, a highly subjective, manual gating procedure is used. Automatic classification of single cells is desirable for an unbiased, reproducible assessment. One proven classification system in machine learning is SAE. We described training and classification procedure for SAEs in detail and presented three different regularization strategies to mitigate overfitting: masking as well as Gaussian noise and L_2-regularization. We defined four distinct network architectures of different complexity. Hyperparameters were optimized using TPE algorithm. Evaluation of classification performance and capacity to assess MRD was presented. Network complexity did not affect classification performance significantly. Considering that the full size (D)SAE already incorpo-

rates 644760 learnable parameters, further increasing network complexity does not seem promising for this task. An important difficulty to tackle is the low precision. One way to address this issue is *cost-sensitive learning*, whereby type I and type II error are weighted differently in the cost function (see [19, 8.6.5]). This becomes especially important with FCM data at a later time in treatment, when MRD declines ever more. After all, an MRD of 0.1% is still considered high, whereas values towards 0.001% correspond to good treatment success [20].

Acknowledgements This work was supported by the European Commission FP7-PEOPLE-2013-IAPP 610872 and by ZIT Life Sciences 2014 (1207843). We want to thank Michael Dworzak and Angela Schumich at the Children Cancer Research Center in Vienna for annotating and providing the FCM data.

References

1. Theunissen P, Mejstrikova E et al (2017) Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia. Blood 129(3):347–357
2. Brüggemann M, Schrauder A, Raff T et al (2008) Standardized MRD quantification in European ALL trials: proceedings of the second international symposium on MRD assessment in Kiel, Germany, 18–20 Sept 2008. (2010) Leuk Off J Leuk Soc Am Leuk Res Fund 24(3):521–535
3. Greenspan H, van Ginneken B, Summers RM (2016) Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging 35(5):1153–1159
4. Guo Z, Li X, Xiang et al (2017) Medical image segmentation based on multi-modal convolutional neural network: study on image fusion schemes. arXiv:1711.00049
5. Moeskops P, Viergever M, Mendrik A et al (2016) Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imag 35(5):1252–1261
6. He K, Zhang X, Ren S, Sun J (2014) Spatial pyramid pooling in deep convolutional networks for visual recognition. In: ECCV. Springer
7. Zhang Y, Sohn K, Villegas R, Pan G, Lee H (2015) Improving object detection with deep convolutional networks via bayesian optimization and structured prediction. In: CVPR. IEEE (2015)
8. Schlegl T, Waldstein S, Vogl WD et al (2015) Predicting semantic descriptions from medical images with convolutional neural networks. In: Ourselin S, Alexander DC, Westin CF, Cardoso M (eds) IPMI. Springer International Publishing, vol 9123. pp 437–448 (2015)
9. Vincent P, Larochelle H, Lajoie I et al (2010) Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 11:3371–3408
10. Bengio Y (2009) Learning Deep Architectures for AI. Mach Learn 2(1):1–127
11. Licandro R, Rota P, Reiter M, Kampel M (2016) Flow Cytometry based automatic MRD assessment in Acute Lymphoblastic Leukaemia: Longitudinal evaluation of time-specific cell population models. In: 14th international workshop on content-based multimedia indexing (CBMI)
12. Ng A (2011) Sparse autoencoder. CS294A Lect Notes 72(2011):1–19
13. Bengio Y, Lamblin P, Popovici D et al (2007) Greedy layer-wise training of deep networks. In: Advances in neural information processing systems, pp 153–160
14. Bishop C (2006) Pattern recognition and machine learning. Information science and statistics, Springer, New York

15. Vincent P, Larochelle H, Lajoie I et al (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Machi Learn Res 11(Dec):3371–3408
16. Bergstra J (2018) Hyperopt—Distributed asynchronous hyperparameter optimization in python. http://hyperopt.github.io/hyperopt/. Accessed 5 Mar 2018
17. Bergstra J, Yamins D, Cox D Making a science of model search. arXiv:1209.5111
18. Bergstra J, Bardenet R, Bengio Y, Kgl B (2011) Algorithms for hyper-parameter optimization. In: Advances in neural information processing systems pp. 2546–2554
19. Han J, Kamber M (2011) Data mining: concepts and techniques, 3rd edn. Elsevier, Burlington, MA
20. Karsa M, Dalla Pozza L, Venn N et al (2013) Improving the identification of high risk precursor B acute lymphoblastic leukemia patients with earlier quantification of minimal residual disease. PLoS ONE 8(10):e76455

Modified Low-Power Built-in Self-test for Image Processing Application

P. Anitha, P. Ramanathan and P. T. Vanathi

Abstract In recent trend, optimization of power without degradation of performance is major concern in application areas like embedded systems digital image and signal processing. The proper selection of test pattern/test image is one of the major issues. Our motivation of this work is to reduce the total power dissipation and area overhead of a Test pattern generator. The proposed BIST uses Negative Edge triggered D-Flip flop (NEDFF) for random pattern generation. When compared to existing LFSR with regular D-FF, our Modified LFSR with NEDFF reduces the count of transistors extensively. BIST using NEDFF is implemented and simulated using Microwind tool with 90 nm technology. The result reveals that significant amount of total power consumption is reduced while testing a VLSI circuit with NEDFF.

Keywords Low-power testing · Less area · NEDFF · Power consumption Switching activity · Medical image processing

1 Introduction

In order to address error tolerance problem in medical image processing, testing techniques has been proposed to determine acceptability of target chips. The biomedical images are usually more complicated for doing error tolerability evaluation based on acquired attribute values. The role of BIST in medical image processing is to

P. Anitha (✉)
Department of ECE, Sri Krishna College of Engineering & Technology, Coimbatore, India
e-mail: anithangptech@gmail.com

P. Ramanathan
Department of ECE, Madanapalle Institute of Technology & Science, Angallu, Andhra Pradesh, India
e-mail: pramanathan2509@gmail.com

P. T. Vanathi
Department of ECE, PSG College of Technology, Coimbatore, India
e-mail: ptvani@yahoo.com

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_20

evaluate the acceptability of target circuits. Generally, more than one million tests are required to obtain good fault coverage. So, testing is very essential to find a faulty and good circuit which produces a quality of product. Design for Testability (DFT) and Built-in Self-test (BIST) are the two popular techniques used. The main problem is design time and effort. Here, we are aimed to identify faults as early in the manufacturing process which reduces, testing time, and delay time and total energy dissipation during the test. The test pattern generation in BIST is usually done using linear feedback shift (LFSR) and Cellular Automata. When the area overhead is concern, LFSR is used because of its simple construction. The structure of cellular automata is quite complex when compared to LFSR.

2 Literature Review

Various types of BIST architecture discussed in this literature review to identify a power efficient design. Kavitha et al. proposed low-power Linear Feedback Shift Register [1]. Bit Swapping technique for dynamic power saving was proposed by Priya [2] offers less dynamic power dissipation. Another unique technique, probability error detection is proposed by Nishat et al., best suited for area and power-efficient testing for digital logic circuits [3]. To achieve very low-power consumption Trupti Patli et al. proposed a logic BIST using low-power linear feedback shift register. To generate low-power test patterns only one flip flop in LFSR is enabled or activated during the shift operation. This lower flip flop activity infers less power loss in BIST module [4]. The BSIT architecture presented in most of the literature do not offer negative edge triggered flip-flop operation. The proposed approach uses the concept of reducing the transitions in the test pattern generated by conventional LFSR. The transition is reduced by increasing the correlation between the successive bits. Balwinder Singh et al. proposed low-power BIST by increasing correlation between successive bits to reduce transition rate of LFSR [5]. Multipliers are the fundamental logic unit of any Computational block. Bharti Mishra et al. have proposed low-power BIST-based Multiplier [6]. Thirunavukkarasu et al. proposed bit swapping LFSR to reduce power consumption in the testing of Universal Asynchronous transmitter and receiver [7].

3 Existing Test Pattern Generator

Test pattern generator is the main contribute to identify a fault in the circuits. With the help of 2:1 MUX and LFSR Test pattern generation generates test patterns. It accepts two primary inputs and produces only one primary output which is given as an input to circuit under test. Circuit under test generates output to test response analysis before that any fault is detected it is just fed back into the BIST control unit

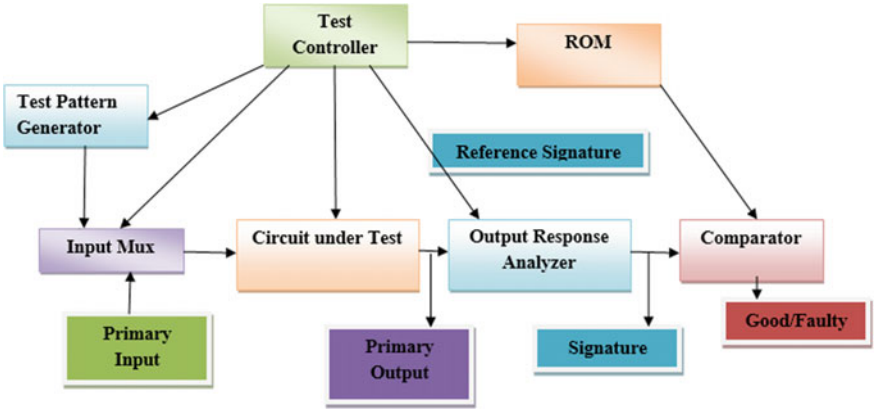


Fig. 1 Internal architecture of BIST

and this unit will act as a feedback circuit until the fault generated is getting clear. Finally, test response analysis produces good-/fault-free output (Fig. 1).

The main advantages of BIST architecture are reduced Testing Time and Test application time is improved. The test controller starts its procedure if the IC is powered up. The status line is made high if there is fault in the circuit. If so, the controller connects the primary inputs to CUT (Circuit under Test) through MUX and makes the operation ready. LFSR so-called Linear Feedback Shift Register. It mainly consists of D flip flop and XOR gates. It acts as a feedback circuit. Input of LFSR is connected to the output of CUT. The peak power is reduced by LFSR by connecting the MUX with LFSR. Switching activity is reduced in LFSR and so power consumption is reduced.

4 Proposed Method

In our modified BIST design, a linear feed back shift register (LFSR) is employed to generate the test pattern. LFSR generates random patterns to identify stuck-at-fault with highest probability. The main drawback of LFSR-based BIST is increased switching activity. It results in massive power dissipation and delay penalty during the test than a normal function. To optimize the total power consumption of the test controller we have implemented negative edge triggered D-FF. In pattern generator stage, this D-FF offers power efficient operation with reduced area. More over clock input is scaled down. So, there is no necessity of Clock input to drive the NEDFF. The operating frequency of NEDFF ranges from 243 to 555 MHz (Figs. 2, 3).

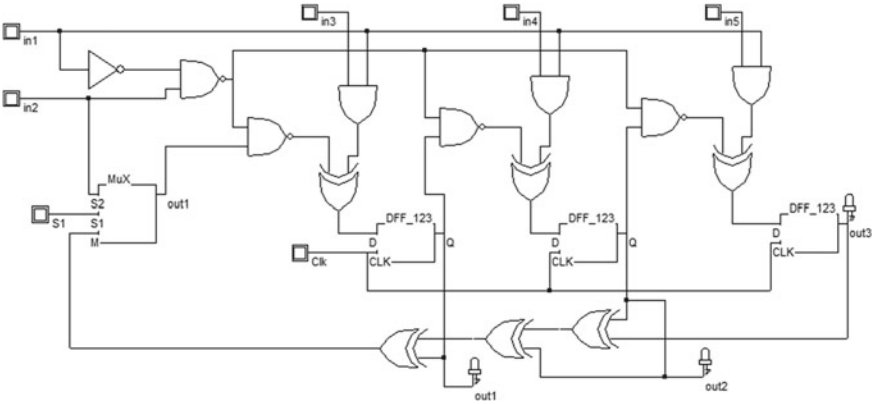


Fig. 2 Implementation of built-in self-test (test controller)

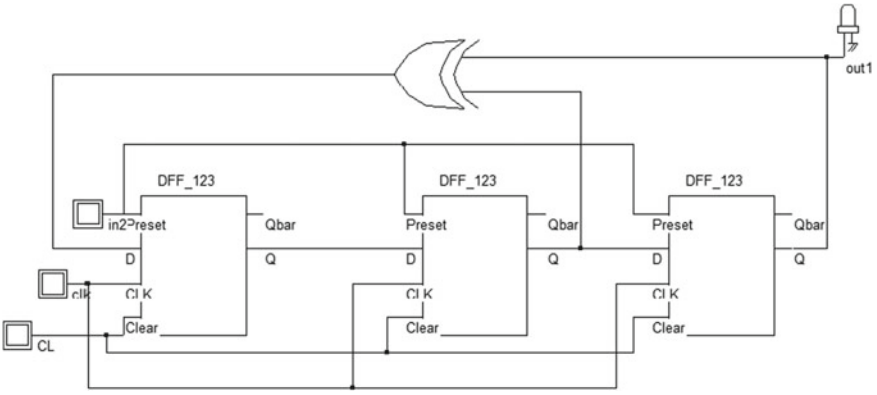


Fig. 3 Test pattern generator using negative edge triggered D-FF

5 Application of BIST in Image Processing

In the era of image process, addition of error degrades the standard of a picture and affects the frequency feature of a picture. The frequency variation is calculable to evaluate acceptability of errors within the image. BIST plays major role to grading the image. The value of RGB color model has ranges from 0 to 255. Each pixel has anyone of 256 colors. 16 bit of 256 different values of RGB will be stored in the memory. The error-free pictures and their threshold values are going to be kept in system on chip of BIST controller. The pixel value of an image has been taken as test data. Applied test data will be compared with the threshold values stored in BIST controller to check the correctness of an image. Suppose, the image has unacceptable faults, then engineer should define corresponding acceptability attribute to that particular fault. If the test data is massive, that information's are compressed then keep in

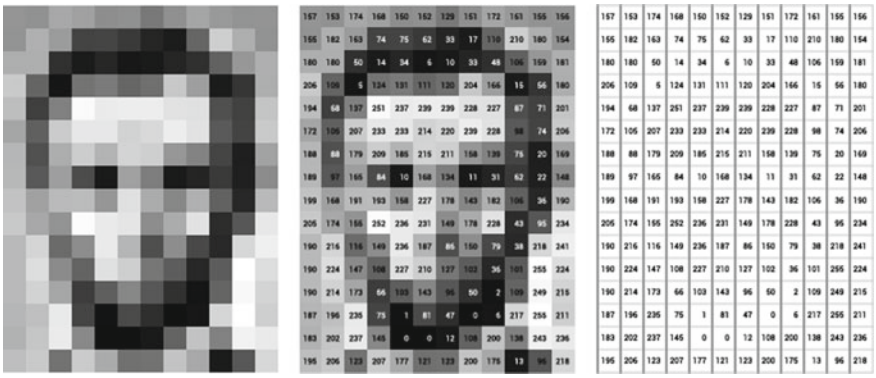


Fig. 4 Extracting pixel information from image

system on chip. During testing the compressed data decompressed then applied to circuit under test (Fig. 4).

6 Result and Discussion

The proposed NEDFF is implemented and simulated using Microwind tool with 90 nm technology. Power Results are tabulated in Table 1 and average dynamic power is tabulated in Table 2. The result reveals that 32.1% of total power consumption and 63.9% of dynamic power is reduced compared to regular BIST Architecture (Figs. 5, 6 and Table 3).

Table 1 Comparison of total power consumption of regular BIST with proposed BIST

Type of the BIST architecture	Total power consumed (mW)
BIST with regular D-FF	1.698
BIST with negative edge triggered D-FF	0.546
Dynamic power of BIST with regular D-FF	4.687
Dynamic power of BIST with NEDFF	2.999
Total power optimized	1.688

Table 2 Dynamic power dissipation of random input

Random input pattern applied	Total power dissipation (mW)	Total power dissipation (mW)
	Regular BIST	BIST with NEDFF
0000 0001	3.026	1.292
0000 1110	2.062	1.913
0001 1111	5.610	3.415
0000 0100	4.287	1.494
0000 0101	4.509	1.494
0000 0011	2.689	1.494
1111 1111	9.897	7.321
1111 0111	10.29	8.885
1111 0011	2.933	2.255
0011 1100	2.880	1.042
1000 0011	7.689	6.341
1101 1010	2.995	2.681
1010 0101	3.011	2.056
0101 1010	4.897	3.014
1001 0101	3.526	1.625

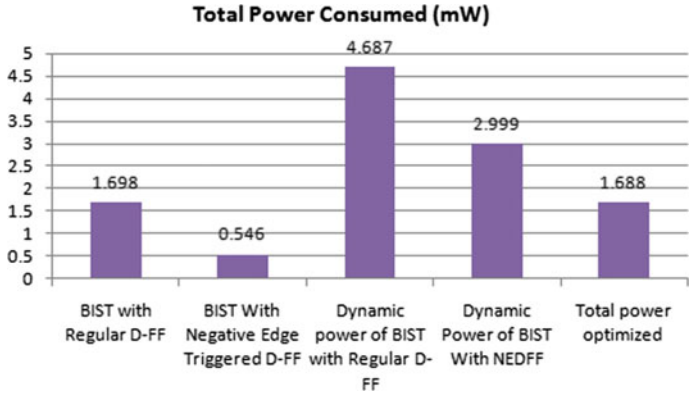


Fig. 5 Performance analysis of proposed BIST using NEDFF

7 Conclusion

The proposed BIST consumes very less power with reduced clock power and the significant of dynamic power is reduced. Hence the design is well suited for low-power testing in image processing circuits.

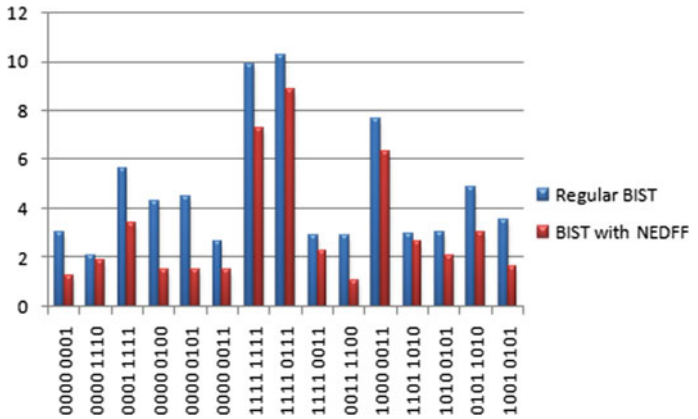


Fig. 6 Comparison of dynamic power dissipation

Table 3 Comparison of regular LFSR with modified LFSR

	Total power reduction (%)	Dynamic power dissipation (%)	Simulation tool used	Circuit complexity
Test patent generator with NEDFF	32.1	63.9	Microwind tool with 90 nm technology	Less
TPG with modified LFSR (R-injection)	5.13	41.7	Xilinx 9.1	High

Acknowledgements We would like to thank the management and principal of Sri Krishna college of Engineering and Technology, Coimbatore for providing the necessary facilities and support.

References

1. Kavitha A, Seetharaman G, Prabakar TN, Shrinithi S (2012) Design of low power TPG using LP-LFSR. In: 2012, IEEE International conference intelligent system modelling and simulation
2. Priya SHKP (2013) Test pattern generator (TPG) for low power logic built in self test (BIST). Int J Adv Res Electr Electron Instrum Eng 2(4)
3. Pandharpurkar NG, Ravi V (2015) Design of BIST using self-checking circuits for multipliers. Indian J Sci Technol 8(19)
4. Patil T, Dhankar A (2016) A review on power optimized TPG using LP-LFSR for low power BIST. In: IEEE sponsored world conference on futuristic trends in research and innovation for social welfare
5. Singh B, Khosla A, Bindra S (2009) Power optimization of linear feedback shift register (LFSR) for low power BIST. In: International advance computing conference, IACC 2009

6. Mishra B, Jain R, Saraswat R (2016) Low power BIST based multiplier design and simulation using FPGA. In: Conference on electrical, electronics and computer science (SCEECS), 2016
7. Thirunavukkarasu V, Saravanan R, Saminadan V (2016) Performance of low power BIST architecture for UART. In: International conference on communication and signal processing (ICCSP)

A Hassle-Free Shopping Experience for the Visually Impaired: An Assistive Technology Application

Sherin Tresa Paul and Kumudha Raimond

Abstract The issues and challenges faced by the Visually Impaired (VI) bring forth an insurmountable inequality of living conditions which needs immediate attention. The objective of this work is to render an in-depth understanding of some of the issues faced by the VI and to propose an Assistive Technology (AT) framework for assisting the VI to have a smooth and carefree shopping experience, including navigation in a supermarket to the sections specified by the VI user, identification of objects of interest, hassle-free billing, and registration for a subscription-based offer notification system based on the user's shopping habits and preferences. The ultimate goal is to improve the Quality of Life (QoL) led by the VI person in as many ways as possible.

Keywords Assistive technology for Visually Impaired (VI) · Blindness
Quality of Life (QoL) · Image processing · Pattern recognition · IoT
Optical head-mounted displays · BLE beacons
Inertial Measurement Unit (IMU)

1 Introduction

The difficulties faced by the Virtually Impaired (VI) are unimaginable even in the highly innovative and technologically enabled world of gadgets. The simplest task of picking up a fruit of their liking to complex tasks like getting around in an unknown place are still viewed by the VI community with apprehension. The 2010 estimate of VI with data collected from six WHO regions [26] shed light on overwhelming facts and numbers, wherein 285 million people all over the world are VI, in which

S. T. Paul (✉) · K. Raimond
Karunya University, Coimbatore, India
e-mail: sherintresapaul777@gmail.com

K. Raimond
e-mail: kraimond@karunya.edu

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_21

39 million with full visual impairment, 246 million having low vision, and 81% of the people under the clinical investigation were above 50 years.

The objective of this work is to render an in-depth understanding of some of the issues faced by the VI and to assist them to have a smooth and carefree shopping experience which include navigation in a supermarket to the sections specified by the VI user, identification of objects of interest, hassle-free billing, and registration for a subscription-based offer notification system which will give notifications of newly arrived and discounted items based on the user's shopping habits and preferences. The ultimate goal is to improve the QoL led by the VI person in as many ways as possible.

1.1 Challenges Faced by VI in a Supermarket

The following lists some of the challenges faced by a VI person in a typical shopping scenario:

- The usual unreliable method used for object identification by touch.
- Products are available in various quality, price, and quantities which makes the shopping experience a tedious one for the VI.
- Unexpected obstacles and the crowd in the supermarket make navigation even with cane, a difficult task.
- The chance of accidental collision is more in a crowded place like a supermarket.

1.2 AT for VI

The field of AT has been growing at an amazing pace and scale owing to the researches all around the world trying to make the lives of the disabled and/or elderlies easier. AT for Blinds (ATB) is one such intriguing field of research where the Daily Life Activities (DLA) and Outdoor Navigation and Way Finding (ONWF) through technological assistance are made happen with sensor-activated canes, headgear with an HD camera, image processing, voice-activated controls, etc. The goal of this work is to increase the QoL of the VI through a navigation system empowered by image processing, IoT, machine learning, and other AT systems.

2 Related Works

The VI community has benefitted considerably from the AT, but most of the innovations include highly expensive electronic parts which a VI person with limited economic conditions may not be able to afford. This section gives a brief layout of

the existing literature on AT-enabled navigation systems and applications. Ebsar [1] is an indoor navigation system which incorporates Arabic language developed for the people who have Arabic language as their native language. It uses QR code technology to locate and calibrate the user's location. The system works in two phases, one being the map-making phase in which the building's indoor pathways are traversed by a person with normal sight and Point of Interest's (POI's) are marked by QR codes and then the same is used for indoor navigation for the VI in the phase two. It uses Google Glass [19] as an I/O device for the VI user. It is an Optical Head-Mounted Display (OHMD) device developed by Google and shows promising results for the VI as well as sighted people. However, the QR code dependency of the program is tedious and need to be dealt with. Making the program platform independent and adding certain capabilities like Arabic OCR, etc, are some of the future directions planned.

The basic concept behind the working of an edge detection algorithm like Canny [4] is the difference in the contrast between the neighboring pixels' luminance values. The algorithm detects and marks an edge when the said contrast difference is above a particular threshold value which means the system is highly dependent on the threshold value. A high threshold gives an incomplete set of detected edges and a low threshold gives far too many falsely detected edges. As a countermeasure, an improved edge detection algorithm for indoor navigation of the VI was developed [5] employing a Modified Sigmoid Function (MSF) [3] and an Inertial Measurement Unit (IMU) [2]. Four scenarios involving two use cases namely, white cane and guide dog were evaluated for the MSF-based edge detection system and some of the standard edge detection systems. The MSF-based system showed significant improvement in case of largely blurred frames and showed improvements in terms of computational complexity and high performance with lesser error values, and hence claims to be of use for real-time applications. Though the results are promising compared to several existing systems, the real-time implementation has many challenges like reducing the false positivity, including an OHMD device, sensors, etc.

With the emergent disruptive technology IoT, Bluetooth Low Energy (BLE) beacons [30] are used in many indoor navigation systems [29]. Beacons are low-energy devices which are basically one-way transmitters of data packets. Each data packet has basically three types of frames, namely, URL frame which announces the URL information, UID frame with the beacon's identity for the purpose of identification of the beacon, and the TLM frame which contains information like beacon's battery level, temperature, etc. The open initiative by Google enabled with BLE beacons gives way to a new arena called physical web [18] where each smartphone device can be made as a beacon to broadcast a particular URL for business or entertainment purposes using beacon toy [14]. Some of the commonly available beacon manufacturers [17] which comply with Google's Eddystone Protocol [15] are Accent Systems, Beaconinside, Blesh, BlueCats, BlueUp, Bluvision, Estimote, Gimbal, Kontakt, Minewtech, Nordic Semiconductor, POI, Radius Networks, Reco, Ruuvi Open-Source Innovations, Signal360, Glimworm, etc. Google's open-source Eddystone API [16] is available on GitHub for download. The smartphone can act as

Table 1 BLE beacon performance evaluation with and without Kalman filters for Estimote, Kontakt, and Glimworm BLE beacons in terms of standard deviation (dBm) [23]

	No filter	Smoothed	V = 2	V = 2.5	V = 3
Estimote	6.54	7.18	4.22	5.22	4.47
Kontakt	5.79	7.09	3.76	4.25	6.46
Glimworm	5.65	5.46	4.91	3.41	5.17

a simple listening device for proximally active beacons or as an interface for remotely controlling IoT-enabled devices.

An evaluation [23] of BLE beacon performance was done with Kalman filters for smart homes and buildings. The evaluation was done for three beacons, namely, Estimote [9], Kontakt [21], and Glimworm [13]. The research sheds light on the issue of possible interference by ambient noise in the vicinity by mobile devices, other beacons, etc. and proposes an improvement by employing Kalman filter and Pedestrian Dead Reckoning (PDR) algorithm. The comparison is done without filtering, with smoothing and with filtering. The results based on the said comparison are shown in Table 1.

A new algorithm named Dead Reckoning algorithm based on Bluetooth and Multiple sensors (DRBM) was proposed [33] which focused on the precision enhancement of PDR algorithm by utilizing bluetooth, multiple sensors as well as Kalman filters. The research was an eye-opener regarding the issue of situation dependency of the current IoT scenarios and tried to bring more position accuracy and adaptability in IoT implementation for indoor navigation. A Real-Time Location System (RTLS) with a wireless charging technique was introduced [31] for mitigating the issues caused by RFID-based Indoor Positioning System (IPS) in a healthcare center use case. The system used the concept of trilateration for generating the position coordinates of the mobile unit. Even with the 8-bit microcontrollers, the system showed highly efficient positioning and minimal error. Lack of situation awareness is one of the many issues faced by the VI which keep them to the indoors. The VI has to rely on another sighted person for getting around unknown places, for shopping, etc. To make the VI people navigate independently, a novel method comprising a wearable unit of camera, haptic feedback belt, and an embedded electronic processor was developed [32]. The system was equipped with a braille display system as well which contributed to the navigation independency for the VI. Furthermore, the system was built on the concept of Stixel World [11] which makes the system capable to find "Free Space" between obstacles. for better and precise navigation. For maximum avoidance of accidental collision and to detect objects like an unoccupied chair, the system used free space parsing and object recognition using depth algorithms. The method has an efficient three-step planning phase which comprised of detection of a collision-free path, planning a safe trajectory of collision-free path for the VI to follow, detection of certain objects and moving towards it. The testing phase evaluated the efficiency of the system to assist tasks such as obstacle detection, navigation through a maze, locating an unoccupied chair, moving in a real environment (MIT

Lab) with and without cane, and/or system assistance. The results showed that the system performs well beyond the cane-based navigation with reduced number of accidental collisions but it needs to be tested and calibrated for a real-world scenario like navigation through a supermarket.

An innovative concept [12] was put forth to monitor children in schools by installing RFID tags and by badge calling. The system helps in real-time monitoring of each of the child without being actually present in the child's vicinity. If in case of a medical or any other emergency, a silent alarm in the form of a text message can be sent out to the staff, authorities, and/or a medical team. With some major adjustments and modifications, this system can be used to monitor the VI children, as they usually find a delay in detecting a danger to themselves or anyone around him/her. Although the concept was practical and relevant to the mobility issues and real-time monitoring of minors, the system should be modified, implemented, and tested for different real-world use cases. An IMU-based open-source system [25] for navigation was proposed. The embedded inertial measurement based navigation unit is called OpenShoe. The hardware and software resources are open source and are available for download [20]. The method is fairly simple but the efficiency is compromised in terms of the occurrence of errors for maintaining the simplicity of the system.

An elaborate review of literature is done focusing on the recent advancements in AT paradigm to increase the QoL of the VI community. It is evident from the literature that there are quite some attempts in the AT field to ease the pain of VI community in getting through the daily activities but the obvious lack of a structured use-case-based framework for the proper installation and administration of the said AT advancements leads to a reduced QoL for a VI person. An AT framework is proposed which utilizes the disruptive technology IoT, an OHMD inspired by Google Glass [19] equipped with cutting-edge image processing and classification algorithms, QR Codes, BLE Beacons [17], Eddystone API [16], etc. The motivation of the proposed AT framework is to assist the VI community in achieving an equality with the sighted community in terms of QoL. The objective of the proposed framework is to assist the VI to have smooth and carefree shopping experience which include navigation in a supermarket to the sections specified by the VI user, identification of objects of interest, hassle-free billing, and registration for a subscription-based offer notification system based on the user's shopping habits or preferences.

3 Proposed AT Framework for Hassle-Free Shopping

The proposed AT framework for hassle-free shopping is shown in Fig. 1 is divided into three Use Cases (UC) which are collision-free navigation through the aisles with the help of an intelligent OHMD device with voice input and output, identification of the specified item by image processing, pattern recognition, classification algorithms, easy billing and subscription for the offer notification system based on the user's shopping preferences.

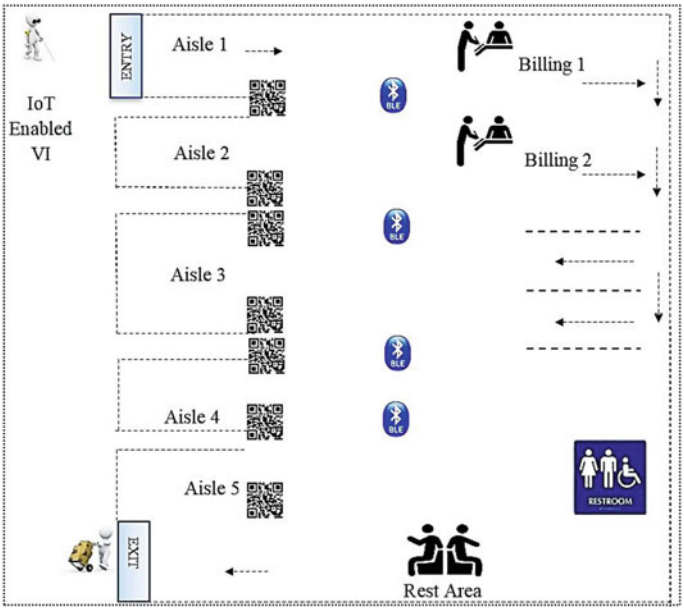


Fig. 1 Proposed AT framework

The OHMD

The OHMD device is inspired by Google Glass [19] which works by voice commands and also by the pressure sensors embedded in the IoT-enabled glove. It is to be noted that an updated floor map of each AT-enabled supermarket is uploaded in the cloud with a unique store ID which will be on the BLE beacon at the entrance so that any user can make use of it for finding the particular store. The voice command is registered as an actionable one using the keyword "Go To". The system will be designed in such a way that whatever comes after the keyword "Go To" is taken as the destination, for example, "Go To" fruits. As soon as the VI user gives this command, the OHMD searches for the particular store's floor map which is updated periodically in the cloud. The floor map with its QR-coded aisles helps the VI user to find the particular section.

3.1 Use Case 1: Collision-Free Navigation

Moving through a crowd of people and another assortment of shelves may come naturally for a person with eyesight while a VI person usually finds it an almost impossible task at hand and may as well choose to not get into such a situation. The Use Case 1 portrays the collision-free navigation scenario in a supermarket. The VI user is equipped with an OHMD device which can be operated with voice input. The

Wi-Fi-enabled OHMD device is inspired by Google Glass [19] which has an HD camera operated by voice input, BLE listener, an earphone for giving voice outputs for the VI user, and a processor unit. A haptic feedback unit is incorporated into the user's shoe sole. A representative figure from an IMU-based open-source navigation system called OpenShoe [17] is shown in Fig. 2.

The supermarket is assumed to be equipped with Estimote BLE beacons with proximity sensors at every junction between aisles. The beacons can be of the following variety according to the budget at hand. A fact table of available Estimote beacons in the market is given in Table 2.

All the aisles in the supermarket are fitted with QR codes of the product information, which is updated as and when a new stock arrives. The navigation is further enhanced by the assistance of QR codes for each aisle. As soon as the OHMD scans the QR code, the information is given as a voice command to the user through the OHMD headgear. It is a prerequisite for the supermarket to be set up with a QR-code-based itemized floor map. The map is built and updated frequently by the supermarket which requires three major steps.

Fig. 2 IMU-based system for navigation unit (fitted in both shoes) [17]

Table 2 Estimote beacons fact table [10]

	Location UWB beacon	Location beacon	Proximity beacon	Sticker beacon	Video beacon
Battery	5 years	5 years	2 years	1 years	USB
Range (m)	200	200	70	7	10
Price	$159	$99	$59	$99	–
iBeacon [24] or Eddystone [15] packets	8 at a time	8 at a time	1	1	2 at a time
Sensors	Motion, light, temperature, pressure	Motion, light, temperature, magnetometer, pressure	Motion, temperature	Motion, temperature	–

- **Step 1**: Floor exploration by a person with normal sight and noting Point of Interest (POI) like fresh fruit aisle entry, vegetable aisle entry, discounted fresh food products, household items aisle, etc., which is then generated as QR codes.
- **Step 2**: Back traversal of explored path, placing the QR codes for the corresponding POI identified in step 1, map construction with defined pathways and generalized product information.
- **Step 3**: Testing the QR-code-based map as a guide for the VI user by the normally sighted and unsighted users for further calibration.

3.2 Use Case 2: Identification of Item

Identifying an item in the supermarket without eyesight is only possible through the perception of touch, hence unreliable and time taking. In this use case, the AT plays a quantifiable part by taking a picture of the item of interest and by executing an object recognition algorithm. A block diagram depicting the navigation to the specified section in the supermarket assisted by the OHMD, for example, fruits section and item recognition algorithm is given in Fig. 3.

In [22] a visual–tactile fusion method utilizing joint sparse coding is proposed for object recognition tasks where the visual access might be limited. The method fuses the visual and tactile feedbacks where the complete coupling of the tactile and visual feedbacks is not necessary. The tactile data is collected from each fingertip through an expert hand unit [8] and then incorporated with the visual data effectively

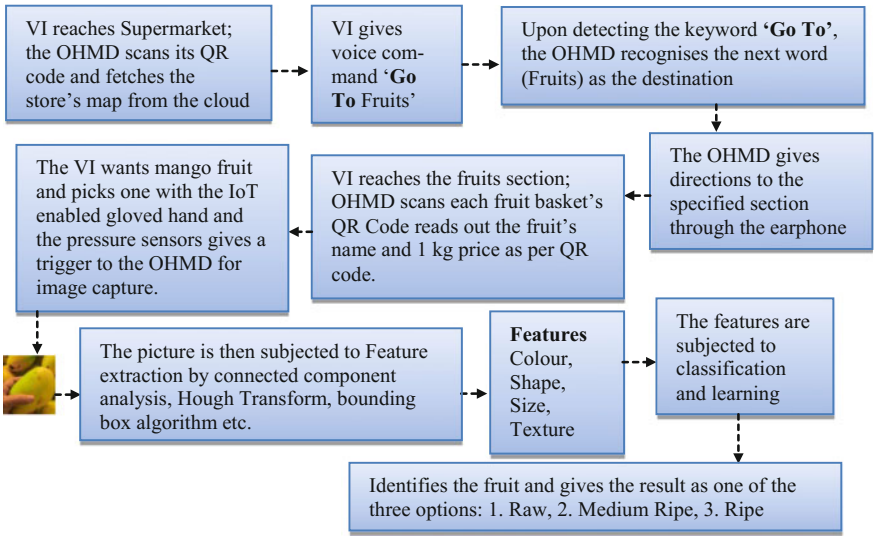


Fig. 3 OHMD's role in navigation and identification of object of interest

resulting in an efficient object recognition system. The IoT-enabled hand glove in the proposed framework is a tactile feedback unit that can be incorporated along with the OHMD through which the identification of the item in the supermarket would be more accurate and faster. A 3D deep net architecture called PointNet [27] utilizing deep learning and Point Cloud Library (PCL) [28] depicts an entirely different perspective for the object recognition with clear-cut and improved experimental results. DeepLab [7] is a cutting-edge technology for upscaled image classification and was proved to be effective in many critical datasets using atrous convolution [6].

3.3 Use Case 3: Easy Billing and Registration for Offer Notification

The billing process is one of the most tedious tasks for even a sighted person, and therefore in a VI person's perspective it is more troubling to be dealt without any assistance. Use Case 3 is depicted in Fig. 4. When the VI person gives voice command "Go To" billing, the OHMD device guides only to the active (BLE beacons still transmitting) queue with less than five persons. As the billing and user registration is done, a record for the user is created with user id Uxxx and at the end of the day the record is updated to cloud. As and when a new stock arrives the item is checked with previous purchases of the registered users and an Item Relevance Factor (IRF) is calculated. A Relevance Factor Threshold (RFT) is set so that whenever the newly arrived item's IRF exceeds the RFT, a notification is sent to the user's mobile.

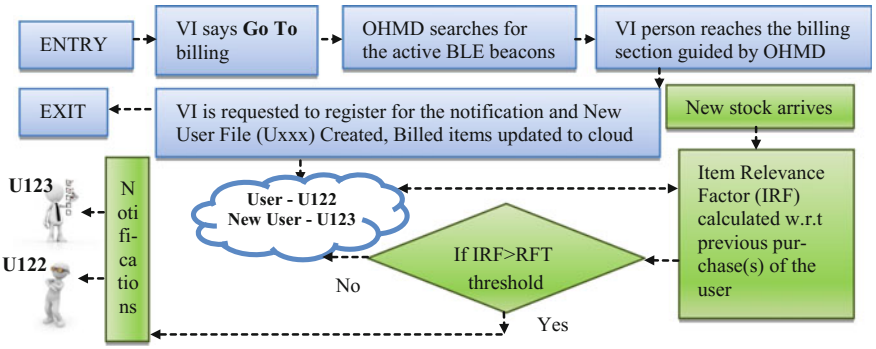


Fig. 4 Billing, offer registration, and notification to the registered user's mobile

4 Conclusion and Future Work

It is literally impossible to understand the difficulties experienced by the VI in this fast-track world. Unfortunately, the technological advancement happening all around us is not reflected in the VI person's lifestyle. An extensive review of literature is done focusing on the recent advancements in AT paradigm to increase the QoL of the VI community. Even if there are quantifiable research and innovation happening in the field of AT which can be used to increase the QoL of the VI, accommodating those AT innovations into the daily lives of VI is still in the budding stage. A lack of proper framework to employ and administer the AT for the VI is recognized and an AT framework for hassle-free shopping is proposed with three use cases namely, collision-free navigation, identification of item, easy billing and registration for offer notification.

The future work entails the development of a prototype, testing it with VI as well as sighted users, and analyzing the results. The algorithm to calculate item relevance factor for the offer notification system will be developed and tested in parallel to the prototype development. The proposed AT framework can also be used for elderlies who have low vision due to cataract and/or other impairments.

References

1. Al-khalifa S, Al-razgan M (2016) Ebsar: indoor guidance for the visually impaired. Comput Electr Eng 54:26–39
2. Ave I (1987) United States Patent (19)
3. Braun GJ, Fairchild MD (1999) Functions. 2:380–393
4. Canny J (1986) A computational approach to edge detection 6
5. Chan KY et al (2016) PT US CR. Expert Syst, Appl
6. Chen LC et al (2017) Rethinking atrous convolution for semantic image segmentation
7. Chen LC et al (2018) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848
8. Chowdhury T et al (2013) Vegetables detection from the glossary shop for the blind 8(3), 43–53
9. Estimote Inc. Estimote. https://estimote.com/
10. Estimote Inc. Estimote products. https://estimote.com/products/
11. Franke U, Pfeiffer D (2009) The stixel world—a compact medium level representation of the 3D-world, pp 51–52
12. Gaba P, Chugh G (2018) Child Saf Syst RTLS 5(1):147–151
13. Glimworm Beacons Glimworm. https://glimwormbeacons.com/
14. Google Beacon Toy. https://play.google.com/store/apps/details?id=com.uriio
15. Google Eddystone. https://developers.google.com/beacons/eddystone#full_support_for_eddystone
16. Google Eddystone API. https://developers.google.com/beacons/overview
17. Google Eddystone supported beacon manufacturers. https://developers.google.com/beacons/eddystone#beacon_manufacturers
18. Google Physical Web. https://google.github.io/physical-web/
19. Google Inc. Google glass. https://developers.google.com/glass/
20. John-Ol of Nilsson, Isaac Skog, Peter Handel1, and K.V.S.H. Openshoe. www.openshoe.org

21. kontakt.io Kontakt. https://kontakt.io/
22. Liu H et al (2016) Visual—tactile fusion for object recognition. IEEE Trans. Autom. Sci. Eng. 1–13
23. Mackey A, Spachos P (2017) Performance evaluation of beacons for indoor localization in smart buildings, pp 823–827
24. Ng TM (2016) From Where I am to Here I am. Accuracy study on location-based services with iBeacon technology, 3733
25. Nilsson J et al (2012) Foot-mounted INS for everybody—an open-source embedded implementation, pp 140–145
26. Pascolini D, Mariotti SP (2012) Global estimates of visual impairment (2010), 614–619
27. Qi CR et al (2016) PointNet: deep learning on point sets for 3D classification and segmentation big data + deep representation learning. Cvpr, 652–660
28. Rusu RB, Cousins S (2011) 3D is here: point cloud library. In: IEEE international conference on robot automation, pp 1–4
29. Shailesh P et al (2017) Indoor navigation system with augmented reality using eddystone beacons, 10–12
30. Singh A (2017) Bluetooth low energy: need of the hour 3, 749–751
31. Urrea WA, Ariza HM (2017) Real time location system (RTLS) focused on the optimization of efficiency for hospital center assistance 12(15), 5248–5253
32. Wang H et al (2017) Enabling independent navigation for visually impaired people through a wearable vision-based feedback system, pp 6533–6540
33. Yu N et al (2017) A precise dead reckoning algorithm based on bluetooth and multiple sensors. IEEE Internet Things J (Accepted Publ. 4662)

Retina as a Biomarker of Stroke

R. S. Jeena, A. Sukeshkumar and K. Mahadevan

Abstract Stroke is one of the significant reasons of adult impairment in most of the developing nations worldwide. Various imaging modalities are used to diagnose stroke during its initial hours of occurrence. But early prediction of stroke is still a challenge in the field of biomedical research. Since retinal arterioles share similar anatomical, physiological, and embryological attributes with brain arterioles, analysis of retinal fundus images can be of great significance in stroke prognosis. This research work mainly analyzes the variations in retinal vasculature in predicting the risk of stroke. Fractal dimension, branching coefficients and angle, asymmetry factor and optimality ratio for both arteries and veins were computed from the processed input image and given to a support vector machine classifier which gives promising results.

Keywords Stroke · Retinal fundus images · Support vector machine

1 Introduction

Stroke [1] is a cerebrovascular ailment affecting the normal blood flow to brain. It is a heterogeneous disorder which has two main pathophysiological divisions: hemorrhagic and ischemic stroke, and ischemic stroke is further classified into lacunar and cortical. There is likelihood for the co-occurrence of both ischemic and hemorrhagic strokes. When a clot is formed in the artery supplying blood to cerebrum, Ischemic stroke results and a burst of a blood vessel results in Hemorrhagic stroke. About 87%

R. S. Jeena (✉) · A. Sukeshkumar
College of Engineering Trivandrum, Trivandrum, Kerala, India
e-mail: jeena_rs@yahoo.com

A. Sukeshkumar
e-mail: drsukeshkumar@yahoo.in

K. Mahadevan
Department of Ophthalmology, Sree Gokulam Medical College & Research Foundation, Trivandrum, Kerala, India

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_22

of all strokes belong to ischemic category with 10% being intracerebral hemorrhage (ICH) and 3% being subarachnoid hemorrhage (SAH) strokes [2]. Clearly, stroke has transformed from a disease pertaining to developed nations to a global risk.

The standard imaging modality used in the initial identification of stroke is a non-contrast computed tomography (CT) scan [3]. Even though CT remains the most accepted brain imaging tool, the crucial sign of ischemic stroke is not perceivable in first few hours after stroke onset. Hence the detection of ischemic stroke is much more complicated. Innovations in MRI have enhanced the potential for perception of minute lesions in brain images [4].

Early analysis of stroke is still an important issue that we are facing in the area of biomedical research. It is very crucial for proper treatment and care. Investigations show that diverse risk factors hold vital information about the occurrence of Stroke. Biomarkers also play a significant role in the prognosis of various ailments. Retina can be considered as a biomarker, as any change in the cerebral blood vessel will be reflected in the retinal blood vessel. This work focuses on retinal imaging for the prediction of stroke.

Retinal fundus imaging is noninvasive, fast and high-resolution technique which produces a two-dimensional representation of the three-dimensional structure of eye [5]. The optic nerve, a cable-like gathering of nerve filaments transmits visual data from the eye to the cerebrum. Any harm of the optic nerve will disjoin the exact transmission of visual data between the retina and brain, leading to vision loss. Since retina is a structural outgrowth of the brain, the retina may be susceptible to the same inflammatory injury that causes neurovascular disorder in the cerebral nervous system. This is clearly depicted in Fig. 1.

Fig. 1 Link between eye and brain (*Courtesy* The Discovery Eye Foundation)

2 Literature Survey

Cardiovascular ailments show itself in the retina in different ways. Hypertension and atherosclerosis cause changes in the ratio between the width of retinal arteries and veins [6]. Tapering of the arteries and broadening of the veins is associated with an enhanced risk of stroke. Research shows that microaneurysms or arteriolar narrowing found in the retina may present hazard factors for cerebrovascular ailments like stroke [7]. Morphological variations in vessel contour [8], branching model, thickness of blood vessels, tortuosity [9], presence of retinal lesions, branching angle and coefficient [10] and fractal dimension are some of the variations in vascular pattern of retina associated with cardiovascular diseases like stroke.

Vision loss is one of the major outcomes of stroke [11, 12]. De Silva et al., proved that those with arteriolar thinning were more prone to neurovascular diseases like stroke [13]. McGeechan et al. showed that broader venule builds the hazard of stroke in human beings whereas the variation of retinal arterioles was not linked with stroke [14]. Research shows that ischemic stroke victims have sparser microvascular retinal vasculature [15]. Neural network based classification has been implemented by the same authors after extracting features from retinal fundus images [16]. This research paper is an extension of their previous work. Research strongly suggests that retinal imaging is a powerful tool which can be used in early prediction of neurovascular ailments like stroke.

3 Methodology

Retinal images of control group and stroke subjects have been collected from Sree Gokulam Medical Centre and Research Foundation, Kerala (control n = 80, Stroke n = 50). The selection of the region of interest to measure the vessel width and angle is a significant phase in retinal image analysis. Centre of the optic disc is detected manually and are divided into three zones. Zone A corresponds to the region between optic disc boundary to 0.5 optic disc diameters (ODD), Zone B represents 0.5–1.0 ODD and Zone C corresponds to 0.5–2.0 ODD from the optic disc boundary.

Cerebral vascular variations are the significant pathological signs of stroke. Research show that cerebral variations are reflected in retinal vasculature and can be analyzed by computing the branching angle, fractal dimension, branching coefficient, and optimality ratio. The features computed from the retinal images are briefly explained below.

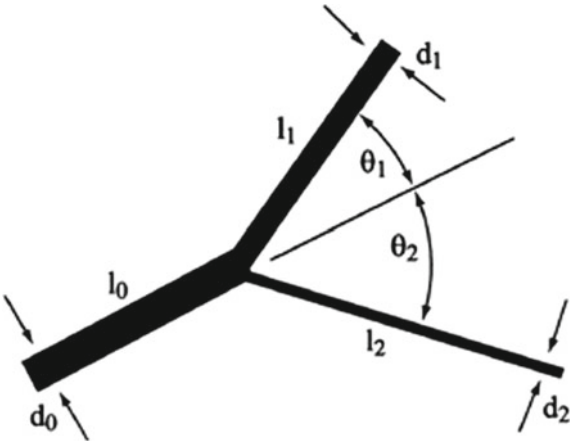


Fig. 2 Retinal bifurcation angle

3.1 Features

Fractal dimension (FD)
Fractal dimension [17, 18] is computed from the skeletonized blood vessel using the Box-Counting method. A pattern is enclosed by boxes of different size (s) and the number of boxes required to envelope the complete pattern is represented as N_s. Fractal dimension is estimated from slope of best fitting straight line obtained by plot of log (N_s) versus log (s). It is given by Eq. (1)

$$\mathbf{FD} = \frac{\mathbf{log(Ns)}}{\mathbf{log(s)}} \tag{1}$$

Fractal dimension will be higher for the retinal fundus images of stroke patients which indicates higher degree of structural complexity.

Branching angle (BA)
Branching angle is the angle between two child blood vessels at the bifurcation which is depicted in Fig. 2. It is computed with the aid of Vampire annotation tool. In this work, the mean value of the branching angles of arteries and veins in Zone C are calculated.

$$\boldsymbol{BA} = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2 \tag{2}$$

Branching coefficient (BC)
The branching coefficient at each bifurcation [19, 20] is defined as

$$\boldsymbol{BC} = \frac{\boldsymbol{d_1^2} + \boldsymbol{d_2^2}}{\boldsymbol{do^2}}, \tag{3}$$

where d_1 and d_2 denote the vessel thickness of each daughter vessel and d_o, thickness of the parent vessel. Thickness of the blood vessels in Zone C has been analyzed in our study using Vampire annotation tool. The mean values of both arteries and veins are calculated separately.

Asymmetry factor (**AF**)
The AF [21] for $d_1 \gg d_2$ is given by Eq. (4). The ratio of the squares of width of the daughter vessels is computed.

$$AF = \frac{d_1^2}{d_2^2} \tag{4}$$

Optimality ratio (Γ)
Optimality ratio or the junction exponent [10, 21] is defined as

$$\Gamma = \sqrt[1/3]{\frac{d_1^3 + d_2^3}{2d_O^3}} \tag{5}$$

The mean value of optimality ratio of blood vessels, both arteries and veins are calculated separately.

3.2 SVM Classifier

SVM [22] is a supervised classifier that performs classification by creating hyperplanes in a multidimensional space. They are based on the idea of decision planes that define decision margins. It can deal with multiple continuous and categorical variables. It merges linear algorithms with various kernel functions (linear, polynomial, quadratic) that make it a significant tool in machine learning and medical imaging applications.

4 Results and Discussion

All the analysis and algorithms were implemented using MATLAB. Each retinal image was converted to gray scale and pre-processed using adaptive histogram equalization. The pre-processed image is shown in Fig. 3a. Image is binarized by thresholding and required morphological operations are done to remove the background. To reduce all objects in the vascular map to lines, skeletonization is done. This leads to the extraction of vessel tree structure, as presented in Fig. 3b. After skeletonization, the box-counting fractal dimension is computed using Eq. (1) in Zone C for both stroke affected and healthy samples.

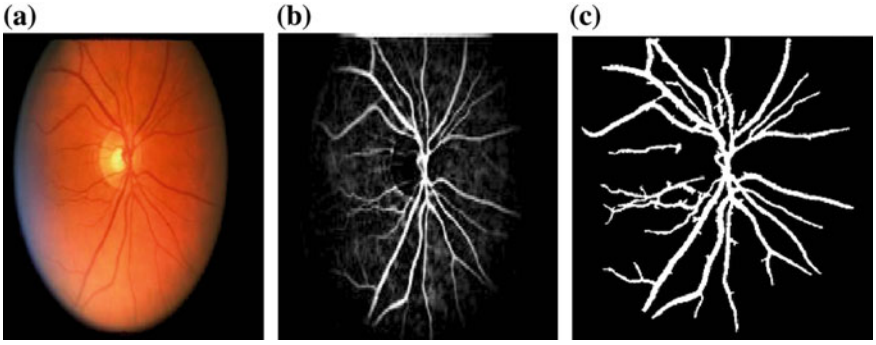


Fig. 3 **a** Input image **b** pre-processed image **c** binarized image

Table 1 Summary of retinal features between healthy and stroke cases

Sl no	Features	Healthy (n = 80) Mean ± SD	Stroke (n = 50) Mean ± SD
1	Fractal dimension (FD)	1.31 ± 0.01	1.73 ± 0.22
2	Branching angle artery (BA_a)	70.4 ± 0.32	71.4 ± 0.14
3	Branching angle vein (BA_v)	74.1 ± 0.03	74.5 ± 0.32
4	Branching coefficient artery (BC_a)	1.46 ± 0.26	1.7 ± 0.39
5	Branching coefficient vein (BC_v)	1.2 ± 0.41	1.5 ± 0.02
6	Asymmetry factor artery (AF_a)	0.80 ± 0.3	0.81 ± 0.02
7	Asymmetry factor vein (AF_v)	0.75 ± 0.04	0.78 ± 0.1
8	Optimality ratio artery (Γ_a)	0.77 ± 0.03	0.80 ± 0.1
9	Optimality ratio vein (Γ_v)	0.74 ± 0.03	0.76 ± 0.02

For the computation of branching angle, branching coefficient, asymmetry factor and optimality ratio separately for arteries and veins, detection of arteries and veins should be done. Vampire annotation tool [23] has been used for assessing arteries and veins in computing branching angles and width of the blood vessels in zone B of the binarized image. It is a complete annotation tool to annotate Optic disc, Fovea, Junctions, and Vessel Widths. This can aid in the calculation of branching angle, branching coefficients, asymmetry factor, and junction exponent using Eqs. (2)–(5) respectively.

The statistics of the computed parameters in the retinal fundus images for both control and stroke cases are given in Table 1. The evaluated parameters had been verified by an ophthalmologist.

These nine input parameters (FD, BA_a, BA_v, BC_a, BC_v, AF_a, AF_v, Γ_a, Γ_v) are given to a classifier for the automatic detection of normal healthy and retinal images of stroke patients. Dataset has 80 normal healthy images and 50 images of stroke patients. 50% of the normal and abnormal images are given (Normal n = 40, Stroke n = 25) for training the SVM classifier. Remaining 50% of the images were used for

testing (Normal n = 40, Stroke n = 25). SVM has been implemented using various kernels (linear, quadratic, polynomial) and the suitable option of kernel for stroke prognosis from retinal images has been investigated. All the kernels exhibit reasonable level of accuracy, but the linear kernel is a better option due to an improved accuracy of 93%.

5 Conclusion

Stroke is a life-threatening cerebrovascular disease. A novel approach for stroke prediction by using retina as a biomarker is implemented in this research work. Prior changes necessary to understand the pathophysiology of stroke is easily demonstrated in the retina by analyzing the retinal vasculature. This simple and non-invasive technique would allow patients to be treated more efficiently and at decreasing care costs. Larger database, effective computational tools and machine learning, when combined in an effective way will definitely add additional insights in stroke prognosis.

References

1. Warlow C, Sudlow C, Dennis M, Wardlaw J, Sandercock P (2003) Stroke. Lancet 362:1211–1224
2. Mozaffarian D, Benjamin EJ, Go AS et al (2015) Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation e29–322
3. Ropper AH, Brown RH (2005) Adams and victor's principles of neurology, 8th edn. The McGraw Hill Companies, Inc, New York, p 662
4. Biller J, Love BB, Schneck MJ (2008) Vascular diseases of the nervous system: ischemic cerebrovascular disease. In: Bradley WG, Daroff RB, Fenichel GM, Jankovic J (eds) Neurology in clinical practice, 5th edn. Elsevier, Philadelphia, p 1165–1166
5. Saine PJ, Patrick J, Tyler ME Ophthalmic photographers' society. In: Fundus photography overview. http://www.opsweb.org
6. Kumar V (2017) Understanding retinal changes after stroke. Open J Ophthalmol 7:281–292
7. Baker ML, Wang JJ, Liew G et al (2010) Differential associations of cortical and subcortical cerebral atropy with retinal vascular signs in patients with acute stroke. Stroke 41:2143–2150
8. Jeena RS, Sukesh Kumar A (2017) A comparative analysis of healthy retina and retina of a stroke patient. In: International conference on modelling and simulation, Kolkata
9. Grisan E, Foracchia M, Ruggeri A (2008) A novel method for the automatic grading of retinal vessel tortuosity. IEEE Trans Med Imag 27:310–319
10. Witt N, Chapmana N, Thom S (2010) A novel measure to characterize optimality of diameter relationships at retinal vascular bifurcations. Artery Res 4(3):75–80
11. Ishikawa H, Caputo M, Franzese N, Weinbren NL, Slakter A et al (2013) Stroke in the eye of the beholder. Med Hypotheses 80:411–415
12. Sabel BA, Henrich-Noack P, Fedorov A, Gall C (2011) Vision restoration after brain and retina damage: the residual vision activation theory. Prog Brain Res 192:199–262
13. De Silva DA, Manzano JJ, Liu EY, Woon FP, Wong WX et al (2011) Retinal microvascular changes and subsequent vascular events after ischemic stroke. Neurology 77:896–903

14. McGeechan K, Liew G, Macaskill P, Irwig L, Klein R et al (2009) Prediction of incident stroke events based on retinal vessel caliber: a systematic review and individual-participant meta-analysis. Am J Epidemiol 170:1323–1332
15. Ong YT et al (2013) Microvascular structure and network in the retina of patients with ischemic stroke. Stroke 44:2121–2127
16. Jeena RS, Sukesh Kumar A (2017) Artificial neural network based classification of healthy retina and retina of stroke patients, lectures on modelling and simulation; a selection from AMSE# 2017-N°2, pp 109–118
17. Aliahmad B, Kumar DK, Hao H, Unnikrishnan P, Azemin C, Zulfaezal M, Kawasaki R, Mitchell P (2014) Zone specific fractal dimension of retinal images as predictor of stroke incidence. Hindawi Publication 2014
18. Jeena RS, Sukesh Kumar A (2017) Stroke prognosis through retinal image analysis. AIVP Adv Image Video Process 5(2)
19. Frost S, Kanagasingam1 Y, Sohrabi H, Vignarajan J, Bourgeat P, Salvado O, Villemagne V, Rowe CC, Macaulay SL, Szoeke C, Ellis KA, Ames D, Masters CL, Rainey-Smith S, Martins RN, AIBL Research Group (2013) Retinal vascular biomarkers for early detection and monitoring of Alzheimer's disease. Transl Psychiatry 3:e233. https://doi.org/10.1038/tp.2012.150
20. Patton N, Aslam TM, MacGillivray T, Deary IJ, Dhillon B, Eikelboom RH et al (2006) Retinal image analysis: concepts, applications and potential. Prog Retin Eye Res 25(1):99–127
21. Joshi VS (2012) Analysis of retinal vessel networks using quantitative descriptors of vascular morphology. PhD thesis, University of Iowa, USA
22. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov. Springer 2(2):121–167
23. MacGillivray TJ, Trucco E, Cameron JR, Dhillon B, Houston JG, van Beek EJR (2014) Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions. Br J Radiol 87(1040):20130832

Distributed Representation of Healthcare Text Through Qualitative and Quantitative Analysis

J. R. Naveen, H. B. Barathi Ganesh, M. Anand Kumar and K. P. Soman

Abstract Many healthcare-related applications use pretrained embeddings, but these are often trained over general corpus which is mostly downstreamed to certain particular application. One problem noticed among such embeddings is that these are not efficient across various health text applications and even less number of research describe evaluation of these embedding for health domain. In this paper, distributional embedding model is performed to acquire a word representation on data crawled from Journal of Medical Case Reports. This distributed embedding model is analyzed qualitatively and quantitatively over crawled corpus. Qualitative evaluation is employed by cosine similarity on different categories and is visually represented. Quantitative evaluation performed with parts of speech tagging and entity recognition. The embedding model attained a cross-validation accuracy of 91.70% in parts of speech tagging for GENIA corpus and ensured 83% accuracy in the entity recognition of i2b2 clinical data.

1 Introduction

A massive effort is being made toward storing medical and health records in digital text form, which unlocked the possibility for computational processing of medical data and its application. This application must be able to derive meaning from the text. If we pass the raw data it cannot derive meaning from the text, because it is

J. R. Naveen (✉) · H. B. Barathi Ganesh · M. Anand Kumar · K. P. Soman
Amrita School of Engineering, Center for Computational Engineering and Networking (CEN), Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: naveenaksharam@gmail.com

H. B. Barathi Ganesh
e-mail: barathiganesh.hb@arnekt.com

M. Anand Kumar
e-mail: m_anandkumar@cb.amrita.edu

H. B. Barathi Ganesh
Arnekt Solution Pvt. Ltd., Pentagon P-3, Magarpatta, Pune, Maharashtra, India

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_23

just a string of character to software. The research field which derives information from midst of this digital text is natural language processing. Various efforts have been made to generate semantic structure from text. A mathematical model or an algorithm wants some form of numeric representation to work with. The naive way of representing a word in vector form is one hot representation but it is a very ineffective way for representing a huge corpus. In a more effective way, we need some semantic similarities to nearby points, thus creating the representation bring beneficial info about the word actual meaning, called word embedding models. These are categorized based on count and predictive word embedding models.

Word2vec is undoubtedly one of the best prevalent word embedding models that gives low-dimensional dense vector (word embedding) of the term (words and phrases). Word2vec with skip-gram architecture published [1] on 2013 had large influence on research group since the model gave effective training of dense word matrix and direct addition into downstream models [2]. So linguistic people and NLP scientists have made practice using word embedding for common NLP problems [3] such as information retrieval (IR), information extraction (IE) [4], sentiment analysis, and text summarization. As a result, the same trends continue in the fields of health care and bioinformatics also. They have been remarkably used to solve the problems in health care like biomedical named entity recognition (NER) [5] medical synonym extraction [6], relation extraction (RE), chemical–disease relation, drug–disease interaction, etc. Most of the time people are using general corpus such as Wikipedia and Google News [7] or the closed set small corpus for training the embedding models.

Most of earlier works for evaluating the word embedding models were focused on different methodologies. Baroni [8] used word embeddings created by context-predicting and context-counting semantic vectors where dependencies between the data and co-relation issues are studied with wide variety of lexical semantic tasks with varied parameter settings. Counter-predicting model got better result against the count-based model. Landauer, Thomas [9] came up with indirect knowledge acquisition from text using latent semantic analysis (LSA) and did research in local co-occurrence as similarities in a space. Turney [10] phrased analogy tasks as classification problem using unsupervised vectors. Different authors tried to modify this unsupervised way of learning text [11]. Ghanny [12] did quantitative analysis of Word2vec model for general English corpora. Recently, Pakhomo [2] evaluated word embeddings in bioinformatics domain.

Pretrained embeddings such as Word2vec trained on Google News and GloVe trained on Wikipedia embedding are available for a large corpus, which are already trained for common English words. These publicly available embeddings are trained for general purpose applications. But for healthcare applications, clinical NLP problems need a domain specific word embeddings [13]. In this paper, we evaluate our trained embedding model on Journal of Medical Case Reports with qualitative and quantitative methods against the standardized corpora.

2 Background

2.1 Word Embedding Model

Word2vec is irrefutable the best prevalent word embedding model, proposed by Mikalov [1] and it is motivated by Hinton's [14] distributed representation model. The fundamental hypothesis in word embedding or distributional methods is that words with similar sense tend to happen in similar context and they capture the relationship between words. Word2vec, basically, is a shallow neural network having one input, output, and a hidden layer that processes the input content and generate word embeddings on text data. In this feature vectors are formed on the co-occurrence possibilities and probabilities of the words present on the input, thus network has the ability to mimic linguistic properties of the data while training this neural network. Input to the model may be large but the generated output vectors are dense and low dimensional in space, with every unique word in the corpus having a corresponding vector linked with it. This makes the representation more computational and memory efficient compared to other models. This unsupervised machine learning framework has two architectures, skip-gram and continuous bag of words.

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{-s\leq k\leq s, k\neq 0} logp(Q_{n+k}|Q_k) \tag{1}$$

Skip-gram method try to maximize the average probability value for the word sequence $Q_1, Q_2, \ldots, Q_N$ given by Eq. 1. Here "s" indicates training context size that is directly related to the center word Q_n and $p(Q_{n+k}|w_k)$ is softmax function. In the skip-gram model, the context or surrounding word is predicted given the center word as the input and in CBOW model, given the surrounding words, the center word is predicted.

2.2 Sequence-Sequence Modeling

2.2.1 Part of Speech Tagging

Parts of Speech Tagging or POS is used for assigning grammatical groupings to the words in a sentence. Syntactic and semantics of nonstandard text can be found out using POS method. Correctness of POS labeling not just specifically influences the performance of pattern-based methods but additionally impacts correctness of parsing which, in general, utilizes the POS tags on the entities as a portion of the input. POS tagging is evaluated trigram Word2vec model followed by LSTM neural network. Trigram feature gives more dependency between words which reduces

disambiguty problem. Since Word2vec is an unsupervised way of embedding the words into numbers, the quality of Word2vec in health care should be tested. We took GENIA corpus for evaluating our Word2vec embedding and its hyperparameters.

2.2.2 Entity Recognition (ER)

Entity recognition (ER) is commonly used as part information retrieval problem like topic modeling, information extraction, and question answering. Detected entities (words, phrases) is subsequently connected to a concept in a knowledge base (referred as entity linking). In this work, entity recognition is used for quantitative evaluation of the word embedding model. i2b2 clinical corpora [15] is used for entity recognition task. This closed set small corpus is annotated in four different categories, namely, problem, treatment, test, and others. Systems with hand-crafted linguistic features tend to have good performance but finding out the relevant features are time-consuming task. But approach of neural network is different, they try to learn from the input itself where the word embedding model plays an important role.

2.2.3 Long Short-Term Memory Networks (LSTM)

Recurrent neural network (RNN) model is popular in sequence data analysis and NLP problems because the other neural networks such as multilayer perceptron (MLP) and convolutional neural networks (CNN) consider each input and output vector as independent while learning the weights and bias terms. In practice, RNN neural network also suffer from vanishing and exploding gradient problems. LSTM is called extended version of RNN because RNN has the problem of forgetting the features or patterns of the past data which might be helpful in classification or forecasting techniques. Here, the problem of RNN was overcome by introduction of LSTM network. LSTM is built up using memory blocks or cells. Each cell consists of input, output, and forget gates. The main function of gate is to control the information provided to the network. These gates help the LSTM network to remember information for longer duration than the RNN.

3 Corpus

Data used for creating Word2vec is scrapped using web crawler program. Contents of the corpus are taken from Journal of Medical Case Reports.[1] WebCrawler is basically a program made by user for collecting useful data from the web page. Initially, the crawler will find all the links on the seed page and follow those links to find the other pages. While collecting the corpus irreverent contents such as figures, tables, etc. are

[1] https://jmedicalcasereports.biomedcentral.com/articles.

Table 1 Statistics for creating word embedding model

Sl. no	Training data	Frequency
1	Total docs	4109
2	Total words	8,400,474
3	Total sent	473,675
4	Avg-words/sent	17.73

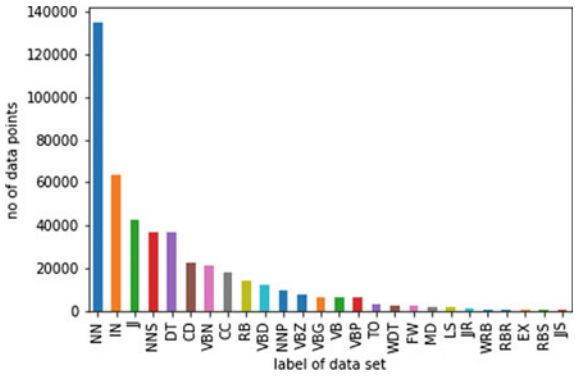


Fig. 1 Data statistics for POS tagging classification

omitted and rest are saved as text files. Like this, 4109 journal contents are crawled for creating healthcare data. Data statistics are explained in Table 1. Figure 1 shows the statistics of GENIA [16] data used for POS tagging. A total of 434,099 sentences from medical journal and 39,563 sentences from standard data set is used for creating word embedding model. Entity recognition collected on i2b2 [15] clinical data set.

4 Experiments

The system is implemented using Python libraries such as Scikit-learn, Gensim, and Keras machine learning libraries with an i7 Intel quad-core processor and NVIDIA GeForce GTX-1060 GPU. Once crawling or parsing is done the next challenge is cleaning corpora which highly depend on the source of data, parsing code, external noise, etc. which influence the preprocessing methods. Data cleaning is done to both the crawled data and test data. In order to clean text corpora, unwanted noise should be removed from raw data. So, we discarded the special characters such as ±, Latin alphabet, etc. which are not encoded by UTF-8 encoding scheme. Then removed the negligible category of class such as ?, =, interjection (UH), and predeterminers (PDT) from the test corpus and converted all the words into lowercase. After these preprocessing steps cleaned data is used for creating word embedding model. This word embedding model is passed to trigram embedding, which avoids ambiguity of word in the evaluation.

In this paper, we did the evaluation for word embedding using two methods, qualitative and quantitative analyses. Qualitative evaluation is employed by cosine similarity on different categories and is visually represented. Quantitative evaluation performed with parts of speech tagging and entity recognition.

4.1 Qualitative Experiment

In qualitative evaluation employed cosine similarity to calculate the similarity between words. Using this, we found five most similar words to target word. Cosine similarity can be defined by Eq. 2.

$$Similarity(V1, V2) = \frac{D1 \cdot D2}{||D1|| \cdot ||D2||} \tag{2}$$

where V1, V2 are the words considered and embedding space vector representation for V1 and V2 are given by D1 and D2, respectively. In this work, we considered three target words each from three different health categories, i.e., drug, symptom, and disorder. Using our proposed method, we found out five similar words for each category and was compared using similar words formed out already from available embeddings such as GloVe and Google News. Overview of the three health categories and the comparison of target word and similar word is given in Table 2.

Table 2 Qualitative evaluation with three categories of health key words

Category	Target word	Trained Word2vec model journal BMC	Word2vec from Google News	GloVe from Wikipedia
Drug	Aspirin	Azathioprine	Dose aspirin	Ibuprofen
		Rifampicin	Ibuprofen	Tamoxifen
		Capecitabine	Statins	Pills
		Doxorubicin	Statin	Statins
		Fluconazole	Calcium supplements	Medication
Symptom	Dyspnea	Fatigue	Dyspnoea	Shortness
		Diarrhea	Pruritus	Breathlessness
		Nausea	Nasopharyngitis	Cyanosis
		Arthralgia	Symptom severity	Photophobia
		Dizziness	Rhinorrhea	Faintness
Disorder	Diabetes	Psoriasis	Diabetics	Hypertension
		Neutropenia	Diabetic	Obesity
		Schizophrenia	Hypertension	Arthritis
		Epilepsy	Diabetes_mellitus	Cancer
		Obesity	Heart_disese	Alzheimer

4.2 Quantitative Experiment

To find the contribution of word embedding in health care, we use qualitative analysis. Here, word embedding was applied to POS tagging and entity recognition (ER). We modeled our analysis into classification task. For POS tagging we have 26 classes, whereas entity recognition we have four classes. We then evaluated the word embedding on a POS tagged representation of GENIA corpus, which is then considered at an application level by performing NER on i2b2 clinical data [15] to ensure the quality of representation. Trigram feature gives more dependency between words which reduces disambugity problem. Entity recognition task also performed with the same LSTM architecture used for POS tagging. Figure 3 shows Word2vec dimension tuning against pos tagging task where dimension of embedding model varied from 50 to 400 among these word embedding dimension 300 got better result. Same hyperparameters are used for both classifications. There were four categories of classes present in the i2b2 clinical corpus [15], namely, treatment, test, drug, and others. Figure 2 shows the common architecture used for POS tagging and entity recognition. We extracted the required data and label from GENIA [16] data set consisting of 23,467 sentences and i2b2 consisting of 16,107 sentences. From that data and category are separated (Fig. 2).

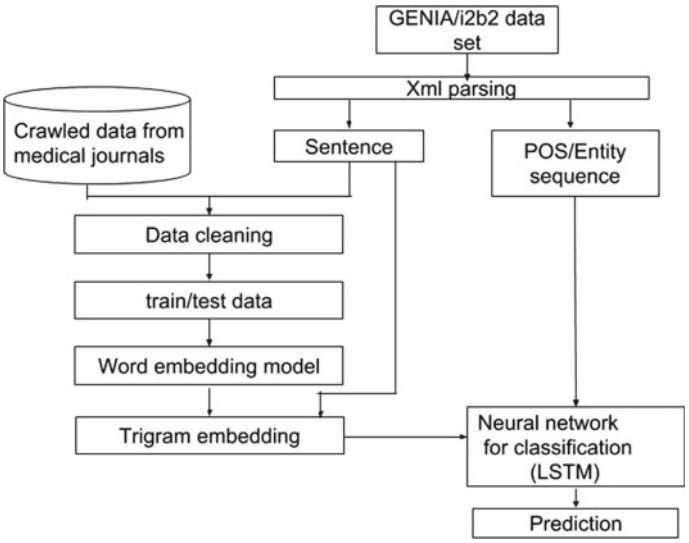


Fig. 2 Methodology for POS tag/entity recognition evaluation

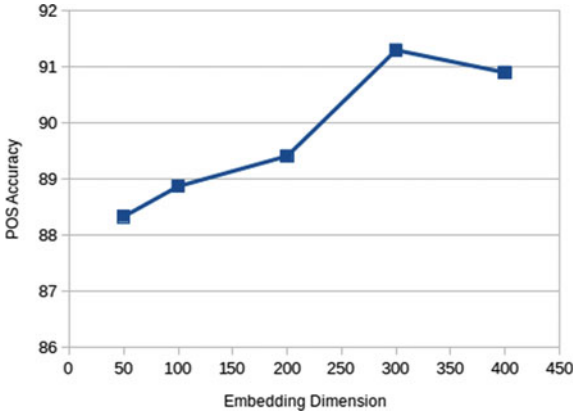


Fig. 3 Word2vec dimension tuning against POS tagging

5 Discussion

Qualitative evaluation is employed on three different categories drug name, symptoms, and disorder. Figure 4 shows the t-SNE[2] map generated by the target word in different categories. The t-distributed stochastic neighbor embedding maps high-dimensional vector space into two dimension. Vectors are close in embedding space can be visualized by t-SNE map where high-dimensional Euclidean distances between vectors converted into conditional probabilities that represent similarities. Here, orange, blue, and green data points represent the categories disorder, drug, and symptoms, respectively. From Fig. 4, it is clearly understood that word embedding maps different categories into different clusters.

From Table 2, disorder category, the word psoriasis is very closely related to the word diabetics since 27% psoriasis people have diabetic as their disorder and the word obesity is related word with diabetic. Rest three are different disorders. GloVe model finds out the word obesity (related word) but pretrained Google News embedding is more focused on morphological terms diabetic and diabetics. In symptoms, Word2vector created from journal BMC and GloVe find out common symptoms in health care. Google News failed to find out relevant terms. In drug category, all three models found out different drugs. GloVe find out only four drugs out of five words. We tested our models quantitatively with two supervised learning tasks POS tagging and entity recognition task. We use single layer LSTM model for classification. The activation function softmax perform better compared to the Relu function. The performance of these classifications is evaluated based on the confusion matrix generated after the classification. The values of precision, recall, and f-score are noted. We got more than 91% accuracy for POS tagging and got better result (83.09%) in entity recognition tasks. POS tagging is done with GENIA [16] corpus and entity recognition done on i2b2 clinical data set [15]. Clinical contents are more ambiguous

[2]t-distributed stochastic neighbor embedding.

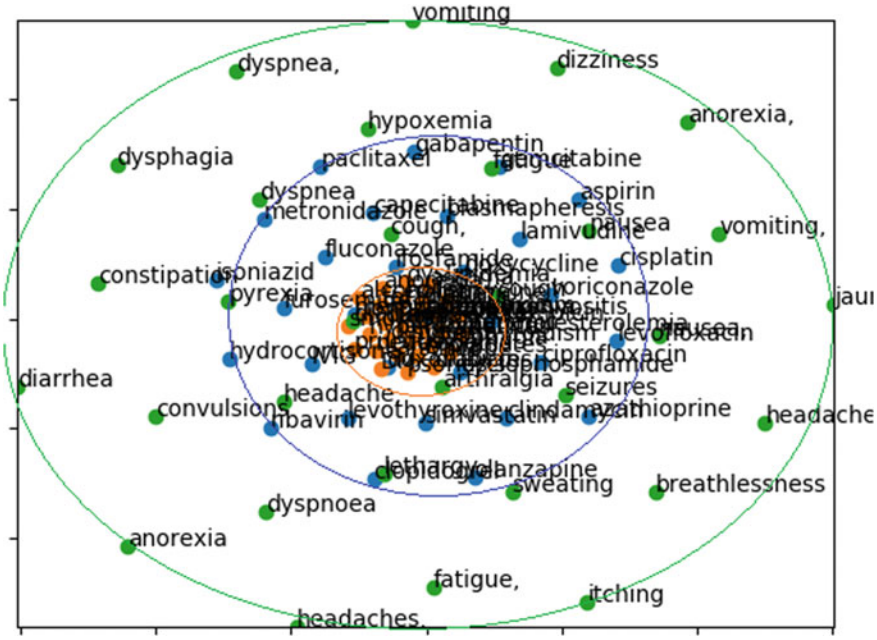


Fig. 4 t-SNE map of Word2vector created from journal BMC with different target word categories

Table 3 Cross-validation accuracy for POS tagging and entity recognition (k-fold)

K-fold accuracy	k = 5	k = 10
POS tag	91.09	91.70
Entity recognition	82.90	83.09

Table 4 Hyperparameters for Word2vec

Hyperparameter	Word2vec from BMC
Learning rate	0.1
Dimension	300
Window size	1
Min-count	1
n-epoch	5

and noisy compared to web literature, so NER perform relatively low in classification which is shown in Table 3.

Word2vec model parameters such as vector size, number of forward + backward pass (epoch), learning rate (determines how quickly or slowly model updates the parameter), and window size (number of training samples required) are fixed based on classifications. The parameters are listed in Table 4.

6 Conclusion and Future Work

The paper evaluates the Word2vec embedding qualitatively and quantitatively in healthcare domain. In qualitative evaluation, we employed cosine similarity to compute the similarity between words. Using this, we found five most similar words to target word, where our embedding model better performed compared to the other two. Quantitative evaluation done with two classification tasks, POS tagging, and named entity recognition using deep learning with the help of LSTM network and got better accuracy for these multiclass problems. More accuracy can be attained by inducing more corpora and by adding deeper layer feature learning capabilities to the model. This work can be extended to annotated Word2vec model in healthcare domain.

References

1. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119
2. Pakhomov SVS, Finley G, McEwan R, Wang Y, Melton GB (2016) Corpus domain effects on distributional semantic modeling of medical terms. Bioinformatics 32(23):3635–3644
3. Barathi Ganesh HB, Anand Kumar M, Soman KP (2018) From vector space models to vector space models of semantics. In: Forum for information retrieval evaluation. Springer, Berlin, pp 50–60
4. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, Liu S, Zeng Y, Mehrabi S, Sohn S et al (2017) Clinical information extraction applications: a literature review. J Biomed Inf
5. Tang B, Cao H, Wang X, Chen Q, Xu H (2014) Evaluating word representation features in biomedical named entity recognition tasks. BioMed Res Int
6. Jagannatha A, Chen J, Yu H (2015) Mining and ranking biomedical synonym candidates from Wikipedia. In: Proceedings of the sixth international workshop on health text mining and information analysis
7. Gurulingappa H, Toldo L, Schepers C, Bauer A, Megaro G (2016) Semi-supervised information retrieval system for clinical decision support. In: TREC
8. Baroni M, Dinu G, Kruszewski G (2014) Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: long papers), vol 1, pp 238–247
9. Landauer TK, Dumais ST (1997) A solution to plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol Rev 104(2):211
10. Turney PD (2008) A uniform approach to analogies, synonyms, antonyms, and associations. In: Proceedings of the 22nd international conference on computational linguistics, vol 1. Association for Computational Linguistics, pp 905–912
11. Barathi Ganesh HB, Reshma U, Anand Kumar M, Soman KP (2017) Amrita_cen_nlp@ irled 2017
12. Ghannay S, Favre B, Esteve Y, Camelin N (2016) Word embedding evaluation and combination. In: LREC, pp 300–305
13. Barathi Ganesh HB, Anand Kumar M, Soman KP (2016) Distributional semantic representation in health care text classification. In: International conference on forum of information retrieval and evaluation, pp 201–204

14. Hinton GE, McClelland JL, Rumelhart DE et al (1986) Distributed representations. Parallel Distrib Process Explor Microstruct Cogn 1(3):77–109
15. Uzuner Ö, South BR, Shen S, DuVall SL (2011) 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. J Am Med Inf Assoc 18(5):552–556
16. Verspoor K, Cohen KB, Lanfranchi A, Warner C, Johnson HL, Roeder C, Choi JD, Funk C, Malenkiy Y, Eckert M et al (2012) A corpus of full-text journal articles is a robust evaluation tool for revealing differences in performance of biomedical natural language processing tools. BMC Bioinf 13(1):207

Detection of Lymph Nodes Using Centre of Mass and Moment Analysis

R. Akshai, S. Rohit Krishnan, G. Swetha and B. P. Venkatesh

Abstract A novel technique to determine specific spot where the lymph node is present and also analyse whether it causes cancer or not. Lymph nodes are small structures that filter harmful substances. Their incessant growth may lead to cancer. The two-dimensional image obtained from CT scan is transformed into three-dimensional cube using 3D slicer tool. The centre of mass is calculated and moment of inertia is obtained from centre of mass. The value of moment of inertia is decreased, until a node is detected or it meets a particular threshold. Thus, the proposed method enhances the node detection accuracy and reduces the time taken for detection using possible simple mathematical computations.

Keywords Computed tomography · 3D slicer · Moment of inertia
Moment analysis · 3D virtual view · N-body system · Particle-particle mesh
Centre of mass · Centroid · Projection plane

1 Introduction

Cancer is a chronic disease which arises when there is abnormal cell growth in a particular part of the body. It spreads to other parts if it is left unattended, and it may even lead to death. One of the symptoms of cancer is swelling of lymph nodes. Lymph node is a kidney-shaped organ and an adaptive immune system. There are several lymph nodes all over the body. Computerized tomography is a technology which is used to scan the body. It generates a two-dimensional image of the scanned part. By analysing the image, the swollen and cancer-causing lymph nodes are detected and appropriate treatment is given.

R. Akshai · S. Rohit Krishnan · G. Swetha (✉) · B. P. Venkatesh
Department of Computer Science & Engineering, Sri Sai Ram Engineering College, Chennai 600044, India
e-mail: swechess97@gmail.com

R. Akshai
e-mail: akshairamesh21@gmail.com

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_24

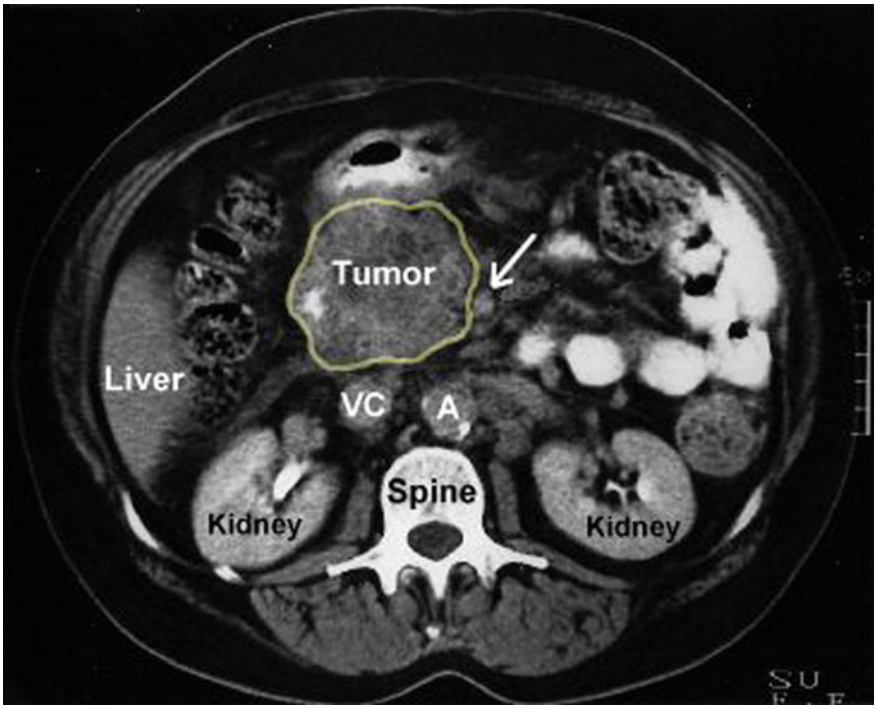


Fig. 1 Observation of lymph nodes in different parts of the body

There are several ways to analyse the CT scan. As analysing the two-dimensional image is a tedious task, usually, the two-dimensional image is converted into three-dimensional image using any of the existing conversion tools for analysing it more efficiently. A method has to be determined to check the presence of lymph nodes. This method should be efficient such that the lymph node is discovered soon and also in a short period of time. CAD is one such powerful tool to analyse the CT scan images in less time with more accuracy. The main disadvantage of analysing two-dimensional image is that it gives more false positives (FP) when compared to analysing 3D images.

Hence to overcome the above-mentioned disadvantage, we devised an algorithm to efficiently detect lymph nodes within a limited time and analyse whether it is a cancer node. The three-dimensional image obtained from the slicer tool is analysed for detecting the lymph nodes by taking into account the centre of mass and moment of inertia. This method helps to quickly identify the node's location (Fig. 1).

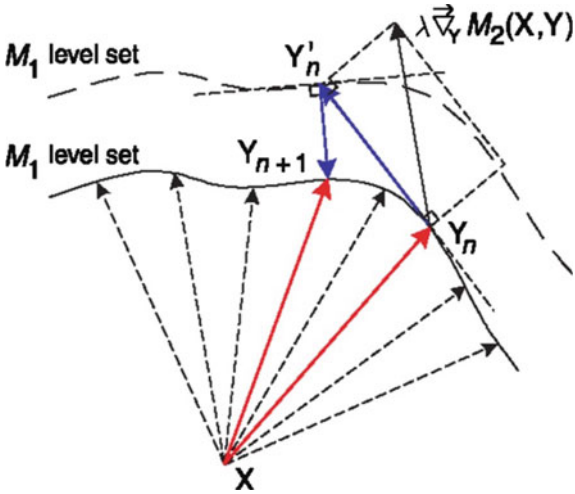


Fig. 2 Computing moment of inertia at different coordinates

2 Proposed System

In this system, we analyse the scanned CT scan image based on the moment analysis. The two-dimensional image will be taken as an input and placed in a three-dimensional cube. The 3D slicer tool is used for the analysis of the image. The centre of mass for all the six faces of the cube is calculated. Then, the coordinates are mapped against each other and the spectral line obtained from the mapping is examined. If we find the lymph along the axis, then we get the coordinates say x, y, and z. If it is not present, then we calculate the moment of inertia from the centre of mass and these values are stored in array list. After storing values of centre of mass, we sort in descending order. Once the table is sorted, we check for the lymph nodes presence from the sorted table. At iteration say i, if the lymph node is detected, the coordinates are obtained and it is converted into two-dimensional points (Fig. 2).

3 Implementation and Algorithm

The detection process involves the conversion of input CT scan result (2D image in DICOM format) into the corresponding 3D image encapsulated within a three-dimensional cube. This process is achieved by using the 3D Slicer tool.

The 3D virtual box has the fixed aspect ratios. Also, we know that the cube's centre of mass lies in the intersection point of cube's diagonals, and this coordinate centre of mass value is taken as reference point. For each face of the cube, the centre of mass is computed separately, leading to the values $r_c(i)$ where $i = 1, 2, 3, 4, 5$, and 6 (Figs. 3 and 4).

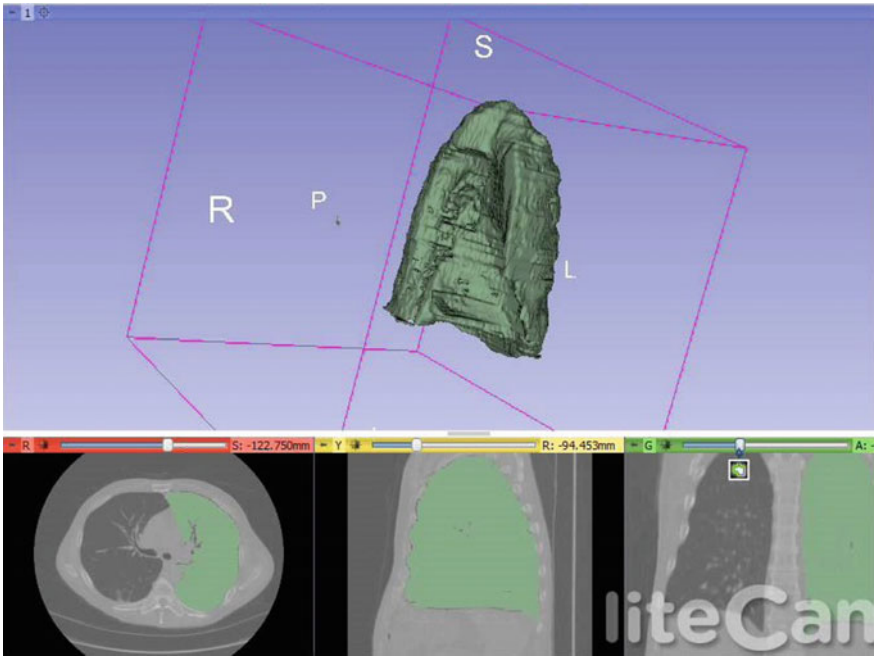


Fig. 3 The 2D image encapsulation in 3D slicer tool

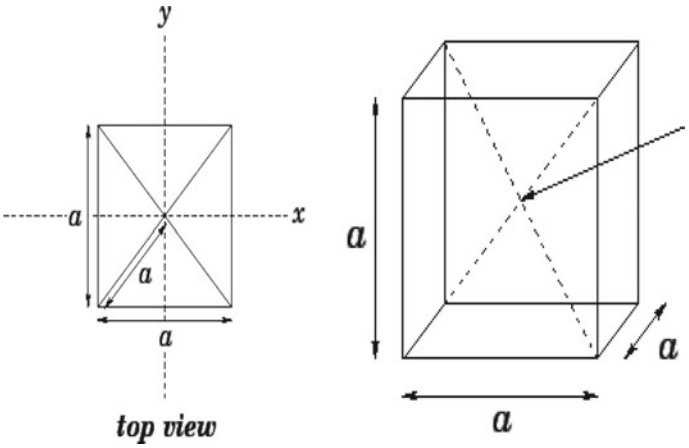


Fig. 4 Initial coordinates using centre of mass across different planes

If the total mass of the object is M, and the object contains N elements, such that each *i*th element has mass m and position vector r_i, then the position vector for the centre of mass can be expressed as

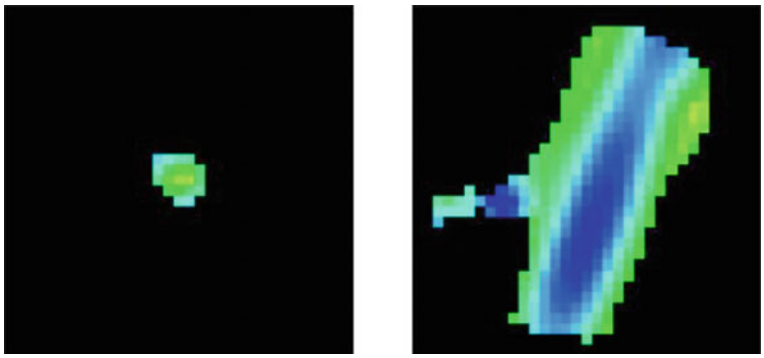


Fig. 5 Result after applying moment of inertia filter

$$\mathrm{r}_{cm} = \frac{1}{M} \sum_{i=1,N} m_i \mathrm{r}_i .$$

Thus, the position vector of the 3D cube is the reference coordinate and the position vectors of each face of the cube are mapped around it. If the lymph node is detected at any of the points in the centre of mass in 2D or 3D projection, the corresponding coordinate combinations (all eight possibilities xyz, yzx, xzy … or all four possibilities xy, yx …) are verified and thus, the partial result and confirmation of presence of lymph nodes is diagnosed.

If no lymph node is detected at the reference point as well as the 2D face coordinates, then the corresponding moment of inertia is computed for each of the coordinates. The values are used as a dataset arranged in descending order of the moment of inertia value.

3.1 Computation of Moment of Inertia

Nodules are the globular regions in the body, hence this can be used to distinguish nodules and blood vessels, since the blood vessels are cylindrical regions. These two regions are differentiated using moment of inertia where voxel intensities are considered as mass after applying the radial suppression filter. First, a gamma correction is applied to f(x, y, z) as follows:

$$\mathrm{f}\gamma(\mathrm{x}, \mathrm{y}, \mathrm{z}) = \{\mathrm{f}(\mathrm{x}, \mathrm{y}, \mathrm{z})\}\gamma,$$

where f γ (x, y, z) is the corrected image and γ is the gamma correction parameter. The moment of inertia tensor I is defined by I (Fig. 5).

$$\mathrm{I} = \big[\big[\mathrm{Ixx\ Ixy\ Ixz}\big]\big[\mathrm{Iyx\ Iyy\ Iyz}\big]\big[\mathrm{Izx\ Izy\ Izz}\big]\big], \text{ (Array of inertia vectors)}$$

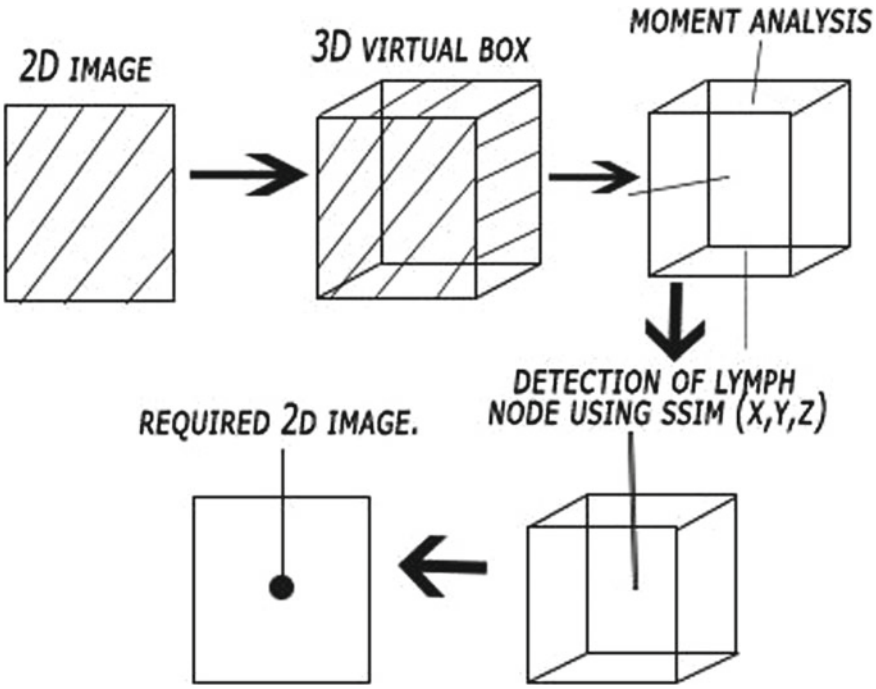


Fig. 6 The process flow for detection of lymph nodes

$$\text{where Ixx} = \text{X(x, y, z)} \in \text{Ksp(rf ix) g } \gamma\text{(x, y, z)(y 2 + z 2)},$$
$$\text{Iyz} = \text{Izy} = -\text{X(x, y, z)} \in \text{Ksp(rf ix)g } \gamma\text{(x, y, z)yz}.$$

Thus for each unique MI value in the dataset, the corresponding coordinates are checked for the presence of lymph node, this is done by using the moment analysis and the SSIM technique.

The moment images from a spectral line cube task moment can be used to produce moment images. The moment axis should be the first or third in the cube (e.g. vxy or xyv order). Reorder the axes if the moment of the second axis is required for any purpose. The moments which are evaluated for each spatial pixel along the velocity axis are defined as

$$M_n = \int I(v) v^n dv$$

where I(v) is the intensity at a given velocity v.

The overall functioning of the system can be summarized as shown in the block diagram Fig. 6 (Fig. 7).

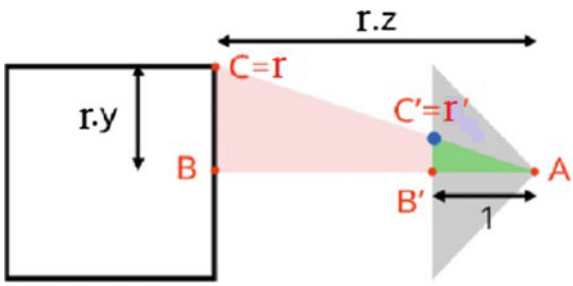


Fig. 7 Mapping of 3D to 2D coordinates

Translation of 3D space coordinates to 2D coordinates:

$$r'.y = r.y/r.z;\ r'.x = r.x/r.z$$

The particle-particle approach is used to detect the next neighbouring lymph node.

4 Conclusion

The proposed method can be used to detect and analyse the lymph nodes which may lead to cancer. As centre of mass and moment of inertia are taken into consideration for detecting the presence of lymph nodes, this makes it simpler and efficient to detect the location of the node. This reduces a lot of time to spot the exact location of the node. Since three-dimensional images are used for analysing the presence of lymph node, it gives more accurate results.

References

1. Hishida T, Miyaoka E, Yokoi K et al (2016) Lobe-specific nodal dissection for clinical stage I and II NSCLC: japanese multi-institutional retrospective study using a propensity score analysis. J Thorac Oncol 11:1529–1537. https://doi.org/10.1016/j.jtho.2016.05.014
2. Zhao W, Shi F (2018) A method for lymph node segmentation with scaling features in a random forest model. Curr Proteomics 15(2):128–134(7)
3. Takizawa H, Yamamoto S, Shiina T (2010) Recognition of pulmonary nodules in thoracic CT scans using 3-D deformable object models of different classes. Algorithms 3:125–144
4. Warfield S, Zou K, Wells W (2004) Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imag 23(7):903–921
5. Consortium for open medical image computing in CAMELYON17 Webpage, 2017, Background, Evaluation
6. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In Proceedings of conference on computer vision and pattern, pp 886–893

7. Song Y, Cai W, Eberl S, Fulham M, Feng D (2011) Thoracic image case retrieval with spatial and contextual information. In: Proceedings of international symposium on biomedical imaging (ISBI) 2011, pp 1885–1888
8. Tunio MA, Al Asiri M, Mohamed R, Al-Dandan S (2013) Supraclavicular lymphadenopathy: initial manifestation of metastasis in carcinoma of cervix. Case Rep Obstet Gynecol 2013 409196[DOI]
9. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. CVPR 1:511
10. Barbu A, Suehling M, Xu X, Liu D, Zhou SK, Comaniciu D (2010) Automatic detection and segmentation of axillary lymph nodes. Med Image Comput Comput Assist Interv 13:28–36

Estimation of Elbow Joint Angle from Surface Electromyogram Signals Using ANFIS

P. Rajalakshmy, Elizabeth Jacob and T. Joclyn Sharon

Abstract The estimation of joint angle finds wide application in fields such as robotics, prosthetics, ergonomics, etc. Various estimation techniques have been discussed in the literature. The estimation of joint angle using Surface Electromyogram (sEMG) signals has gained importance due to its ability to recognize continuous human motion. This requires feature extraction from the acquired sEMG signals in order to develop an estimation model. In this paper, an attempt has been made to extract a few significant time domain features from sEMG signals acquired from the biceps brachii for four different subjects. The feature signals obtained using sliding window technique is further used to develop an estimation model using a suitable training algorithm. The trained model is eventually used to control the elbow joint angle.

Keywords Joint angle · sEMG · Estimation · Feature extraction · Sliding window ANFIS

1 Introduction

Surface Electromyogram (sEMG) is a noninvasive biosignal. An assist device such as the upper and lower limb exoskeletons can be controlled using surface Electromyography to support human being [1]. By estimating force, torque, and angle, the EMG signal is used to detect user intended movement [2]. The EMG signals are acquired from biceps brachii to estimate the elbow joint angle. In the course of recognizing

P. Rajalakshmy (✉) · E. Jacob · T. Joclyn Sharon
Department of Electronics and Instrumentation, Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: rajalakshmy@karunya.edu

E. Jacob
e-mail: elizabethjacob@karunya.edu.in

T. Joclyn Sharon
e-mail: joclynsharon@gmail.com

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_25

extended human motion estimation from surface Electromyogram, three key steps are elaborated [3]:

(i) Feature signals from the sEMG are extracted.
(ii) Extended human motion from the feature signals of sEMG is estimated.
(iii) The motor is driven to control the elbow joint angle.

The amplitude of the EMG signal is relevant to the isometric muscle force. Isometric muscle force is usually the signal which is used to calculate the EMG signal-to-noise ratio [4]. Basically, the raw sEMG signal is very weak so it will produce the energy in the frequency range of 13–500 Hz. In order to get information about the elbow joint angle during flexion and extension motions, the recorded electromyogram signals have to be processed using the time domain features after completing the signal acquisition. To estimate the elbow joint angle time domain features were investigated.

The time domain features are Integrated EMG (IEMG) and Zero Crossing (ZC).

The time domain feature signals, acquired using the sliding window function, are estimated directly from the surface EMG signal. The joint angle is estimated using machine language algorithm after the feature extraction stage. To determine the correctness of the estimation, an appropriate feature is chosen. Thereby, the acquired signals are used to drive the motor and control the elbow joint angle. The applications of elbow joint angle estimation are robotics, rehabilitation, and ergonomics.

2 Methods

2.1 Signal Acquisition

EMG signals are widely used to provide information regarding contracting of muscles, muscle fatigue, and reflex action. These signals find widespread use in many biomedical applications. With the help of three electrodes, the sEMG signals are measured from the biceps brachii of an individual. The electrodes used are Ag/AgCl electrodes which are of the disc type. The measuring electrodes are placed with a distance of 1 cm between them on the biceps brachii muscle and the reference electrode is placed on the flexor carpi ulnaris muscle. The electrodes are given as input to the Biokit which performs signal conditioning operations on the input sEMG signals [5]. The analog voltage from the Biokit is given to the National Instruments myDAQcard which acts as the interface between the electrodes and the LabVIEW Development System. Parallely, the angular displacement is measured using a three-axis accelerometer and read using LabVIEW.

In total, four healthy subjects with an average age of 28 years and an average weight of 62 kg are subjected to signal acquisition. The signal acquisition is done when the subjects are seated comfortably for periods of 10 s. The flow diagram of the entire system is shown in Fig. 1.

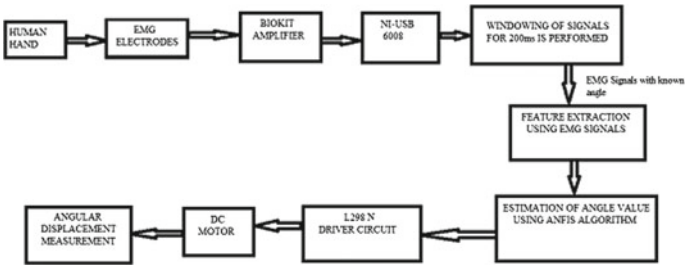


Fig. 1 Block diagram of the system

2.2 Feature Extraction

The sEMG signals obtained from the biceps muscles are segmented into windows of length 200 ms each. The windowed signals are given as input to the time domain features. The extraction of time domain features like Integrated EMG (IEMG) and Zero Crossing (ZC) is done from each window. In a similar way, even the angular displacement values are also segmented into 200 ms windows and an average of these values is taken. These features are represented with the help of formulae. In each of these formulae, x_i indicates the ith EMG signal, x_{i+1} indicates the $(i+1)$th count of EMG signal, and N indicates the length of the signal.

(1) **Zero Crossing (ZC)**

The number of times a given signal crosses certain threshold value is termed as zero crossing. The formula for Zero Crossing is [3]:

$$ZC = \sum_{i=1}^{N} sign\left(-x_i^* x_{i+1}\right) \text{ and } |x_i - x_{i+1}| \geq 0$$

$$sign(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases}$$

(2) **Integrated EMG (IEMG)**

The absolute summation of EMG samples over a window of length N of the signal is termed as Integrated EMG. The formula for Integrated EMG is [3]

$$IEMG = \sum_{i=1}^{N} |x_i|$$

The signals obtained after feature extraction are given to the machine learning algorithm for training and testing purposes.

2.3 Estimating the Elbow Joint Angle Using ANFIS

The estimation of joint angle is done from the feature signals extracted by choosing the best relevant features from the given time domain features [6]. By experimentation, it is found out that IEMG and ZC vary in their value with the change in angular displacement. The proposed algorithm for the estimation of joint angle is ANFIS.

An adaptive neuro-fuzzy inference system or Adaptive Network-Based Fuzzy Inference System (ANFIS) is a kind of artificial neural network that is based on Takagi–Sugeno fuzzy inference system. Since it takes into account of both neural networks and fuzzy logic principles, it has the ability to capture the benefits of both in a single frame. Its inference system correlates to a set of fuzzy IF-THEN rules that have learning capability to relate to nonlinear functions. Hence, ANFIS is considered to be a universal estimator.

Sugeno-type fuzzy inference systems are a type of FIS, wherein ANFIS uses a hybrid learning algorithm to identify the membership function parameters of single-output. An amalgamation of least squares and backpropagation gradient descent methods are used for training FIS membership function parameters to model a given set of input/output data [7].

2.4 Control of Elbow Joint Angle

After the known joint angle is found, proper control of the DC gear motor for joint motion should be established [8]. A gear motor is a specific type of electrical motor that is designed in such a way that it produces high torque while maintaining a low-speed motor output [6]. An arm arrangement which is open ended on one end is attached to a large plate with gear teeth. The arm slides on small wheels along a groove. The control signals given to gear motor cause the gear assembly to rotate, hence, triggering the gear train which causes motor speed to decrease and facilitates smooth movement [9].

This helps in the movement of the attached arm to the desired angle given by the control signals obtained from LabVIEW Software.

3 Results

In this section to improve the performance of motion estimation, the multiple time domain feature signals of sEMG are verified. The raw sEMG signal is acquired from the biceps brachii muscles of human forearm after amplifying and filtering is shown in Fig. 2. This raw electromyogram signal is then windowed into 200 ms using a sliding window prior to which two-time domain features, Integrated EMG (IEMG), Zero Crossing (ZC), are calculated from each window. The acquired surface

electromyogram signals and the angular displacement data are processed offline for feature extraction and evaluation.

In Fig. 2, the feature values increase from the movements of rest to flexion–extension. The features, IEMG and ZC, vary with the varying angular displacement data and average angular velocity data obtained from the forearm. This is because of the reason that the amplitude of the signal increases with increase in the number of motor units. IEMG also increases with increase in sEMG amplitude. Since IEMG is the mathematical integral of absolute value of sEMG signal. Also with increase in sEMG signal amplitude, the count of number of times the signal crosses the zero reference line increases, leading to an increase in ZC value.

The feature signals obtained by using sliding window are fed as input to the ANFIS model and the control signals, output, are used as commands to drive the motor to the required joint angle.

The ANFIS model is initially trained for two different angles 45° and 90°. The estimated angle values for a test input are shown in Fig. 3 and the values are tabulated in Table 1. The graphical representation of the estimated angle values is shown in Figs. 4 and 5.

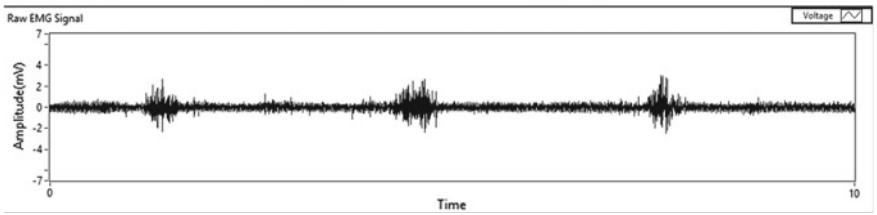


Fig. 2 The raw sEMG signal

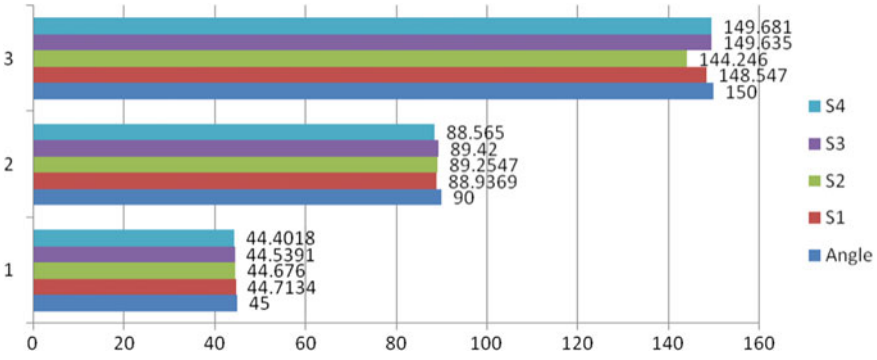


Fig. 3 The comparison plot for different subjects for different angles

Table 1 The table comparing the known and estimated angle values

Subject	Estimated angle	
	Actual angle = 45°	Actual angle = 90°
Subject 1	44.7134	88.9369
Subject 2	44.676	89.2547
Subject 3	44.5391	89.42
Subject 4	44.4018	88.565

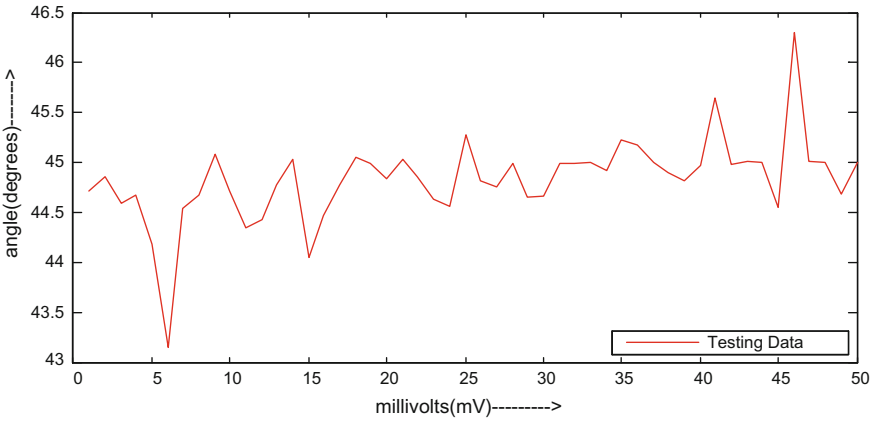


Fig. 4 The estimated angle obtained when comparing training and testing data, known angle = 45°, S1

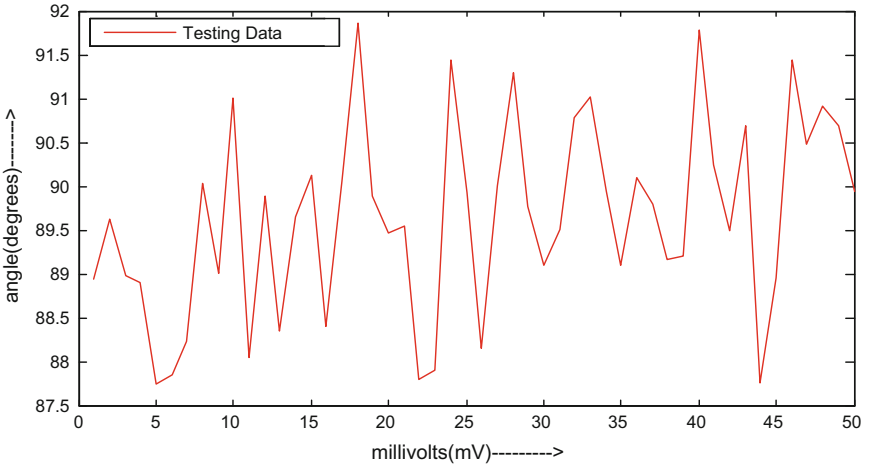


Fig. 5 The estimated angle obtained when comparing training and testing data, known angle = 90°, S1

4 Conclusion

In this work, a surface Electromyogram signal is used to control the human elbow joint motion using a motor controlled arm. The sEMG signals are acquired from the biceps brachii during the flexion–extension of the forearm using surface electrodes that are placed on the muscles of human subjects. From the acquired signals, time domain features like IEMG and ZC are extracted. These time domain features along with the manipulator kinematics, here, the angular position of the robot arm forms the training data set to the neuro-fuzzy algorithm to form the estimation model. During the testing phase, the time domain features are presented to the estimation model which predicts the angular position of the robot arm and the real-time hardware moves the arm to the desired position.

Acknowledgements The technical support and laboratory facilities provided by the parent institution for completing this project are gratefully acknowledged. All the staffs who guided to the best of their knowledge are gratefully acknowledged. The authors of all the journals listed in the reference section are also acknowledged gratefully.

References

1. Farina D, Merletti R, Enoka RM (2004) The extraction of neural strategies from the surface EMG. J Appl Physiol 96:1486–1495
2. Shaw L, Bagha S (2012) Online EMG signal analysis for diagnosis of neuromuscular diseases by using PCA&PNN. Int J Eng Sci Technol 4:4453–4459
3. Choi Y, Lee AY, Yu HJ (2011) Human elbow joint angle estimation using elctromyogram signal processing. Inst Eng Technol 5:767–775
4. Hussain MS, Mohd-yasin F, Reaz MBI (2006) Technique of EMG signal analysis: detection processing, classification and applications, vol. 24, pp 1342–1350
5. Einzinger PD, Livshitz LM, Mizrabi J (2001) Interaction of array of finite electrodes with layered biological tissue: effect of electrode size configuration. IEEE Trans Neural Syst Rehabil Eng 9:355–361
6. Hanung Adi Nugroho H, Wahunggoro O, Triwiyanto (2017) An investigtion into time domain features of surface electromyography to estimate the elbow joint angle. Adv Electr Electron Eng 15:448–458
7. Al-Jumaily A, Aung YM (2013) Estimation of upper limb joint angle using surface EMG signal. Int J Adv Robot Syst 10:369–377
8. Sivanandan KS, Ramakrishna R, Raj R (2016) A real time surface electromyography signal driven prosthetic hand model using PID controlled DC motor. In: The Korean society of medical & biological engineering and Springer, vol 6, pp 276–286
9. Raj R, Sivanandan KS (2015) Estimation of elbow joint angle from Time domain features of SEMG signals using fuzzy logic for prosthetic control. Int J Current Eng Technol 5:2078–2081

Video Stabilization for High-Quality Medical Video Compression

D. Raveena Judie Dolly, D. J. Jagannath and J. Dinesh Peter

Abstract A medical video analysis provides a forum for the dissemination of diverse results in the area of medical field. The swiftly sprouting changes in the health care entail the modern advances with imaging. The effect of camera motion from a video stream leads to improper diagnosis in case of medical video processing. Hence, video stabilization technique enables the video stream to generate a stabilized video. Compression also plays a very vital role in the field of clinical and image analysis, since it carries massive information thereby reducing the file size. To further ensure effectiveness in diagnosis, a video stabilization is performed before compression. The results proved that the proposed methodology offered a substantial increase in the quality of video by **1.25 dB** peak-signal-to-noise ratio and **0.06** structural similarity index.

Keywords Video stabilization · Medical video compression · Intra-frame Predictive frame · Bidirectional frame · Group of pictures (GoP)

1 Introduction

To alleviate the undesired motions between successive frames in a video, stabilization must be performed. A video stabilization algorithm for hand-held camera videos is suggested using differential motion estimation [1]. In order to ascertain the

D. Raveena Judie Dolly (✉) · D. J. Jagannath
Electronics and Communication Engineering, Karunya Institute of Technology & Sciences, Coimbatore 641114, India
e-mail: dollydinesh@karunya.edu

D. J. Jagannath
e-mail: jagan@karunya.edu

J. Dinesh Peter
Computer Science and Technology, Karunya Institute of Technology & Sciences, Coimbatore 641114, India
e-mail: dineshpeter@karunya.edu

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_26

Fig. 1 Video stabilization system

undesirable motion component, a filtering algorithm was proposed that takes into consideration the system constraints with respect to the amount of the corrective motion that can be applied on each frame [2] as shown in Fig. 1.

Two motion blurred videos are considered in [3]. FAST detection method is applied for feature point detection as suggested in [3]. The use of FAST detection method displays enhanced outcome in terms of stabilization. FAST detection technique is suitable in enhancing the quality of low-grade video surveillance cameras. A motion smoothing technique is suggested to alleviate the uncomfortable effect due to camera motion perturbations, and to temporally smooth out the camera motions for the whole video as suggested in [4]. After the survey of various techniques, it is suggested that motion smoothing adaptive filter is better than normal filters as mentioned in the article [5].

2 Methodology

It is desirable to adopt video stabilization before performing compression for medical videos. This may result in exact diagnosis by the doctors. Stabilization is performed to eliminate the undesired motion or degradations and to prevent local dynamics as suggested in [6, 7]. The proposed methodology includes raw videos collected from various sources. Those videos are converted into video sequences for stabilizing the video as shown in Fig. 2. The stabilized video enters into a process of compression to get a compressed video.

2.1 Video Stabilization

In each succeeding video frame, the movement of the target relative to the previous frame is determined to eliminate undesirable translational camera motions thereby

Fig. 2 Process flow of the proposed methodology

to generate a stabilized video. Motion estimation is finding the motion vectors from neighboring frames in a video sequence. The motion vectors may be the entire frame or an explicit part, for instance, rectangular blocks, arbitrary-shaped patches, or a single pixel. Hence, the motion estimation is considered to be the most important stage. Camera motion often comprises global transformation. Therefore, it is necessary to get an accurate estimate of the global motion while performing motion estimation as recommended in [8]. Secondly, the motion smoothening is performed in stabilization to eliminate the jitters and undesired movements, usually considered as high-frequency motion such as vibration in the captured video. These undesired movements can be effectively eliminated by low-pass filters. Motion smoothing is often mentioned as motion filtering which is performed to reduce spatial redundancy between frames. Lastly, motion compensation is performed to eliminate temporal redundancy between frames so as to produce stabilized video.

2.2 Video Compression

The stabilized video undergoes a most important step of reducing the file size. Hence, the basic video processing algorithms of MPEG-2 are implemented. The video taken for implementation are converted into video sequences in the IBP format as group of pictures (GoP) as mentioned in [9]. Spatial redundancy is abridged by applying eight times horizontally and vertically a 8×1 DCT transform. Quantization is performed for the transform coefficients, thereby reducing high-frequency coefficients, which are scanned in zig-zag order proceeding from the DC coefficient at the upper left corner of the block and coded using Huffman tables referred also as variable length coding (VLC). They are applied to macroblocks according to the IBBP group of picture since it considerably reduces the bit rate and achieves high compression rates [10]. It identifies the current macroblock in previously transmitted pictures at the exact position in the image or displaced by the "motion vector". The macroblock with lowest average error is chosen as the reference, since exact match does not appear. The "error macroblock" is then processed for reducing the spatial redundancy. Motion compensation is performed for reducing the temporal redundancy.

3 Implementation Results and Discussion

A series of MRI dicom slices recordings obtained from different patients from various clinical laboratories around Coimbatore district, Tamil Nadu, whose age were between 19 and 62 were considered for testing. Table 1 indicates the outcome of few case studies. It is observed that the quality was maintained without much degradation and the structural similarity index was observed to yield better result while executing video stabilization with MPEG-2 compared to conventional MPEG-2 alone.

Table 1 PSNR and SSIM of medical videos case study

Cases	MPEG-2		Video stabilization with MPEG-2	
	PSNR	SSIM	PSNR	SSIM
MRI brain recording 1	32.25	0.7	33.26	0.74
MRI brain recording 2	33.67	0.67	34.05	0.73
MRI brain recording 3	31.91	0.68	33.72	0.71
MRI brain recording 4	31.77	0.67	33.51	0.73
MRI brain recording 5	32.13	0.66	33.41	0.77

From the table, it is observed that implementation with video stabilization with MPEG-2 yields good result.

4 Conclusion

Video compression plays a major part in the medical field to store massive information of various patients. Moreover, major concern lies with proper diagnosis from the observations. To assist the doctors in this regard, clinical image analysis is highly important. Hence, video stabilization is suggested before performing video compression in the medical arena for better subjective observations. The implementation results indicate that the average PSNR and SSIM of video stabilization with MPEG-2 to be 33.59 dB and 0.73 and for MPEG-2 to be 32.34 dB and 0.67. A superiority of **1.25 dB PSNR** and **0.06 SSIM** is observed while performing video stabilization with MPEG-2. Hence, to aim for more exact diagnosis, research in the field of compression can be performed in alignment with video stabilization.

Acknowledgements We sincerely thank Karunya Institute of technology and sciences for providing us with all the infrastructure and research facilities.

References

1. Rawat P, Singhai J (2011) Review of motion estimation and video stabilization techniques for hand held mobile video. Signal Image Process Int J (SIPIJ) 2(2)
2. Tico M, Vehvilainen M (2007) Robust method of video stabilization. In: 15th European signal processing conference (EUSIPCO 2007), Poznan, Poland, 3–7 Sept 2007
3. Kulkarni S, Bormane DS, Nalbalwar SL (2017) Video stabilization using feature point matching. J Phys Conf Ser 787:012017
4. Chang H-C, Lai S-H, Lu K-R (2004) Robust and efficient video stabilization algorithm. 0-7803-8603-5/04/$20.00 ©2004 IEEE
5. Somasekhar Reddy K, Lokesha H, Akhila S (2014) A survey on video stabilization algorithms. Int J Adv Inf Sci Technol (IJAIST) 31(31). ISSN: 2319:2682

6. Raveena Judie Dolly D, Josemin Bala G, Dinesh Peter J (2017) Performance enhanced spatial video compression using global affine frame reconstruction. J Comput Sci 18:1–11
7. Raveena Judie Dolly D, Josemin Bala G, Dinesh Peter J (2017) A hybrid tactic model intended for video compression using global affine motion and local free-form transformation parameters. 1–15
8. Kulkarni SM, Bormane DS, Nalbalwar SL (2014) Coding of video sequences using block matching motion estimation three step search algorithms. Int J Innov Res Electr Electron Instrum Control Eng 2(6):651–1655
9. Sriram B, Eswar Reddy M, Subha Varier G (2013) Study of various motion estimation algorithms for video compression/coding standards & implementation of an optimal algorithm in LabVIEW. Int J Emerg Technol Adv Eng 3(4):372–381
10. Jain R, Baraskar T (2014) The comparative analysis of various approaches used in motion estimation. Int J Adv Found Res Comput (IJAFRC) 1(12):1–9

Significance of Global Vectors Representation in Protein Sequences Analysis

Anon George, H. B. Barathi Ganesh, M. Anand Kumar and K. P. Soman

Abstract Understanding the meaning of protein sequences is tedious with human efforts alone. Through this work, we experiment an NLP technique to extract features and give appropriate representation for the protein sequences. In this paper, we have used GloVe representation for the same. A dataset named Swiss-Prot has been incorporated into this work. We were able to create a representation that has comparable ability to understand the semantics of protein sequences compared to the existing ones. We have analyzed the performance of representation by the classification of different protein families in the Swiss-Prot dataset using machine learning technique. The analysis done by us proved the significance of this representation.

1 Introduction

Finding the meaning of biological patterns are always difficult due to the large data it contains. Nature's language is difficult to understand for an ordinary human. The progress in artificial intelligence had brought humans the ability to create algorithms that can interpret the language in a better way. Use of NLP techniques claims to be a solution to those problems in understanding. Labeling the data is difficult due to a large amount of data available in it, and this can be considered as the main motivation for moving toward the unsupervised approaches rather than the supervised ones.

A. George (✉) · H. B. Barathi Ganesh · M. Anand Kumar · K. P. Soman
Amrita School of Engineering, Center for Computational Engineering and Networking (CEN), Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: anongeorge007@gmail.com

H. B. Barathi Ganesh
e-mail: barathiganesh.hb@arnekt.com

M. Anand Kumar
e-mail: m_anandkumar@cb.amrita.edu

H. B. Barathi Ganesh
Arnekt Solution Pvt. Ltd., Pentagon P-3, Magarpatta, Pune, Maharashtra, India

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_27

There exist different amino acids which combine together in a specific way to form polypeptide chains. These chains are structurally arranged to form proteins. Asgari and Mofrad [2] observed that even though the 3D structure is important in forming the characteristics, linear sequences are sufficient enough for the creation of protein representations.

Extracting the good features from a protein sequence is more important than changing the classifiers. A good set of features will provide better results when combined with appropriate classifiers. In this work, we aim to create a vector representation that contains good features for the further classification works. Existing vector representation was done with a skip-gram model using 546,790 sequences of Swiss-Prot database which has created promising result [2]. Here, we are trying to create n-dimensional vectors with widely used NLP technique called GloVe. The proposed model is evaluated using SVM classifier. Different models with varying hyperparameters are made and the evaluation is done by classification to obtain the best GloVe model. The GloVe model with selected hyperparameter needs to be trained only once and it can be used for various purposes including family classification, structure prediction, and interaction predictions.

2 Related Works

Asgari and Mofrad [2] have created Prot-Vec and Gene-Vec for protein sequences and gene sequences, respectively. They have used Word2vec models like skip-gram to represent the protein sequences. They have divided the sequences into n-gram model which consider three letters occurring together as one word. They have used distributed representation of words to obtain the semantics. They have classified proteins into different families as well as structured and unstructured proteins very efficiently. They have used training data as 546,790 sequences of Swiss-Prot database for their model and used it for the classification of 7027 protein families of 324,018 protein sequences in Swiss-Prot that resulted in an accuracy of $93\% \pm 0.06$ [2].

Barathi Ganesh et al. [4, 7] identified that distributional representation has made better results compared to the frequency-based text representation methods.

Searls [15] proposed that DB interweaves the field of linguistic metaphor for extending their knowledge of understanding about the language of life. In this work, the author has described the relationship between human genome and these metaphors by considering the notion of DNA as the language of the genome. Motomura et al. in [12] have treated amino acid protein sequence as a stack of English words for extracting the structural sequence information of protein sequence.

In [10], they tried deep learning methods with longitudinal probability densities for identifying the phenotypic patterns from the clinical data.

In traditional methods, people had used supervised learning, whereas in this approach they preferred unsupervised learning model.

Bork et al. [5] untangled the complexity in mapping the phenotype and genotype. For this goal, they had used information from expression data, proteomics, and also from fully sequenced genomes. Cai et al. in [6] used SVM for classifying protein into its respective functional families. Here, they have developed a new software named SVMPRO which was trained with a large number of proteins with their functional families and seed proteins. Huynen et al. in [9] have conducted a comparative study of proteins based on their type of genes. Here, they have chosen the genes by considering the genomic context. Murzin et al. [13] have developed a protein database which contains the information about different protein structures. It provides exhaustive and detailed information about the proteins and accounts the relationship between the protein structures. Ando et al. in [1] have developed a method which helps to identify the potentially disordered proteins between the distinctly allied species. Usage of pretrained GloVe for the emotion detection tasks in [8] proved that it can understand the semantics of the language. Barathi Ganesh et al. [3] used SVM classifier for text classification after representing the words as vectors which gave better classification accuracy.

3 Global Vectors

Global Vectors (GloVe) is a type of unsupervised machine learning technique used to obtain meaningful vectors corresponding to each word in a corpus [14]. Use of word–word co-occurrence while forming the vector representation improves the model rather than observing a window of surrounding words in the database as done in skip-gram and CBOW models before. Skip-gram uses a word to predict the nearby words, thereby adjusts the weights in the shallow neural network producing word to vector representation. CBOW uses the surrounding words to foresee the specific word, then the weights are adjusted as mentioned in the Skip-gram model. Feature extraction is done in a better way by utilizing these methods. Two main types of vector representations include global matrix factorization methods and local context window methods, such as the skip-gram model [11]. Global vector representation combines the best of those two models to create a new one. GloVe keeps up the statistical information present in the corpus. More information is captured by the word–word co-occurrence count on the complete set of words that are used for training.

$$\hat{C} = \sum_{i,j} g\left(X_{ij}\right)\left(w_i^T \tilde{w}_j - log X_{ij}\right)^2 \tag{1}$$

In Eq. 1, $\hat{C}$ is the cost function and X is the matrix that has the word–word co-occurrence counts. The number of times of occurrence of the word j appears near the word i is represented as X_{ij}. Here, $w \in \mathbb{R}^d$ are the word vectors and $\tilde{w} \in \mathbb{R}^d$ are separate word vectors. A weighting function $g\left(X_{ij}\right)$ is used in the cost function [14].

Table 1 Family classification metadata

Swiss-Prot-accession ID	Long ID	Protein name	Family ID	Family description
Q6GZX4	001R_FRG3G	Putative transcription factor 001R	Pox_VLTF3	Poxvirus late transcription factor VLTF3 like
Q6GZX3	002L_FRG3G	Uncharacterized protein 002L	DUF230	Poxvirus proteins of unknown function
Q6GZX0	005R_FRG3G	Uncharacterized protein 005R	US22	US22 like
Q91G88	006L_IIV6	Putative KilA-N domain-containing protein 006L	DUF3627	Protein of unknown function (DUF3627)
Q197F3	007R_IIV3	Uncharacterized protein 007R	DUF2738	Protein of unknown function (DUF2738)

4 Datasets

Dataset [2] consists of 324,018 protein sequences and their corresponding 7027 protein families. Sample family classification metadata is given in Table 1. The dataset also contains protein sequences for each of the 324,018 proteins. Protein families provided in the dataset are biased and most of the families contain only few protein sequences. Global vectors are created using the entire sequences in the dataset. Vectors of every sequence are also made using the GloVe created and that can be used for protein family classification. The most occurring protein families are alone considered for the classification purpose.

5 Methodology

Creating a vector representation of protein sequence that starts with the collection of the dataset and preprocessing steps. The preprocessing steps consist of getting all the sequences of the entire proteins in the dataset. These sequences are split into three letter words called as 3-g. Three different sets of words are created for each sequence as in [2]. These entire words are saved into a text file which is considered as the input for training the GloVe. After training GloVe model, corresponding vectors are obtained for all the words present in the corpus. Vectors corresponding to each sequence in the dataset is made from the obtained GloVe. For obtaining the vectors,

we first split the sequence into 3-g and vectors from the GloVe model is taken for the corresponding words in the sequences. All the vectors corresponding to each word are taken as an element-wise sum to give a vector for each sequence.

Hyperparameters changed for GloVe creation are vector dimension, window size, and minimum vocabulary count. The experiment started by arbitrarily fixing the window size and minimum vocabulary count as 4 and 5, respectively, and changing the vector dimension from 100 to 1000 with an increase of 100 dimensions. The accuracy of most occurring 10 class classification using the "one versus rest" SVM classifier was calculated for all the cases. Vector dimension that gave the highest accuracy is chosen and repeated the experiment by varying window sizes from 3 to 10. Classification accuracy is calculated as before and the best window size is chosen. Minimum vocabulary size is varied from numbers 1 to 5 and the best is selected.

After creating the vector for each protein sequence using the best hyperparameters, further evaluation is done by classification for a different number of classes. The data is split into 80% training and 20% testing data and testing accuracy is noted. Process flow diagram for the proposed model is as shown in Fig. 1. The protein names and its family details are utilized for classification [2]. SVM classification accuracy is calculated for the evaluation of the created model.

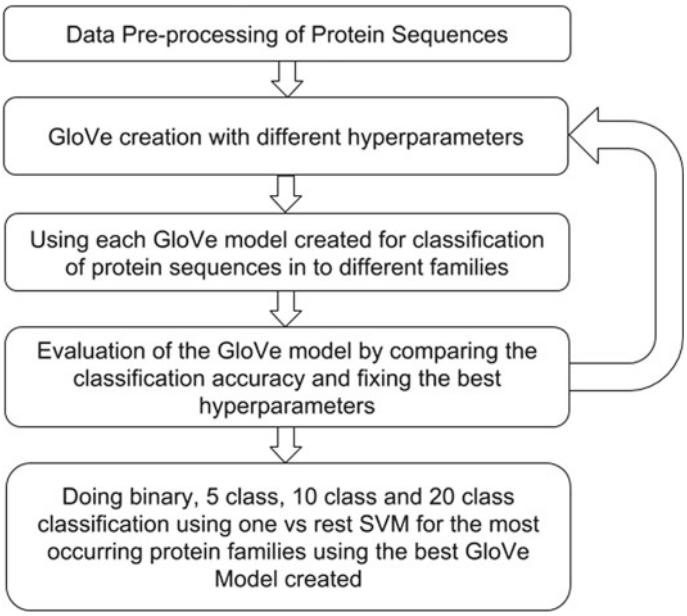


Fig. 1 Process flow diagram of methodology

5.1 Support Vector Machine (SVM)

Support vector machine is used commonly in classification problems as it creates decision boundary based on the uttermost data. This borderline is called hyperplane. Hyperplane has one dimension less than the dimension of original data space. Different classes of data are kept apart by this popular SVM classifier. Decision boundary is found out optimally so that it gives better accuracy for new unseen data [16]. Support vectors are found in such a way that they push apart the nearby classes for maximum separation. We have used one versus rest SVM for multiclass classification.

6 Results and Observation

Classification using the GloVe model gave better results than the existing algorithms. Protein families which consist of most numbers of proteins are considered for classification and evaluation. Using SVM classifier for classification of top 2, 5, 10, and 20 protein families resulted in an accuracy of 97.89%, 96.01%, 92.6%, and 90.86%, respectively, for the 20% test set. Statistics indicate that the semantics of protein sequences are well captured by our model. Obtained results for our system are depicted in Fig. 2.

It has taken around half an hour for the creation of GloVe with 15 iterations when executed in Intel i7 processor with 16 GB RAM. Since the difference in observed results is negligibly less for the classification of 15 and 100 iterations, we have chosen 15 as the primary iteration count for the entire experiment.

Amount of computational power required for creating the GloVe was getting higher as the vector size increased from 100 to 1000. Classification accuracy using the GloVe model with different vector dimensions and a constant window size of 4 and vocabulary minimum count of 5 is shown in Fig. 3. Preprocessing of the obtained vectors for classification took more memory due to the huge amount of protein sequences data. The 100-dimensional vector consumed nearly 3 GB RAM for the preprocessing and arranging the data for classification. It was increased almost linearly till the 1000-dimensional vector consumed nearly 13 GB of RAM.

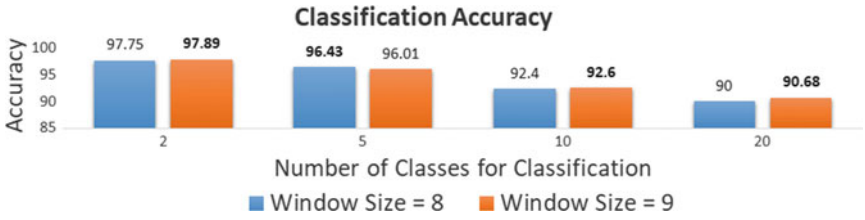


Fig. 2 Accuracy of classification for different number of classes for our best models with window sizes 8 and 9

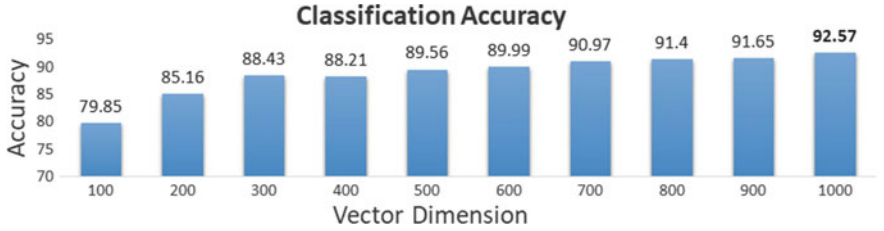


Fig. 3 Variation of accuracy with the changes in vector dimension

The time required for the classification using 100–1000 dimension vector increased as the number of features were high. They produced better results even though it took more time for the computation. Changing the window sizes, which is the number of words to be considered as neighbors, while the creation of GloVe also gave significant changes in accuracy when the vector dimension was lower. But it has not produced many differences with higher dimensional vectors. Results of classification accuracy for different window sizes, constant vector dimension of 10, and vocabulary minimum count of 5 are depicted in Table 2. Classification accuracy for different vocabulary minimum counts is illustrated in Table 3.

The addition of a 3-g model into the vocabulary proceeds only when it repeats a specific number of times. Here, this vocabulary minimum count is varied from 1 to 5. This is to see the difference in performance when the vocabulary size changes. The

Table 2 Evaluation of accuracy for different window sizes

Window size	SVM classification accuracy	
	100D	1000D
3	78.56	91.32
4	79.57	91.75
5	81.13	91.99
6	81.06	91.88
7	79.92	91.86
8	**82.99**	**92.4**
9	**81.43**	**92.6**
10	80.68	92.34

Table 3 Evaluation of accuracy for different vocabulary minimum counts

VOCAB_MIN_COUNT	SVM classification accuracy
1	92.55
2	92.49
3	92.43
4	92.56
5	**92.6**

experiment indicates that the difference observed in accuracy is negligible. Finally, we have created a GloVe representation by finding the best hyperparameters through experiments that is able to understand the semantics of the protein sequences and able to classify multiple protein families with high accuracy.

7 Conclusion and Future Work

The widely accepted NLP technique, GloVe, for the protein sequence classification produced promising results. Classification can be done for any number of families available in the dataset even though the accuracy reduces as we increase the number of classes. Dataset is biased and most of the families contain only few protein sequences which cannot be split into training and testing for getting a good accuracy value.

The model that we have built can be used for many applications like missing protein sequences prediction, classification of FG-Nup sequences from structured protein in the protein data bank, and also it is applicable in the separation of unstructured DisProt sequences from structured kind.

The GloVe model created can be used to classify a number of protein families using the same method that we followed for 2, 5, 10, and 20 classes. Different machine learning algorithms like logistic regression, naive Bayes, kNN, K-Means, and gradient boosting algorithms can be used for improving the classification accuracy for the different protein sequences. Recent improvements in deep learning techniques have proven that algorithms like CNN, RNN, and LSTM will produce better results for classification and sequence prediction and it can enhance our model accuracy.

References

1. Ando D, Colvin M, Rexach M, Gopinathan A (2013) Physical motif clustering within intrinsically disordered nucleoporin sequences reveals universal functional features. PloS One 8(9):e73,831
2. Asgari E, Mofrad MR (2015) Continuous distributed representation of biological sequences for deep proteomics and genomics. PloS One 10(11):e0141,287
3. Balakrishnan BGH, Vinayakumar AKM, Padannayil SK. Nlp cen amrita@ smm4h: health care text classification through class embeddings
4. Barathi Ganesh H, Anand Kumar M, Soman K (2016) Distributional semantic representation in health care text classification. CEUR 1737
5. Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y (1998) Predicting function: from genes to genomes and back1. J Mol Biol 283(4):707–725
6. Cai C, Han L, Ji ZL, Chen X, Chen YZ (2003) Svm-prot: web-based support vector machine software for functional classification of a protein from its primary sequence. Nucleic Acids Res 31(13):3692–3697
7. Ganesh HB, Kumar MA, Soman K (2016) From vector space models to vector space models of semantics. In: Forum for information retrieval evaluation. Springer, Berlin, pp 50–60

8. George A, Soman K et al (2018) Teamcen at semeval-2018 task 1: global vectors representation in emotion detection. In: Proceedings of the 12th international workshop on semantic evaluation, pp 334–338
9. Huynen M, Snel B, Lathe W, Bork P (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res 10(8):1204–1210
10. Lasko TA, Denny JC, Levy MA (2013) Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. PloS One 8(6):e66,341
11. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119
12. Motomura K, Fujita T, Tsutsumi M, Kikuzato S, Nakamura M, Otaki JM (2012) Word decoding of protein amino acid sequences with availability analysis: a linguistic approach. PloS One 7(11):e50,039
13. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) Scop: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247(4):536–540
14. Pennington J, Socher R, Manning C (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543
15. Searls DB (2002) The language of genes. Nature 420(6912):211
16. Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300

Texture Analysis on Thyroid Ultrasound Images for the Classification of Hashimoto Thyroiditis

S. Kohila and G. Sankara Malliga

Abstract The biopsy using Fine Needle Aspiration (FNA) is a major procedure/testing method which has been regularly recommended, at a time, exactly when a thyroid nodule is suspected or else identified. The FNA will usually reveal if a nodule is benign or malignant. Histopathology is also sometimes recommended. Another regular test is the ultrasound. Yet, the ultrasound cannot recognize or distinguish the thyroid disorders. Hashimoto thyroiditis is the most widely recognized kind of inflammation of the thyroid gland. The motto of this work is to identify the Hashimoto's thyroiditis disorder using only ultrasonogram images without going for any painful examination. In this paper, features are studied using the Neighborhood Gray Tone Difference Matrix (NGTDM), Statistical Feature Matrix (SFM), and Laws' texture energy measures methods. The salient features from the above procedures are helpful to identify and in separating the two types of ultrasonic thyroid images as normal and Hashimoto's thyroiditis. The student two-tailed unpaired T-test method is employed to classify the two groups. A major difference between the two groups ($p < 0.001$) was observed. The results are correlated with the histopathology results. The results prove that the Hashimoto thyroiditis can be identified using the ultrasound images.

Keywords Texture analysis · T-test · Hashimoto thyroiditis · NGTDM · SFM Laws' texture energy measures · Thyroiditis

S. Kohila (✉)
Department of Electronics and Communication Engineering,
Sri Sairam Engineering College, Chennai, India
e-mail: kohila.ece@sairam.edu.in

S. Kohila
VELS University, Chennai, India

G. Sankara Malliga
Department of Electronics and Communication Engineering,
Anand Institute of Higher Technology, Chennai, India
e-mail: sankaramalliga@gmail.com

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_28

1 Introduction

The thyroid gland is a kind of endocrine gland which is bigger like many types. It will be found in neck region, underneath the thyroid cartilage. Common thyroid disorders are hyperthyroidism, hypothyroidism, thyroiditis, and thyroid nodules. In this study, an application of a technique is suggested which is used for the automatic findings of Hashimoto's thyroiditis using only the ultrasonogram images [1].

Thyroiditis is an inflammation of the thyroid organ or gland that can emerge due to a variety of reasons. There are two sorts of thyroiditis where at first hyperthyroidism presents that is trailed before the occurring time matched to hypothyroidism. There exist postpartum thyroiditis and Hashimoto's thyroiditis.

Generally, thyroid issues can be analyzed from clinical examinations utilizing immunological, metabolic, and also hormonal investigations of examining blood as well as cytological testing utilizing the FNA biopsy. Nevertheless, the commonly utilized demonstrative and observing tool for the thyroid issues happens to be the ultrasonogram imaging.

Nowadays, through clinical methods like histopathology examination and via FNA cytology, the swelling in thyroid nodule has been determined. Despite the fact that the consequences of FNA cytology and histopathological tests are exact, they have disadvantages as arrangements to provide anesthesia, hospital admission, and needle inclusion. But the ultrasonography is a noninvasive technique. It also has the uniqueness like real-time imaging, low cost, ableness to analyze different dimensions and the effect of volume, mobile nature, and the quality of not having side impact. Thus, the ultrasonogram imaging has become the widely utilized strategy in imaging for examining and executing follow up actions for thyroid-related disorders.

Just a couple of studies have concentrated on creating computational ways to deal with thyroid features analysis by ultrasound image examination. Image intensity details are utilized with a view to have the recognizable proof of the disorder of thyroid Hashimoto's [2], with a view to mark the identification of lesions in nodular thyroid [3], and also with a view related to tumor characterization [4]. The textural information extracted from the image using the co-occurrence matrix characteristics [5] are used to mark recognizable proof for identifying the chronic inflammations of thyroid gland [6] as well as for segregation process among the very usual and pathologic tissues [7]. Texture features using GLCM and GRLM for the ultrasonogram images of Hashimoto thyroiditis are discussed [8].

The evaluation of sonographic identification of deeply chronic inflamed thyroid tissue is much related to examiner's understanding and experience. This paper shows that correct assessment is possible that inflamed and healthy tissues can be distinguished by the methods explained. This work is an initial step to quantitative investigation of the chosen area of vital importance. A smaller number of texture features computed from the ultrasonogram images proved that it is possible to achieve high classification rate to discriminate the Hashimoto's thyroiditis and healthy tissue.

In this paper, the attributes to classify the Hashimoto's thyroiditis and normal images of ultrasound images are focused. They are currently observed using either

FNAC or histological methods. The different features from the ultrasound images are extracted using the methods NGTDM, Laws' texture energy measures, and statistical feature matrix texture analysis. The student two-tailed unpaired T-Test is used for the analysis of the extracted features. Significant differences are observed between the two groups normal and Hashimoto's thyroiditis.

This presentation is structured like below. The materials and methods have been discussed in Sect. 2. Texture analysis using NGTDM and the statistical feature matrix is presented in Sects. 3 and 4, respectively. Section 5 deals Laws' texture energy measures. Results are discussed in penultimate Sect. 6, and the last Sect. 7 has the summation with conclusions.

2 Materials and Methods

2.1 Data Acquisition

To begin with, the ultrasound images of 14 patients affected by Hashimoto's thyroiditis and simultaneously 25 people with non-complaints related to thyroid were accumulated from Bharat Scan Centre, Royapettah, Tamil Nadu. These images were analyzed and the findings of Hashimoto's thyroiditis had been confirmed by means of the reports of histopathology.

An ultrasound image of a normal thyroid gland (without nodules) is exhibited in Fig. 1. The automatic detection and segmentation of thyroid images with and without nodule have been complex and the segmentation processing has not been in the observation of this paper. A tool which is interacting is utilized. The boundaries of nodules are generally marked by the physician as illustrated by the outline in Fig. 2. Number of M $\times$ M non-overlapping images are manually cropped and selected as samples. The features have been extracted from the accumulated images using the aforementioned methods like, Neighborhood Gray Tone Difference Matrix (NGTDM), texture analysis using statistical feature matrix, and Laws' texture energy measures techniques. Significant features were selected using student T-test.

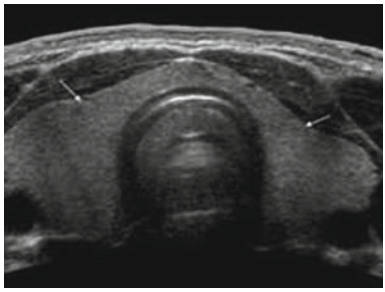


Fig. 1 Ultrasound image of thyroid gland (without nodules)

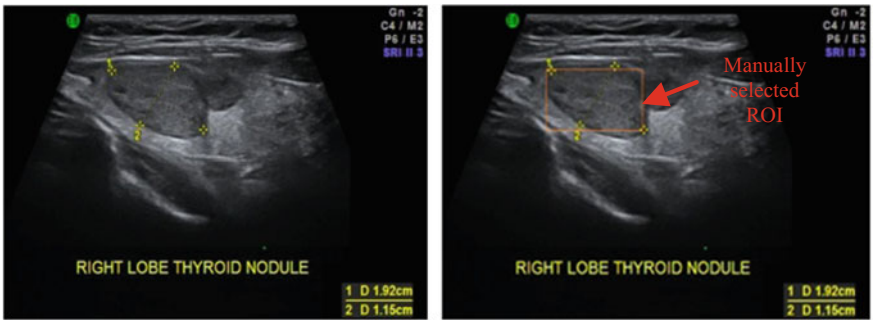


Fig. 2 Ultrasound image of thyroid gland (with nodules)

3 Neighborhood Gray Tone Difference Matrix (NGTDM)

The properties that are used to separate between different textures include complexity, coarseness, contrast, busyness (fineness), directionality, shape, and texture strength. Therefore, for the general texture measures and performance analysis in automatic texture classification, the aforementioned textural properties are included.

The visual properties which are characterized using coarseness, busyness, complexity, contrast, and texture strength are considered as very vital texture descriptions in view of concepts by people in a broader way [9]. These properties were expressed by equations using the neighborhood gray tone difference matrix technique.

Coarseness refers to the roughness of the image and related to the small number of large primitives. The computational measure for coarseness property is given in Eq. 1

$$f_{\text{coar}} = \left[\varepsilon' + \sum_{i}^{G_m} p_i + C(i) \right]^{-1} \tag{1}$$

where G_m is the maximum gray level present in the image and ε' is a constant that prevents f_{coar} attaining infinite. For an M × N image, P_i is the probability of occurrence of intensity value i, and is given by Eq. 2

$$\mathrm{P_i} = \mathrm{N_i}/\text{Total number of pixels} \tag{2}$$

The f_{coar} is the coarseness.

Contrast means the distinction of the things which are brought for comparison. The computational form for this property is given in Eq. 3

$$f_{\text{con}} = \left[\frac{1}{N_g(N_g-1) \sum_{i=0}^{G_n} \sum_{j=0}^{G_n} P_i P_j (i-j)^2} \right] \left[\frac{1}{n^2 \sum_{i}^{G_n} S(i)} \right] \tag{3}$$

where N_g is the number of different intensity levels present in the image and can be expressed as in Eq. (4)

$$N_g = \sum_{i}^{G_k} Q_i \tag{4}$$

where

$$Q_i = \begin{cases} 1, \text{ if } P_i \neq 0 \\ 0, \text{ otherwise} \end{cases}$$

Busyness is appearing to be actively engaged in an activity. The computational measure is given in Eq. 5

$$f_{\text{bus}} = \sum_{i}^{G_n} P_i S(i) / \sum_{i}^{G_n} \sum_{i}^{G_n} i P_i - j P_j \tag{5}$$

where $Pi \neq 0$, $Pj \neq 0$

Complexity refers to the quality of the image and useful in determining the compression techniques. The computational measure is given in Eq. 6

$$f_{\text{comp}} \sum_{i}^{G_n} \sum_{j}^{G_n} \{(|i - j|)/(n^2(P_i + P_j))\}\{P_i S(i) + P_j S(j)\} \tag{6}$$

where $Pi \neq 0$, $Pj \neq 0$

The feature strength is expressed in Eq. 7

$$f_{str} = \left[\sum_{i}^{G_n} \sum_{j}^{G_n} (P_i + P_j)(i - j)^2 \right] / \left[+ \sum_{i}^{G_n} S(i) \right] \tag{7}$$

where $Pi \neq 0$, $Pj \neq 0$.

Among these five NGTDM features, the four features coarseness, complexity, contrast, and strength are employed in the work to point out various features from thyroid ultrasound images of Hashimoto's thyroiditis.

4 Statistical Feature Matrix (SFM)

The visual textures suggested by the author [10, 11] are extracted and analyzed. SFM measures the statistical properties of two pixels within an image at various distances. It depends on the size of the matrix rather than the gray value. The following features are extracted using SFM:

Coarseness	When there is a difference, only in scale between the two patterns, the magnified one is coarser.
Contrast	Image contrast is either shrinking or stretching the intensity scale.
Periodicity	It is due to the regularity of repetitive patterns.
Roughness	It occurs because of mock energy changes in gray levels.

The above features are extracted and analyzed.

5 Laws' Texture Energy Measures

To measure the amount of variation in the image a texture energy approach was developed by Laws [12, 13]. The energy depends on fixed size window. To extract energy from the normal and abnormal images, local averaging, edge and spot mask of size 1×3 are used. A center weighted local averaging mask, $L = (1, 2, 1)$, an edge mask $E = (-1, 0, 1)$, and spot detection mask $S = (-1, 2, -1)$ are applied to obtain texture energy from the images. By convolving L, E, and S masks with each other the 3×3 mask kernel of LL, EE, SS, LS, LE, and ES are obtained. Further, 3×3 mask, LL, EE, SS, LE, SE, and LS are convolved with images and texture energy is extracted. The extracted texture feature is analyzed to differentiate the normal and Hashimoto thyroiditis tissues.

6 Results

A sum of 585 salient feature values is extracted using MATLAB coding. 15 features for each of 39 samples had been computed, extracted, and the above features are utilized for the purpose of analysis. Student T–test was conducted between the ultrasound images of Hashimoto's thyroiditis and normal thyroid ultrasound images and is presented in Table 1. The features that express significant differences between two groups are explicitly shown in Table 2. The significant features that distinguish the two groups are selected based on the T-test results. They are selected with a greater significance of $p < 0.005$. Depending on the T-test analysis, the features coarseness, contrast, complexity, strength, periodicity, roughness and the Laws' TEM features can be utilized to distinguish the normal and Hashimoto's thyroiditis using ultrasonogram images. The varied observations between the two groups for different p-values ($p < 0.05$, $p < 0.005$, and $p < 0.0005$) were presented in Table 2.

The evaluation of inflamed chronic thyroid using ultrasound images in medical practice is based on experience of the examiner or medical practitioner. But, the literature studies show that the correct assessment is possible, and the normal and inflamed thyroid tissues can be differentiated by texture analysis [1–8]. Similarly, the proposed texture features extracted from ultrasonogram images were also shown higher significance toward classification abnormality in the thyroid glands. Acharya

Table 1 T-test analyses on the features extracted

Texture features	T-test	*p*-value B/N normal and Hasi
NGTDM	Coarseness	0.00170
	Contrast	0.00164
	Busyness	0.02860
	Complexity	0.00000
	Strength	0.00000
SFM	Coarseness	0.00042
	Contrast	0.00000
	Periodicity	0.00000
	Roughness	0.00000
Laws' TEM	LL	0.00000
	EE	0.00000
	SS	0.00000
	LE	0.00002
	ES	0.00000
	LS	0.00001

Table 2 Variation between the groups for different *p*-values ($p < 0.05$, $p < 0.005$, and $p < 0.0005$)

Variables/Groups	Normal (n = 25)	Hasi (n = 14)	*p*-value (a)
Age	44.72 ± 13.24	35.57 ± 12.14	0.069
Coarseness	59.11 ± 16.92	42.86 ± 7.77	0.00170
Contrast	0.04 ± 0.02	0.09 ± 0.08	0.00164
Busyness	0.02 ± 0.04	0.07 ± 0.02	0.02860
Complexity	11836.66 ± 18745.56	63402.29 ± 30811.92	0.00000
Strength	495803.11 ± 298823.43	1409014.74 ± 558705.02	0.00000
Coarseness	14.51 ± 4.05	9.96 ± 2.23	0.00042
Contrast	6.16 ± 2.97	12.44 ± 3.32	0.00000
Periodicity	0.76 ± 0.08	0.64 ± 0.03	0.00000
Roughness	2.14 ± 0.05	2.29 ± 0.08	0.00000
LL	64886.34 ± 8808.52	108732.25 ± 27489.09	0.00000
EE	580.48 ± 344.46	1232.45 ± 314.67	0.00000
SS	54.04 ± 49.34	187.25 ± 66.57	0.00000
LE	5025.06 ± 1652.83	7764.23 ± 1741.15	0.00002
ES	166.51 ± 134.93	463.87 ± 134.8	0.00000
LS	1288.11 ± 691.91	2482.28 ± 635.23	0.00001

et al. [1] extracted similar kind of features from ultrasonogram images for classification of Hashimoto's thyroiditis. The vital texture features such as roughness, strength, and morphology were already used in other clinical image analysis application such as classifying osteoporosis from images of X-ray as well as the Computed Tomography (CT) images [14].

7 Conclusion

This paper shows that Hashimoto's thyroiditis can be a different kind from a normal or very usual kind of thyroid gland using the findings through the texture analysis process of ultrasound images. Since this can be evaluated with a number of features (nine), the result confirms that the feature characteristics which are extracted from the ultrasound images of thyroid can be used to characterize the image. Best classification results can be obtained by applying the extracted features as input for a classifier.

References

1. Acharya UR, VinithaSree S, Mookiah MRK, Yantri R, Molinari F, Zieleźnik W, Małyszek-Tumidajewicz J, Stępień B, Bardales RH, Witkowska A, Suri JS (2013) Diagnosis of Hashimoto's thyroiditis in ultrasound using tissue characterization and pixel classification. Proc Inst Mech Eng [H] 227(7):788–798
2. Chang CY, Chen SJ, Tsai MF (2010) Application of support-vector-machine-based method for feature selection and classification of thyroid nodules in ultra sound images. Pattern Recogn 43(10):3494–3506
3. Haralick RM, Shanmugam K (1973) Textural features for image classification. IEEE Trans Syst Man Cybern 6:610–621
4. Tomimori EK, Camargo RY, Bisi H, Medeiros-Neto G (1999) Combined ultrasonographic and cytological studies in the diagnosis of thyroid nodules. Biochimie 81(5):447–452
5. Maroulis DE, Savelonas MA, Iakovidis DK, Karkanis SA, Dimitropoulos N (2007) Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images. IEEE Trans Inf Technol Biomed 11(5):537–543
6. Kousarrizi MN, Seiti F, Teshnehlab M (2012) An experimental comparative study on thyroid disease diagnosis based on feature subset selection and classifiion. Int J Electr Comput Sci IJECS-IJENS 12(01):13–20
7. Tsantis S, Dimitropoulos N, Cavouras D, Nikiforidis G (2009) Morphological and wavelet features towards sonographic thyroid nodules evalution. Comput Med Imag Graph 33(2):91–99
8. Kohila S, Malliga GS (2016) Classification of the thyroiditis based on characteristic sonographic textural features and correlated histopathology results. In: IEEE international conference on signal and image processing (ICSIP). IEEE, pp 305–309
9. Amadasun M, King R (1989) Textural features corresponding to textural properties. IEEE Trans Syst Man Cybern 19(5):1264–1274
10. Wu CM, Chen YC (1992) Statistical feature matrix for texture analysis. CVGIP: Graph Models Image Process 54(5):407–419
11. Chang CY, Tsai MF, Chen SJ (2008) Classification of the thyroid nodules using support vector machines. In: 2008 IEEE international joint conference on neural networks, IJCNN 2008. (IEEE world congress on computational intelligence). IEEE, pp 3093–3098

12. Laws KI (1979) Texture energy measures. In: Proceedings of a workshop on Image understanding workshop, pp 47–51
13. Wu CM, Chen YC, Hsieh KS (1992) Texture features for classification of ultrasonic liver images. IEEE Trans Med Imaging 11(2):141–152
14. Sapthagirivasan V, Anburajan M (2013) Diagnosis of osteoporosis by extraction of trabecular features from hip radiographs using support vector machine: an investigation panorama with DXA. Comput Biol Med 43(11):1910–1919

Cluster Based Paddy Leaf Disease Detection, Classification and Diagnosis in Crop Health Monitoring Unit

A. D. Nidhis, Chandrapati Naga Venkata Pardhu, K. Charishma Reddy and K. Deepa

Abstract Rice is the staple food in the southern parts of India. Its yield directly affects the economy of the state. Rice blast, brown spot, and bacterial blight are the major diseases that can be seen in the leaves of the paddy crop which also greatly affect their productivity. The farmers, with limited knowledge of the disease, add irrelevant pesticides in inappropriate quantities, which not only diminishes the quality of the crop but also leads to the degradation of the soil quality and other environmental degradations. The proposed method uses image processing techniques to detect what disease has affected the leaves and also calculate the severity of the infection by calculating the percentage of the affected area, which can be used in controlling the excessive use of pesticides. The above mentioned diseases are the predominant diseases occurring in paddy crops among other diseases that vastly affect the yield and in turn affect the economic income of the farmers.

Keywords Rice · Leaf diseases · Diagnosis · Image processing · Segmentation

1 Introduction

India is an agricultural country with very slow advancements in agricultural technology. Lack of awareness and modern techniques impact the production and also degrade the ecosystem. Controlled sprinkler system for irrigation with continuous monitoring is the best solution to manage water resources. Appropriate use of pesticides and other minerals only in the quantities required based on the analysis of the plant health and water availability can maintain soil fertility and provide better crop yield.

The objective of the planned system is to develop a user-friendly unit that will certainly ameliorate the productivity of the yield and help the farmers to improve their

A. D. Nidhis · C. N. V. Pardhu · K. C. Reddy · K. Deepa (✉)
Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India
e-mail: k_deepa@blr.amrita.edu

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_29

income. The scope of the suggested application is to advocate a better opinion to the farmers for controlling and making use of precise amount of pesticides which reduce soil contamination and also regulating water usage which helps in water management. The proposed system has enormous significance in the field of automation, precise supervision in plant growth, and disease identification process which uses clustering and feature extraction technology.

A brief description of the collected works is presented in this segment. Islam et al. presented the scheme to quantize the severity of the affected area [1]. This paper talks about the acquisition, storing, and processing of the image from an external source like a camera, and calculating the percentage of the affected area using k-means clustering. But the drawback of this method would be the inability to be able to differentiate among the rice blast, brown spot, and bacterial blight diseases. According to the paper [2], the detection of disease at the primary stage can be done, so that the necessary actions can be taken for minimizing the effect of disease on productivity of the yield. This is done by converting RGB image into gray image and extracting the histogram using MATLAB functions. Narmadha and Arulvadivu [3] have discussed the detection of disease using image processing techniques; in this paper, K-clustered algorithm was used for mitigation of noise and unwanted spots. For recognizing two-dimensional shapes and types in the images [4, 5], image segmentation is used for obtaining regions conforming to individual objects followed with determination of shape factor for recognizing the shape. In paper [6], after the segmentation, mostly green colored pixels are identified based on a particular threshold value that is obtained with the help of Otsu's method. And in the next stage, pixels observed with zero red, green, and blue values are detached from the boundary pixels of the disease-ridden cluster. Histogram matching helps in determining whether the leaf is disease ridden or not. Article [7] discusses the real-time controlling of traffic system by acquiring the image through live video stream, which is then given as input to MATLAB for processing. Article [8] gives a brief insight on using multiclass SVM for disease classification and k-means for segmentation. In paper [9], the procedure for detecting the disease of the plant and classification of diseases is provided. This can be achieved using clustering for segmentation and ANN for cataloguing. With the collected knowledge of all the literature surveys, a better method that incorporates detection and discrimination of the disease and the severity percentage calculations in real time has been proposed, which will be clarified in the forthcoming segments.

2 Suggested Methodology

In this paper, the main focus is to analyze and differentiate the diseases using image processing techniques. The stages of techniques are as follows: [A] image acquisition, [B] image preprocessing, [C] image segmentation, [D] disease detection, [E] transformation to binary image, and [F] calculation of affected percentage area of leaf.

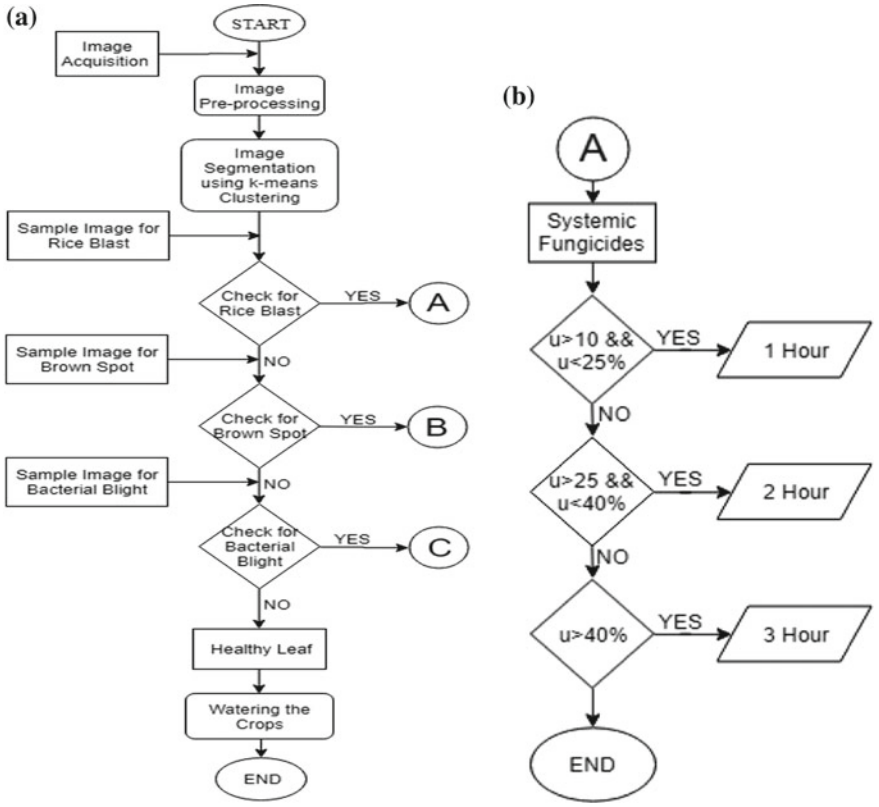


Fig. 1 Flowchart of the entire system

The flowchart of the entire system is given in Fig. 1. The procedure as given in Fig. 1b will be continued for brown spot and bacterial blight. Each of the stages is detailed below.

2.1 Image Acquisition

The process of getting the image from a source is called image acquisition. The source can either be a camera or any other means, and the images obtained from the source are saved in a folder with *.jpg extension.

2.2 Image Preprocessing

The images read from the folder may be of different sizes. Hence, resizing the images to a fixed dimension is mandatory for further processing.

2.3 Image Segmentation

The image is segmented into different regions depending on the intensity of the color. The segmentation is done to segregate different regions of the leaf, hence differentiating the affected area from the unaffected area of the leaf. There are many image segmentation approaches which include thresholding method, color-based segmentation, transform method, and texture method. A better way to obtain an enhanced result from image segmentation is to use algorithm, segmentation tools, and data analysis techniques. In this paper, the image segmentation method considered is k-means clustering [1].

K-means clustering: K-means clustering is a method of vector quantization that is popular for cluster analysis in data mining. This involves the following process: (1) Each cluster's mean is calculated. (2) The remoteness between individual points and every cluster point is premeditated, and its closeness to the cluster average, respectively, is calculated. Allocation of every point to its closest cluster is attained. (3) Steps 1 and 2 are repeated until the summation of the squared Euclidean distance cannot be reduced further. A random point is initially assigned to the cluster. The image is segmented into three cluster mapping, two scripts and the background. The sequence of procedures for the proposed methodology is: acquisition of RGB image, transfiguration of the input image from RGB to gray scale, disguising of the green pixels, deduction of the masked green pixels, segmentation using k-means clustering, and attainment of the useful segments.

2.4 Transformation to Binary Image

The obtained clustered image is then converted into a binary image so that the image contains only two colors (black and white).

2.5 Disease Detection (Feature Extraction)

The disease detection is done through point feature matching. The sequence of process is: reading original image and the sample image, detection of feature points, extraction of feature descriptors, finding of putative points, matching points between

the two images, and calculating of length of surf points and a threshold of 11 matching putative points which are necessary for the infection to be identified as a disease. If matching conditions are satisfied, the disease is recognized and the severity of the disease is calculated, as mentioned in section F. If the conditions are not satisfied, then the sample image of the next disease is taken and steps (a) to (e) are repeated for the next disease.

2.6 Percentage of the Affected Area of Leaf

The percentage of the affected area of the leaf is attained using the given equation: $U = (Ia/(Ia + Iu)) * 100$, where Ia are white pixels of the affected area, Iu are white pixels of the unaffected area, and U is percentage value. From this calculation, we obtain the severity of the disease in terms of affected area percentage, which will act as an input to the automatic treatment unit to dispense the appropriate pesticides.

3 Result and Discussion

To validate the algorithm, four cases are considered, and the simulation results for all the four cases are carried out and presented in this section. The cases considered are as follows: Case I: leaf detected by bacterial blight, Case II: leaf detected by brown spot, Case III: leaf detected by rice blast, and Case IV: healthy leaf.

Case I: Testing of Bacterial Blight

This case study considers a leaf affected by bacterial blight as shown in Fig. 2a. This test image undergoes four stages, and the obtained clusters are as listed in Fig. 2b–e. This test image is compared with the sample image of rice blast. The image processing results for rice blast are depicted in Fig. 2f–h. These images help in calculating the affected area and unaffected area of the first sample image "rice blast" 1. Now, the test image and the sample image are ready for comparison, the image with the points considered are shown in Fig. 2i, j, and the mapping is shown in Fig. 2k. The result from Fig. 2 clearly indicates that the disease is not rice blast, and hence, the program is directed to reiterate the process for the next disease. The same process carried out for Case 1 (rice blast) is used for Case 1 (brown spot) and Case 1 (bacterial blight). The FIGs for the same are in Figs. 3a–d and 4a–d. Case I goes through all the three disease checking as it was not interrupted in between. The result of the analysis was that the disease was identified as "bacterial blight", and the calculation of affected and unaffected was carried out by the MATLAB program.

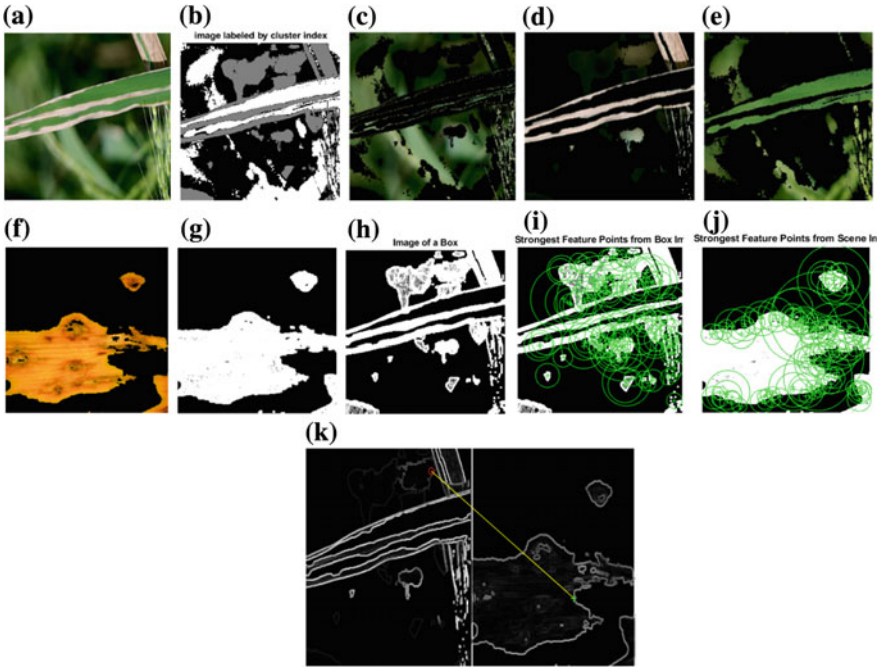


Fig. 2 **a** Bacterial blight test image. **b** Clustered image. **c** Cluster I. **d**. Cluster II. **e** Cluster III. **f** Sample image of rice blast. **g** Gray scale image of sample image. **h** Affected area of the original image taken as box image. **i** 100 strongest points on box image. **j** 300 strongest points on scene image. **k** Matched points

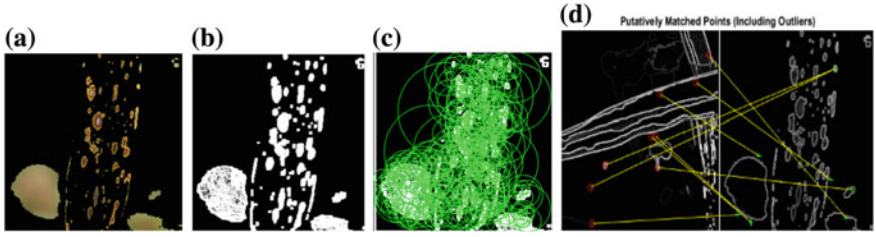


Fig. 3 **a** Sample image of brown spot disease. **b** Gray scale image of sample brown spot image. **c** 300 strongest points on scene image. **d** Matched points

Case II: Brown Spot Disease

The process carried out in Case I, Fig. 2a–k, was reiterated for Case II, and images obtained are from Fig. 5a–h. The process that was done in Case I Fig. 3a–d is repeated in Case II, and results are from Fig. 6a, b.

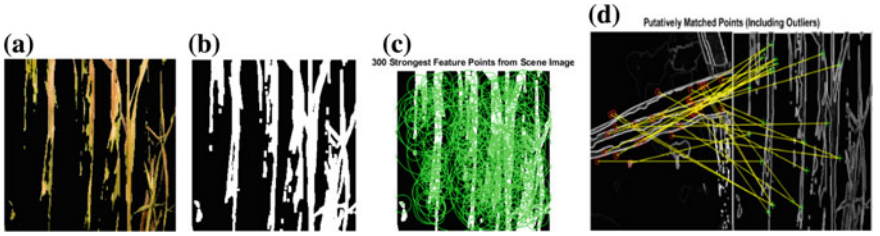


Fig. 4 **a** Sample image of bacterial blight disease. **b** Gray scale image of sample bacterial blight disease. **c** 300 strongest points on scene image. **d** Matched points

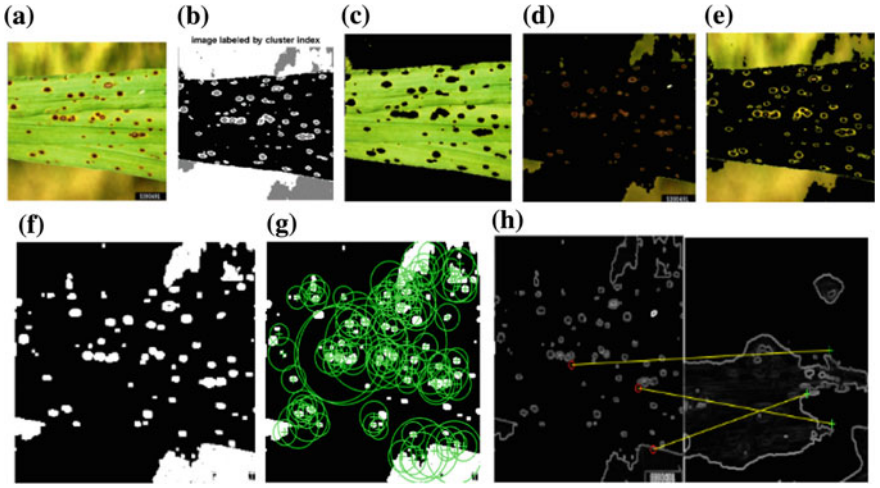


Fig. 5 **a** Test image. **b** Clustered image. **c** Cluster I. **d** Cluster II. **e** Cluster III. **f** Affected area of the original image taken as box image. **g** 100 strongest points on box image. **h** Matched points

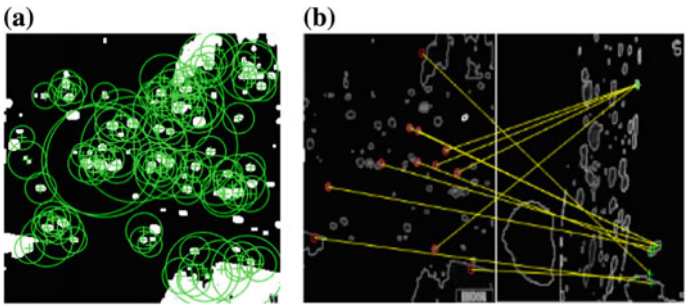


Fig. 6 **a** 100 strongest points on box image. **b** Matched points

Case III: Rice Blast Disease

The same process which was carried out in the Case I Fig. 2a–k is done in Case III, and obtained images are from Fig. 7a–h.

Case IV: Healthy Leaf

The same process which was carried out in Case I Fig. 2a–k is done in Case IV, and obtained images are listed from Fig. 8a–h. The same process that was done in Case I, Fig. 3a–d, is repeated in Case IV, and the obtained image is Fig. 9. The same process that was done in Case I, Fig. 4a–d, is repeated in Case IV, and the obtained image is Fig. 10. The overall result is summarized in Table 1.

As mentioned earlier, the threshold of 11 matching putative points needs to be exceeded in order for it to be classified as a disease. From Table 1, it is clear that the algorithm differentiates the disease correctly. Case IV in Table 1 indicates that the number of matching points is less than the threshold, and hence, it has been concluded

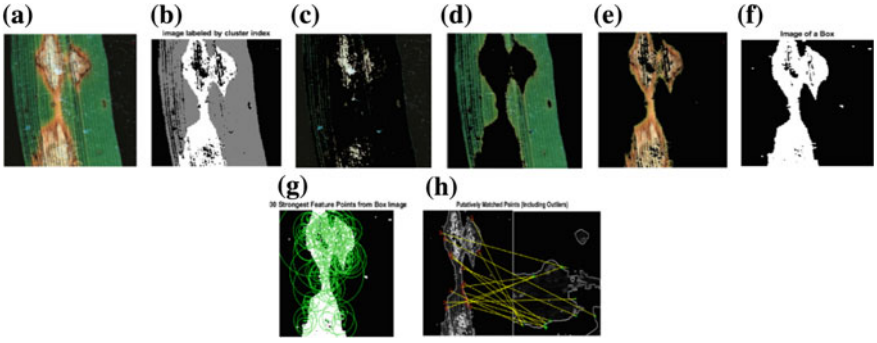


Fig. 7 **a** Test image. **b** Clustered image. **c** Cluster I. **d** Cluster II. **e** Cluster III. **f** Box image is affected image obtained from clustering. **g** 100 strongest points on box image. **h** Matched points

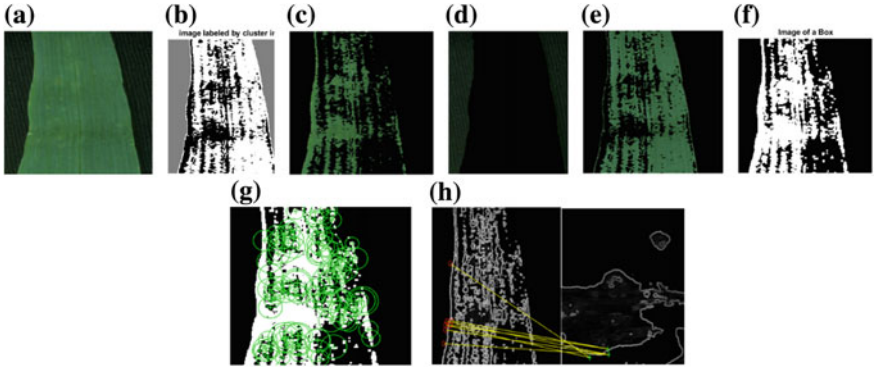


Fig. 8 **a** Test image. **b** Cluster image. **c** Cluster I. **d** Cluster II. **e** Cluster III. **f** Box image is affected image obtained from clustering. **g** 100 strongest points on box image. **h** Matched points

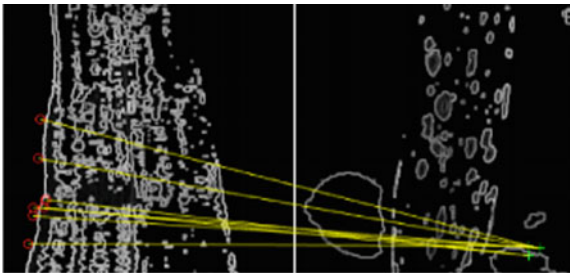


Fig. 9 Matched points for Brown Spot Disease

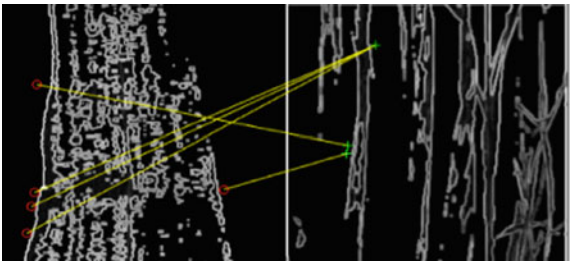


Fig. 10 Matched points for Bacterial Blight Disease

Table 1 Summary of percentage affected area

	Rice blast (no. of matching points)	Brown spot (no. of matching points)	Bacterial blight (no. of matching points)	% of affected area
Case I	1	9	**31**	BP = 15367, BP1 = 18616 u = 45.2197
Case II	3	**12**	4	BP = 6941, BP1 = 35610 u = 16.3122
Case III	**16**	1	10	BP = 12678, BP1 = 29050 u = 30.3825
Case IV	7	7	5	BP = 1354, BP1 = 25544 u = 5.0906

that the crop is not affected. The three diseases considered are rice blast, brown spot, and bacterial blight. The same can be extended to any number of diseases provided the images of the same is uploaded to MATLAB. The proposed image processing scheme is the heart of the PV-based crop health monitoring system shown in Fig. 11 [10]. This system takes the input of the disease detected and the severity percentage from the image processing unit to select the appropriate pesticide and decide the quantity to be automatically dispensed using a microcontroller which controls a solenoid valve to carry out the process. The diseases are identified and the severity of the disease is calculated. The controller initiates the selection of pesticides, as mentioned in Table 2, for the particular leaf disease detected and based on the severity, the duration of the valve is controlled. The time duration for which the valve opens depends on the percentage of the affected area "u" obtained, as explained in Table 2.

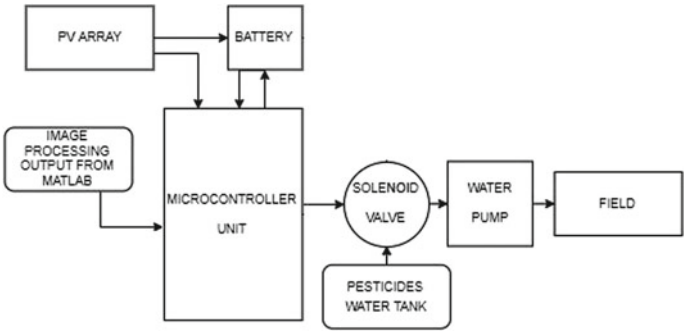


Fig. 11 Block diagram of PV-based crop health diagnosis and treatment unit

Table 2 Pesticides dispensing duration according to the severity

Disease	Pesticides	Severity (%)	Duration (h)
Rice blast	Systemic fungicides, e.g., tricyclazole	10–25	1
		25–40	2
		>40	3
Brown spot	Fungicides like iprodione	10–25	1
		25–40	2
		>40	3
Bacterial blight	Carboxymethyl cellulose (CMC)	10–25	1
		25–40	2
		>40	3
No disease	Water		1

4 Conclusion

The suggested system utilizes leaf features to detect and discriminate between rice blast, brown spot, and bacterial blight diseases by segmenting the infected region. The gray image derived from the RGB image provides clarity on the affected region and is also helpful to extract size, color, proximity, and centroids. These features served as inputs for the classification of three different leaf diseases, which shows accurate classification. The extension of the work will focus on developing algorithms in order to increase the recognition rate of the classification process.

References

1. Islam R, Rafiqul Islam M (2015) An image processing technique to calculate percentage of disease affected pixels of paddy leaf. Int J Comput Appl (0975–8887) 123(12) 2015
2. Mukherjee M, Pal T, Samanta D (2012) Damaged paddy leaf detection using image processing. J Glob Res Comput Sci 3(10)

3. Narmadha RP, Arulvadivu G (2017) Detection and measurement of paddy leaf disease symptoms using image processing. In: International conference on computer communication and informatics (ICCCI-2017), Coimbatore, India
4. Abadi NE, Sadi LA (2013) Automatic detection and recognize different shapes in an image. IJCSI Int J Comput Sci Issues 10(6, no 1)
5. Veni S, Narayanankutty KA (2014) Vision-based hexagonal image processing using Hex-Gabor. Signal Image Video Process 8:317–326
6. Smita N, Niket A (2013) Advances in image processing for detection of plant diseases. Int J Appl Innov Eng Manage (IJAIEM) 2(11)
7. Parthasarathi V, Surya M, Akshay B, Murali Siva K, Shriram Vasudevan K (2015) Smart control of traffic signal system using image processing. Indian J Sci Technol Indian Soc Educ Environ 8(16) (2015)
8. Srunitha K, Bharathi D (2018) Mango leaf unhealthy region detection and classification. Lecture Notes Comput Vis Biomech 28:422–436
9. Anand R, Veni S, Aravinth J (2016) An application of image processing techniques for detection of diseases on brinjal leaves using K-means clustering method. In: IEEE international conference on circuit, power and computing technologies, ICCPCT
10. Nidhis, AD, Venkata Pardhu CN, Charishma Reddy K, Deepa K (2018) Smart crop health diagnosis and treatment unit powered by green fuel. J Green Eng 8(3):389–410

Detection of Glaucoma Using Anterior Segment Optical Coherence Tomography Images

P. Priyanka, V. Norris Juliet and S. Shenbaga Devi

Abstract Glaucoma is one of the major abnormalities in the eye which leads to permanent vision loss over time if left untreated. Prolonged high level of intraocular pressure (IOP) causes optic nerve damage leading to glaucoma. Glaucoma can be detected by analyzing the characteristics of the optic nerve head (ONH) and retinal nerve fiber layer (RNFL). The quantitative analysis of the ocular details obtained by the imaging techniques can help in disease management. The algorithm used here automatically extracts clinical features such as anterior chamber width, iris endpoint width, chamber height, lens vault, angle opening distance, and trabecular iris angle from AS-OCT images to find the condition of eye, namely, the presence or absence of glaucoma.

Keywords AS-OCT · Angle closure glaucoma · Glaucoma · Anterior segment

1 Introduction

Glaucoma is one of the eye diseases that leads to blindness by damaging the optic nerve. The eye continuously produces aqueous humor, and it must drain from the eye to maintain healthy pressure in the eye [1]. In the case of glaucoma, the aqueous humor gets stored inside the anterior segment of the eye which increases the intraocular pressure. Glaucoma initially leads to peripheral vision loss leading to tunnel vision which untreated leads to permanent vision loss. Glaucoma is a condition of struc-

P. Priyanka (✉) · V. Norris Juliet
Department of Electronics and Communication Engineering,
College of Engineering Guindy, Anna University, 600025 Chennai, India
e-mail: priyankapanneer@gmail.com

V. Norris Juliet
e-mail: sweetnorris@gmail.com

S. Shenbaga Devi
Centre for Medical Electronics, Anna University, 600025 Chennai, India
e-mail: s_s_devi@annauniv.edu

J. D. Peter et al. (eds.), *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, Lecture Notes in Computational Vision and Biomechanics 31,
https://doi.org/10.1007/978-3-030-04061-1_30

tural damage of retinal ganglion cells (RGC), and the ability to detect this damage is fundamental in the diagnosis and management of glaucoma [1]. The current standard for diagnosis of glaucoma is gonioscopy which is subjective and requires a trained eye specialist for examination, and also it is not a patient-friendly procedure [2].

Glaucoma can also be diagnosed using fundus image, optical coherence tomography (OCT), and ultrasound biomicroscopy (UBM) of which the anterior segment of the eye is analyzed by anterior segment OCT (AS-OCT) and UBM. Since the UBM has good penetration, it can give the details about ciliary body, zonules, etc. But this procedure also needs an experienced trainer and being a contact procedure, it will create discomfort in patients [3]. Optical coherence tomography is a noncontact procedure which takes less time compared to other procedures and due to this advantage, the OCT image is most widely used in the diagnosis of glaucoma [4].

In any of these imaging procedures, manual evaluation of glaucoma takes a long time, and also, it is subjective being dependent on the experience of the ophthalmologist. To reduce these problems, this paper discusses an algorithm to automatically detect parameters of AS-OCT which helps the ophthalmologist in detecting glaucoma easily.

2 Methodology

The methodology follows the procedure, namely, preprocessing of AS-OCT image, feature extraction followed by classification of the image to find the presence or absence of glaucoma, without any manual intervention. The methodology is explained below.

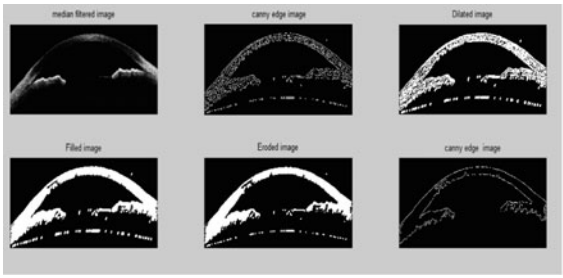
3 Preprocessing

3.1 Filtering

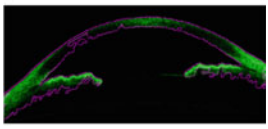
AS-OCT images are given as input to the system, and preprocessing has to be performed on the image to remove noise and enhance the contrast. Here, the median filter is used to remove noise from the images, which helps to improve further processing.

3.2 Edge Detection

Edge detection is an image processing technique for finding the boundaries of an image. Canny operator gives more information about the edges. So canny edge detected image is taken for further processing. Clear edge information is needed for finding the parameters which help in assessing glaucoma. For this purpose, the



(a) Results of pre-processing on input image



(b) Original image with identified edge

Fig. 1 Edge detection

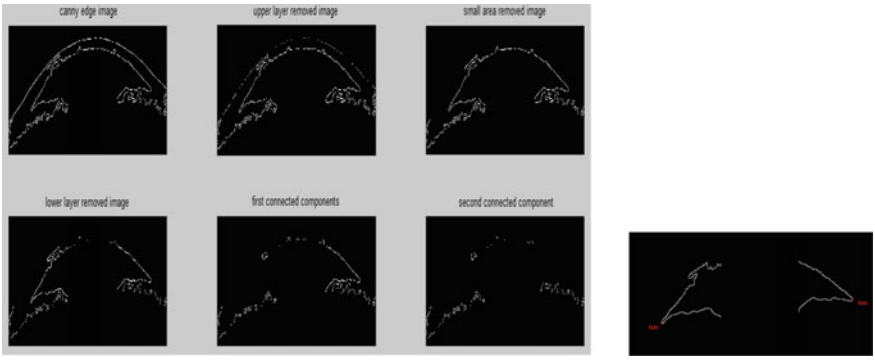
morphological operation is applied on the canny edge detected image. To smoothen the edge, first dilation process is done, and holes are filled on dilated image which makes the intensity even over the region of interest (ROI). Followed by the filling process, erode operation is done on the image which removes the unwanted pixels from the boundary. Finally again, the canny edge detection is done on the eroded image to get enclosed information of the image by its boundaries. Figure 1a, b shows the results of preprocessing steps on input AS-OCT image and the detected edge being overlapped on the original image to see the accuracy of the detected edges.

4 Feature Extraction

After the preprocessing, the clinical features are to be extracted from images which will give the information about the disease. The features, namely, anterior chamber width (ACW), iris endpoint width (IEW), chamber height (CH), lens vault (LV), angle opening distance (AOD), and trabecular iris angle (TIA) help to identify the normal or abnormal anterior segment from AS-OCT images [5].

Scleral spur is used as a reference point for identifying various anterior segment parameters [6], and it is a protrusion of the sclera into the anterior chamber. In a medical image, scleral spur can be identified as a small protrusion near the meeting point of corneal endothelium layer and the iris. This meeting point is called apex [7]. So, scleral spur can be easily identified based on the apex points. Hence, the algorithm is developed to identify the sclera spur which is given below.

After the removal of upper and lower layers from the canny edge detected image, small areas having less than 30 pixels are removed. This is followed by extraction of consecutive two biggest connected components yielding the right and left apex points. The first edge pixels, respectively, from right and left sides are marked as apex points. Figure 2a, b shows the apex finding process and the identified apex point on canny edge detected image of input.



(a) Apex finding process (b) Identified apex points

Fig. 2 Detection of apex

In an AS-OCT image, scleral spur is identified by thresholding the AS-OCT image to identify the intensity variation of sclera and ciliary body. Canny edge detection is done on thresholded image. A circular mask with a radius of twenty pixels is applied on the canny edge detected image with apex point as center to extract the area of interest on both the sides for further process. From the masked image, the consecutive pixels with nonzero intensity along the x-direction are tracked and are marked as scleral spur in extracted edge image. The consecutive pixels with nonzero values represent that there is no intensity variation in the particular direction which represents the scleral spur. Figure 3 shows the process and the identified sclera spur point.

The features, namely, ACW, IEW, CH, LV, AOD, and TIA, are identified using either scleral spur points or the apex points using the algorithms explained below.

4.1 Anterior Chamber Width (ACW)

Anterior chamber width (ACW) is the horizontal distance between two scleral spur (SS) points [5]. Here, ACW is identified by drawing a horizontal line between the two scleral spur points and finding the distance. The distance between the two points is calculated based on Euclidean distance.

4.2 Iris Endpoint Width (IEW)

Iris endpoint width (IEW) is the shortest distance between the two iris end points [5]. After removing the upper layer and small areas from the canny edge detected image, the iris end point is identified. From the center column of the image, the first edge

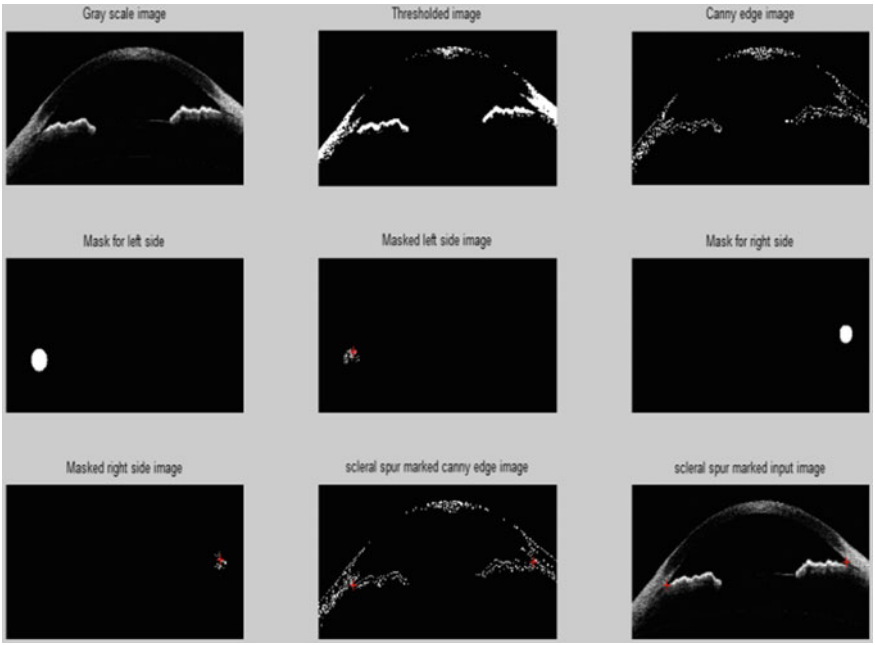


Fig. 3 Identification of scleral spur

point at the left and right sides is identified and marked as iris end points. Horizontal Euclidean distance between the two iris end points gives iris endpoint width (Fig. 4).

4.3 Chamber Height (CH)

Chamber height (CH) is the perpendicular distance between the anterior chamber and the horizontal line joining the two scleral spur points [5]. To identify the chamber height, a perpendicular line is drawn from the midpoint of the ACW line, so as to meet the cornea. The vertical Euclidean distance between the intersecting point on

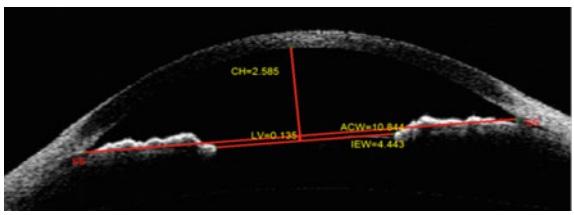


Fig. 4 Image showing values of ACW, IEW, CH, and LV

the corneal endothelium layer and the midpoint of ACW line gives the chamber height (Fig. 4).

4.4 Lens Vault (LV)

It is the perpendicular distance between IEW line and the horizontal line joining two scleral spur points [5]. Figure 4 shows the above four features and their values marked in an AS-OCT image.

4.5 Angle Opening Distance (AOD)

Angle opening distance is the perpendicular distance between the cornea and the iris. The perpendicular line has been drawn from the point which is 500 μm anterior to the sclera spur [5].

For identifying the AOD, first, an arc is drawn with scleral spur point as center with a radius of 500 μm on both the sides of the canny edge detected image. The arc meets the canny edge image at two places, and a perpendicular line is drawn from the intersecting point on corneal endothelium (point A) such that it meets the iris upper boundary, and the intersecting point (point B) is marked. The perpendicular distance between the above-identified points A and B gives angle opening distance (AOD) (Fig. 5).

4.6 Trabecular Iris Angle (TIA)

Trabecular iris angle is the angle between the apex in the iris recess and the arms of the angle passing through a point on the trabecular meshwork, 500 μm from the scleral spur and the point on the iris perpendicularly [5]. Here, the trabecular iris angle is found between two lines. The first line is between the apex point and point A, and the second line is between the apex point and point B. Points A and B are

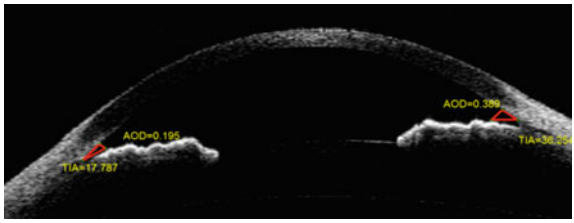


Fig. 5 Image showing value of AOD and TIA

the same points referred in AOD. Figure 5 shows the AOD and TIA marked input image.

5 Classification

The classifier is used to group the data based on its homogeneity in the parameters. Here, the classifier is needed to distinguish the normal and abnormal AS-OCT image based on its anterior segment parameters. Support vector machine (SVM) classifier is used here. SVM works on the principle of fitting the boundary to a region of points which are alike.

In this work, 30 images are considered. Out of 30, 20 images are used to train the classifier. The remaining 10 images are used for testing the performance of the classifier. All the 10 images have been correctly classified by the classifier. So, this algorithm can be used for identifying the glaucoma parameters in AS-OCT images. Six features are extracted from each image. Among the six features, AOD and TIA are identified on both sides of image, as it may vary. The other four parameters are ACW, IEW, CH, and LV. The features obtained from left and right side of an eye image are considered to be separate datasets. Parameter values computed from the images are in pixels. These pixel values are converted into the real-time distance by multiplying with the calibration factor. The calibration factor here is 1 pixel being equal to 0.025 mm. Table 1 contains all the parameter values obtained for images used for training as well as testing. While training the datasets, the data points are set to center at their mean and standard deviation is set to one.

Each side of the image is to be classified as normal and abnormal AS-OCT image based on these feature values. As given in the table, there is a possibility that only one side of an eye is affected with glaucoma. In such a case, the common parameters (such as ACW, IEW, CH, and LV) will be the same for both normal and abnormal sides of the eye. These are considered to be overlapping. But the other parameters (AOD and TIA) show a significant difference to classify as normal or abnormal. All the feature values of the other normal cases are in the same range. Hence, the classifier is able to classify the datasets correctly.

6 Conclusion

Initially, preprocessing was done on the image which gave an enhanced image with good contrast. From the preprocessed image, clinical features are extracted for identification of glaucoma. Based on the extracted features, classification process identifies normal and abnormal images. Results show that the developed algorithm helps to detect glaucoma for different AS-OCT images which are obtained from normal and glaucoma patients.

Table 1 Extracted values of anterior segment parameters

Images no.	ACW (mm)	IEW (mm)	CH (mm)	LV (mm)	Right/left side	AOD (mm)	TIA (°)	Normal/abnormal
1	10.8439	4.4430	2.5848	0.1352	Left	0.1953	17.7872	Abnormal
					Right	0.3889	36.2538	Normal
2	10.7315	4.9953	2.6097	0.2260	Left	0.5551	35.1915	Normal
					Right	0.3260	30.9411	Normal
3	10.6494	3.3613	2.4130	0.1854	Left	0.3202	24.8142	Normal
					Right	0.2305	28.3007	Normal
4	9.5276	3.3835	1.7529	0.0664	Left	1.1118	10.3048	Abnormal
					Right	0.2305	30.5102	Normal
5	10.5300	2.2558	2.5676	0.2192	Left	0.2850	37.6059	Normal
					Right	0.4070	38.7117	Normal
6	10.5452	2.3939	2.5101	0.2703	Left	0.2151	25.8577	Normal
					Right	0.4301	26.7606	Normal
7	11.3684	3.1394	2.8647	0.2186	Left	0.2358	23.2860	Normal
					Right	0.3750	35.2175	Normal
8	10.7865	3.2993	2.5622	0.1429	Left	0.3553	29.8814	Normal
					Right	0.2658	23.7429	Normal
9	9.400	4.9751	1.900	0.1058	Left	0.1031	17.1250	Abnormal
					Right	0.6543	43.5312	Normal
10	10.9638	1.8313	2.8767	0.2572	Left	0.3363	35.4475	Normal
					Right	0.4802	45.0000	Normal
11	10.4705	1.8312	2.5828	0.3548	Left	0.6915	40.5184	Normal
					Right	0.6021	50.7470	Normal
12	11.3526	2.9790	2.6699	0.2931	Left	0.1768	29.0546	Normal
					Right	0.3052	40.3031	Normal
13	10.9184	4.9930	2.4467	0.3454	Left	0.2574	39.5589	Normal
					Right	0.1820	18.2604	Abnormal
14	11.6392	4.8023	2.8278	0.3547	Left	0.5706	45.2768	Normal
					Right	0.4070	34.8358	Normal
15	11.3684	5.3505	2.3755	0.3643	Left	0.5506	50.0796	Normal
					Right	0.4828	24.6559	Normal

Acknowledgements The authors wish to thank Dr. Ronnie Jacob George, Director Research, Vision Research Foundation, Chennai for his valuable suggestions.

References

1. Arevalo JF, Krivoy D, Fernandez CF (2009) How does optical coherence tomography work basic principles. In: Arevalo JF (ed) Retinal angiography and optical coherence tomography Springer, New York, NY
2. Campbell P, Redmond T, Agarwal R, Marshal LR, Evans BJ (2015) Repeatability and comparison of clinical techniques for anterior chamber angle assessment. Ophthalmic Physiol Opt 35(2):170–178

3. Filipe HP, Carvalho M, Freitas MDL, Correa ZM (2016) Ultrasound biomicroscopy and anterior segment optical coherence tomography in the diagnosis and management of glaucoma. Vision Pan-Am Pan-Am J Ophthalmol 15(2):37–42
4. Boyd K (2017) American academy of ophthalmology. https://www.aao.org/eyehealth/disease/what-is-glaucoma
5. Fu H, Xu Y, Lin S, Zhang X, Wong DWK, Liu J, Frangi AF, Baskaran M, Aung T (2017) Segmentation and quantification for angle-closure glaucoma assessment in anterior segment OCT. IEEE Trans Med Imag 36(9):109–117
6. Ang M, Chong W, Tay WT, Yuen L, Wong TY, He M-G, Mehta JS (2012) Anterior segment optical coherence tomography study of the cornea and anterior segment in adult ethnic South Asian Indian eyes. Investig Ophthalmol Visual Sci 53(1):120–125
7. Ni SN, Marziliano P, Wong H-T (2014) Angle closure glaucoma detection using fractal dimension index on SS-OCT images. In: 36th international conference of the IEEE engineering in medicine and biology society, Chicago, USA, pp 3885–3888