

# Artificial Intelligence in Medical Applications

Lead Guest Editor: Yung-Kuan Chan

Guest Editors: Yung-Fu Chen, Tuan Pham, Weide Chang, and Ming-Yuan Hsieh





---

# **Artificial Intelligence in Medical Applications**

Journal of Healthcare Engineering

---

## **Artificial Intelligence in Medical Applications**

Lead Guest Editor: Yung-Kuan Chan

Guest Editors: Yung-Fu Chen, Tuan Pham, Weide Chang,  
and Ming-Yuan Hsieh



---

Copyright © 2018 Hindawi. All rights reserved.

This is a special issue published in “Journal of Healthcare Engineering.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Editorial Board

Saverio Affatato, Italy  
Hasan Ayaz, USA  
William Bertucci, France  
Olivier Beuf, France  
Hongwei Chen, USA  
Daniel H.K. Chow, Hong Kong  
Gianluca Ciardelli, Italy  
Elena De Momi, Italy  
Tiago H. Falk, Canada  
Mostafa Fatemi, USA  
Jesus Favela, Mexico  
Joseph Finkelstein, USA  
Jesus Fontecha, Spain  
Shahram M. Ghanaati, Germany  
Luca Giancardo, USA  
Antonio Gloria, Italy  
Kheng-Lim Goh, Singapore  
Pedro Gomes, Portugal  
Carlos Gómez, Spain  
Philippe Gorce, France  
Sophia Z. Gu, Australia  
Vincenzo Guarino, Italy  
Valentina Hartwig, Italy

Andreas H. Hielscher, USA  
Norio Iriguchi, Japan  
Zhongwei Jiang, Japan  
Rashed Karim, UK  
Pasi A. Karjalainen, Finland  
John S. Katsanis, Greece  
Terry K.K. Koo, USA  
Panagiotis Kosmas, UK  
Michel Labrosse, Canada  
Jui-Yang Lai, Taiwan  
Xiang Li, USA  
Feng-Huei Lin, Taiwan  
Yuan-Pin Lin, Taiwan  
Maria Lindén, Sweden  
Dongfei Liu, Finland  
Francisco Lopez-Valdes, Spain  
Zufu Lu, Australia  
Andreas Maier, Germany  
Mehran Moazen, UK  
Rafael Morales, Spain  
David Moratal, Spain  
Angkoon Phinyomark, Canada  
Vincenzo Positano, Italy

Alessandro Reali, Italy  
Lei Ren, UK  
Jose Joaquin Rieta, Spain  
Emanuele Rizzuto, Italy  
Sébastien Roth, France  
Hélder A. Santos, Finland  
Emiliano Schena, Italy  
Maurizio Schmid, Italy  
Jiann-Shing Shieh, Taiwan  
Tiago H. Silva, Portugal  
Jinshan Tang, USA  
Vinoy Thomas, USA  
Ioannis G. Tollis, Greece  
Uche Wejinya, USA  
Damon W.K. Wong, Singapore  
Ying Yang, UK  
Ping Yu, Australia  
Haihong Zhang, Singapore  
Hongbo Zhang, Finland  
Ping Zhou, USA  
Loredana Zollo, Italy

# Contents

## Artificial Intelligence in Medical Applications

Yung-Kuan Chan , Yung-Fu Chen , Tuan Pham , Weide Chang, and Ming-Yuan Hsieh  
Editorial (2 pages), Article ID 4827875, Volume 2018 (2018)

## 3D Shape-Weighted Level Set Method for Breast MRI 3D Tumor Segmentation

Chuin-Mu Wang , Chieh-Ling Huang , and Sheng-Chih Yang   
Research Article (15 pages), Article ID 7097498, Volume 2018 (2018)

## Strabismus Recognition Using Eye-Tracking Data and Convolutional Neural Networks

Zenghai Chen, Hong Fu , Wai-Lun Lo, and Zheru Chi  
Research Article (9 pages), Article ID 7692198, Volume 2018 (2018)

## A Novel Multiscale Gaussian-Matched Filter Using Neural Networks for the Segmentation of X-Ray Coronary Angiograms

Ivan Cruz-Aceves , Fernando Cervantes-Sanchez, and Maria Susana Avila-Garcia   
Research Article (11 pages), Article ID 5812059, Volume 2018 (2018)

## Effective Pneumothorax Detection for Chest X-Ray Images Using Local Binary Pattern and Support Vector Machine

Yuan-Hao Chan, Yong-Zhi Zeng, Hsien-Chu Wu , Ming-Chi Wu, and Hung-Min Sun   
Research Article (11 pages), Article ID 2908517, Volume 2018 (2018)

## Design of a Clinical Decision Support System for Fracture Prediction Using Imbalanced Dataset

Yung-Fu Chen , Chih-Sheng Lin, Kuo-An Wang, La Ode Abdul Rahman, Dah-Jye Lee ,  
Wei-Sheng Chung, and Hsuan-Hung Lin   
Research Article (13 pages), Article ID 9621640, Volume 2018 (2018)

## Automatic Semantic Segmentation of Brain Gliomas from MRI Images Using a Deep Cascaded Neural Network

Shaoguo Cui , Lei Mao, Jingfeng Jiang , Chang Liu, and Shuyu Xiong  
Research Article (14 pages), Article ID 4940593, Volume 2018 (2018)

## A Deep Belief Network and Dempster-Shafer-Based Multiclassifier for the Pathology Stage of Prostate Cancer

Jae Kwon Kim , Mun Joo Choi, Jong Sik Lee, Jun Hyuk Hong, Choung-Soo Kim , Seong Il Seo,  
Chang Wook Jeong, Seok-Soo Byun, Kyo Chul Koo, Byung Ha Chung, Yong Hyun Park, Ji Youl Lee,  
and In Young Choi   
Research Article (8 pages), Article ID 4651582, Volume 2018 (2018)

## An Evolutionary Computation Approach for Optimizing Multilevel Data to Predict Patient Outcomes

Sean Barnes , Suchi Saria, and Scott Levin  
Research Article (10 pages), Article ID 7174803, Volume 2018 (2018)

## Outlier Removal in Model-Based Missing Value Imputation for Medical Datasets

Min-Wei Huang , Wei-Chao Lin , and Chih-Fong Tsai   
Research Article (9 pages), Article ID 1817479, Volume 2018 (2018)



---

**Comparative Analysis of Hybrid Models for Prediction of BP Reactivity to Crossed Legs**

Gurmanik Kaur, Ajat Shatru Arora, and Vijender Kumar Jain

Research Article (13 pages), Article ID 2187904, Volume 2017 (2018)

**Clinical Decision Support System for Diabetes Based on Ontology Reasoning and TOPSIS Analysis**

Rung-Ching Chen, Hui Qin Jiang, Chung-Yi Huang, and Cho-Tsan Bau

Research Article (14 pages), Article ID 4307508, Volume 2017 (2018)

## Editorial

# Artificial Intelligence in Medical Applications

**Yung-Kuan Chan** <sup>1</sup>, **Yung-Fu Chen** <sup>2</sup>, **Tuan Pham** <sup>3</sup>, **Weide Chang**,<sup>4</sup>  
**and Ming-Yuan Hsieh**<sup>5</sup>

<sup>1</sup>Department of Management Information Systems, National Chung Hsing University, No. 250, Kuokuang Rd., Taichung 402, Taiwan

<sup>2</sup>Department of Dental Technology and Materials Science, Central Taiwan University of Science and Technology, No. 666, Buzih Road, Beitun District, Taichung 40601, Taiwan

<sup>3</sup>Department of Biomedical Engineering, Linköping University, 462 Entrance 71, Floor 12, Room 403, Campus US, Linköping, Sweden

<sup>4</sup>Department of Computer Science, California State University, 6000 J Street, Sacramento, CA 95819-6021, USA

<sup>5</sup>Department of International Business, National Taichung University of Education, No. 140, Minsheng Rd., West District, Taichung 40306, Taiwan

Correspondence should be addressed to Yung-Kuan Chan; [ykchan@nchu.edu.tw](mailto:ykchan@nchu.edu.tw)

Received 26 June 2018; Accepted 28 June 2018; Published 15 July 2018

Copyright © 2018 Yung-Kuan Chan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medical artificial intelligence (medical AI) mainly uses computer techniques to perform clinical diagnoses and suggest treatments. AI has the capability of detecting meaningful relationships in a dataset and has been widely used in many clinical situations to diagnose, treat, and predict the results. In the research and studies of medical AI, we primarily focus on the viability and feasibility to incorporate various computer AI techniques in medical information modeling and clinical procedure deployments. The state-of-the-art AI methodologies have shown great capabilities and capacities in recognition of meaningful data patterns and thus been widely experimented as tools for clinical trials, especially, to aid the decision making in each phase for diagnoses and subsequent treatments, as well as prognoses and projections.

The main focus of this special issue is on the proposal of techniques for medical artificial intelligence, expert systems, data mining, machine learning, and image processing which could be built on top of them. This special issue summarizes the most recent developments in the field, with a special emphasis given to the improvements and results obtained within the last several years. With highly measurable improvements, the MAI issue demonstrates the great potential and promises of applying AI techniques to pragmatic clinical

procedures and analytical medical informatics, especially in the following areas:

- (i) Artificial Intelligence Techniques in Medicine
- (ii) Data Mining and Knowledge Discovery in Medicine
- (iii) Medical Expert Systems
- (iv) Machine Learning-Based Medical Systems
- (v) Medical Signal and Image Processing Techniques

Thirty-three papers were submitted for this special issue. Our distinguished reviewers from respective research fields narrowed the field to eleven papers which were finally accepted. The following is a short summary of the findings of each of these papers.

Avila-Garcia et al. demonstrated the use of a neural network-based multiscale Gaussian matching filter for detection and segmentation on coronary angiogram X-ray images and thus effectively enhanced the results of image classification.

Cui et al. proposed a cascaded neural network composed of a Tumor Localization Network to localize the brain tumor from slices of MRI images and an Intra-Tumor Classification Network to label tumor regions. With advanced technologies used to train and optimize the cascaded neural network,

it attained better accuracy and computation efficiency over other neural network-based methods.

Fu et al. published their results of using a Convolutional Neural Network (CNN) to recognize strabismus. The CNN was trained from the data collected by an eye-movement tracker, including Gaze Deviation (GaDe) image data from both subjects of normal and strabismic visions. After trained with a large number of GaDe images, their CNN performed successfully for strabismus recognition.

Chan et al. deployed a support vector machine to effectively detect common pneumothorax by using the local binary patterns obtained through a multiscale intensity texture analysis on the chest X-ray images.

Chen et al. designed a clinical decision support system to predict fractures in hip bones and vertebrates caused by medications for treatments of chronic respiratory diseases. The system was designed with an integrated genetic algorithm and support vector machine through training with balanced datasets obtained from random and cluster-based undersampling methods, as well as tested with imbalanced datasets.

Barnes et al. proposed a data-driven predictive modeling framework for decision support systems, based on evolutionary computation techniques to optimize multilevel data. The framework could be built from open-source software and flexible to include other evolutionary algorithms.

Kim et al. proposed a Deep Belief Network and Dempster-Shafer- (DBN-DS-) based multiclassifier for the pathologic prediction of prostate cancer, tested with data from thousands of patients, and obtained high accuracy (81.27%) against the Partin table (64.14%).

Kaur et al. examined in detail the performances of hybrid computing models used to predict blood pressures (BP). Each hybrid model was formed with principal component analysis (PCA) and one of the following: Forward Stepwise Regression (FSWR), Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Least Square-Support Vector Machine (LS-SVM). The BP predicted was associated with reactivity to crossed legs among normotensive and hypertensive participants. The LS-SVM model achieved significant improvements.

Yang experimented with 3D Shape-Weighted Level Set Method (3D-SLSM) to carry out precise segmentation of tumors from 3D medical images. The segmentation results of most 3D segmentation algorithms were affected largely by errors and noise; however, 3D-SLSM added 3D shape-weighted values in each iterative process according to the volume changes. The accuracy of 3D-SLSM was tested with MRI and computer-generated images and had the highest accuracy and lowest false-positive rate when compared with the standard tumor model.

Lin et al. researched on effects of outlier removals in medical datasets, that is, whether performing instance selection to filter out noisy data from a given training medical dataset has a positive impact on the final imputation results. By comparing the classification performance obtained through the processes combining instance selection and imputation and the baseline imputation process, three types of medical datasets are used: categorical, numerical, and

mixed types. Three different instance selection methods and three model-based imputation algorithms were compared. The experimental results showed that performing instance selection first mostly improves the imputation result over the types of medical data. Further, the negative impact was to consider instance selection before imputation when the dataset contains lower dimensionalities and numbers of classes. However, for numerical datasets, the combined instance selection and imputation process would perform better than the baseline imputation process for most datasets with different missing rates. They also concluded that for mixed datasets, the instance selection effect was between that for categorical and numerical datasets, which meant that combining instance selection and imputation could be a better choice when algorithms were carefully chosen.

Chen devised a personalized antidiabetic medication recommendation system that tailored HbA1c levels to mitigate medical differences for the selection of antidiabetic medications. By combining fuzzy logic and ontology reasoning, the system could promote a “patient-centric diabetes therapy” besides aiding clinical diagnosis and assist family practitioners in prescribing medications.

*Yung-Kuan Chan  
Yung-Fu Chen  
Tuan Pham  
Weide Chang  
Ming-Yuan Hsieh*

## Research Article

# 3D Shape-Weighted Level Set Method for Breast MRI 3D Tumor Segmentation

Chuin-Mu Wang <sup>1</sup>, Chieh-Ling Huang <sup>2</sup>, and Sheng-Chih Yang <sup>1</sup>

<sup>1</sup>Department of Computer Science and Information Engineering, National Chin Yi University of Technology, Taichung 41170, Taiwan

<sup>2</sup>Department of Interaction Design, Chang Jung Christian University, Tainan 71101, Taiwan

Correspondence should be addressed to Sheng-Chih Yang; [scyang@ncut.edu.tw](mailto:scyang@ncut.edu.tw)

Received 23 July 2017; Accepted 30 April 2018; Published 13 June 2018

Academic Editor: Weide Chang

Copyright © 2018 Chuin-Mu Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Three-dimensional (3D) medical image segmentation is used to segment the target (a lesion or an organ) in 3D medical images. Through this process, 3D target information is obtained; hence, this technology is an important auxiliary tool for medical diagnosis. Although some methods have proved to be successful for two-dimensional (2D) image segmentation, their direct use in the 3D case has been unsatisfactory. To obtain more precise tumor segmentation results from 3D MR images, in this paper, we propose a method known as the 3D shape-weighted level set method (3D-SLSM). The proposed method first converts the LSM, which is superior with respect to 2D image segmentation, into a 3D algorithm that is suitable for overall calculations in 3D image models, and which improves the efficiency and accuracy of calculations. A 3D shape-weighted value is then added for each 3D-SLSM iterative process according to the changes in volume. Besides increasing the convergence rate and eliminating background noise, this shape-weighted value also brings the segmented contour closer to the actual tumor margins. To perform a quantitative analysis of 3D-SLSM and to examine its feasibility in clinical applications, we have divided our experiments into computer-simulated sequence images and actual breast MRI cases. Subsequently, we simultaneously compared various existing 3D segmentation methods. The experimental results demonstrated that 3D-SLSM exhibited precise segmentation results for both types of experimental images. In addition, 3D-SLSM showed better results for quantitative data compared with existing 3D segmentation methods.

## 1. Introduction

In the process of breast cancer screening or medical treatment, the size and shape of a tumor is often an important basis for the diagnosis or treatment strategy. Segmenting the tumor from a medical image can improve the diagnostic accuracy of doctors, and become a guide for the surgery. In the past, many two-dimensional (2D) image segmentation techniques have been developed, such as the active contour model (ACM) [1], region growing, zero crossing [2], thresholding [3], region-based segmentation [4], watershed [5], fuzzy c-means (FCM), texture features, and the level set method (LSM) [6, 7]. However, breast MRI has a relatively low resolution, and tumor boundaries are often indistinct as tumors infiltrate surrounding healthy tissue. Consequently, breast MRI tumor segmentation has always been a challenging task. According to the literature on image segmentation, some segmentation

methods are based on the brightness; however, these methods are easily affected by noise. Some methods are gradient-based and may result in errors if the boundaries are not clear. In addition, some methods are based on local features; however, they are often dependent on the suitability of the features extracted as well as the image consistency. If these methods are used in highly variable medical images, their segmentation accuracy is also relatively unstable. However, some segmentation methods demonstrate superior performance, such as LSM and multispectral detection technology [7] that do give good results for tumor segmentation in 2D medical images. It is noteworthy that in order to meet the needs of multispectral detection technology, different parameters (such as T1, T2, and PD) must be used to produce multispectral MRIs.

Using three-dimensional (3D) models would undoubtedly determine tumor size and shape more accurately than using 2D imaging. However, most of the current

medical imaging instruments present only 2D images. Although there are some relatively expensive instruments that can stack the original 2D images directly into a 3D model, the original 2D images usually contain mixed complex background tissues, often making the object difficult to recognize in 3D. Therefore, the segmentation of 3D medical images has become a computer-aided diagnostic technology in dire need of development [8]. Due to the technologies of 2D image segmentation and contour detection are now relatively mature, some 3D image segmentation methods use 2D segmentation as the foundation to carry out segmentation on 2D sections before stacking these 2D segmentation results into a 3D segmentation model [9–11]. As these 3D segmentation methods lack association between upper and lower sections (Z-axis), the accuracy of these methods in 3D segmentation applications is not as high as in 2D segmentation. In addition, as these methods calculate every 2D section one-by-one, and not the total 3D calculations, the computational burden is significantly increased. At present, there are some existing 3D image segmentation techniques. For example, Rebouças et al. [12] developed ACM into 3D-ACM and compared it with 3D region growing. Although the experimental results showed that 3D-ACM had the better performance, it was closely related to the settings of the initial contour and was difficult to use as a clinical diagnostic aid. In addition, Gangsei and Kongsro [11] extended Dijkstra’s algorithm to a 3D algorithm, and conducted 3D segmentation for bone CT images. Despite the fact that satisfactory segmentation results were obtained, this method is not suitable for use in breast MRI with its various mixed tissues and low contrasts. This is because there is a high contrast between the target and background tissues in bone and vascular imaging, a fact that is exploited by other 3D image segmentation algorithms [13, 14] for bones and blood vessels. Some methods carry out segmentation at three orthogonal 2D planes before combining these segmentation results into a 3D segmentation model [14, 15]. Although these methods significantly decrease the computational burden, they are still based on 2D segmentation and are not strictly true 3D segmentation methods.

Based on the aforementioned analyses, this paper proposes a new 3D tumor segmentation method, namely, the 3D shape-weighted level set method (3D-SLSM). In comparison with past methods as well as our previous research results in [9], 3D-SLSM has three major advantages. Firstly, it is evolved from 2D-LSM, whose tissue segmentation has been confirmed to give good results for 2D breast MRI [7]. Secondly, 3D-SLSM operates directly on the entire 3D model, which not only reduces the computation time but also ensures the association and interaction between each pixel and its neighboring points on all three axes (X, Y, and Z). Thirdly, 3D-SLSM adds a characteristic shape-weighted model in each update, so that the contour converges rapidly towards the surface of the target object, effectively eliminating the mixed surrounding background noise. In order to verify experimentally the feasibility of the proposed method, experimental data are assigned to two groups: computer-simulated images and breast MRI cases with actual tumors. Computer-simulated images help us to observe based on quantitative analysis while breast MRIs are used to confirm

efficiency in clinical applications. In the course of the experiment, the segmentation of experimental images using various algorithms will be carried out. However, as the 3D segmentation results of the various algorithms are 3D point matrices that cannot be directly viewed, rendering techniques must be first employed to convert these 3D point matrices into visual 3D image models to facilitate observation and actual applications. A quantitative evaluation is then conducted using a standard model (delineated by physicians for actual MRI cases) as a basis. Besides evaluating the performance of 3D-SLSM, the accuracy and error rates of 3D-SLSM are also compared to existing algorithms, such as traditional ACM and 2D/3D-LSM, in order to validate the contributions of 3D-SLSM.

This paper is divided into five sections, which are as follows: Section 2 introduces the new 3D-SLSM method proposed in this paper. Section 3 describes the reconstruction and system evaluation methods for experimental data. Section 4 presents experimental results and discussion. Finally, conclusions are presented in Section 5.

## 2. Methods

*2.1. Three-Dimensional Level Set Method (3D-LSM).* The traditional LSM was first proposed in 1988 by Osher and Sethian [16], and it is still used widely in many disciplines today. LSM has already been confirmed to have superior performance in the segmentation of 2D medical images. Therefore, while developing 3D-SLSM, we selected LSM as a foundation, and first converted traditional 2D-LSM into 3D-LSM, which can be used for overall calculations in 3D models. In addition to increasing the computational efficiency, this conversion also significantly increases the accuracy of 3D segmentation as the computation process considers the association of pixels on upper and lower sections (Z-axis). Conventional 2D-LSM calculations require the construction of an initial 2D target region, and  $\varphi$  is used to express the height level of every image pixel. After the boundaries of the initial target region were taken as horizontal lines ( $\varphi = 0$ ), a height distribution map is then constructed according to the characteristics of the various positions in the images. The proposed 3D-LSM first upgrades the 2D initial target region into a 3D initial target object before taking the contour surfaces of the initial target object as horizontal planes.  $\varphi = 0$  indicates that the point is on its contour surface,  $\varphi > 0$  indicates that the point is located inside the target, and  $\varphi < 0$  indicates that the point is located outside the target, as shown in Figure 1.

In order to increase the association of the Z-axis for conversion into a 3D algorithm, a new formula was redefined for 3D-LSM, as shown in the following equation:

$$\Delta\varphi = \delta(\varphi) \left( \mu \cdot \operatorname{div} \left( \frac{\nabla\varphi}{|\nabla\varphi|} \right) - \lambda_1 (\mu_0(x, y, z) - c_1)^2 - \lambda_2 (\mu_0(x, y, z) - c_2)^2 - \nu \right), \quad (1)$$

where  $\lambda_1$ ,  $\lambda_2$ ,  $\mu$ , and  $\nu$  are all weighted coefficients.  $c_1$  and  $c_2$  represent the average grayscale values inside and outside the

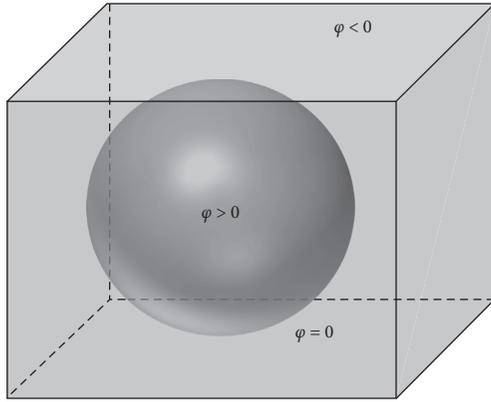


FIGURE 1: Initial contour surface of the 3D level set method.

contour, respectively, while  $\mu_0$  represents the pixel gray-level value. The function  $\delta(\varphi)$  is a Dirac delta function, which we approximate in our implementation as follows:

$$\delta(\varphi) = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + \varphi^2}, \quad (2)$$

where  $\varepsilon$  is a constant that is used to control the sharpness of the contour plane.  $\text{div}(\nabla\varphi/|\nabla\varphi|)$  in (1) is used to smooth out the entire contour plane, and can be obtained from the divergence of the various pixel-gradient ( $\nabla$ ) directions on the contour plane. As 3D-LSM increases the association between gradients on the Z-axis of the contour plane, it is therefore redefined as follows:

$$\text{div}\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right) = \frac{\gamma}{(fx^2 + fy^2 + fz^2)^{1.5}} * (fx^2 + fy^2 + fz^2)^{0.5}, \quad (3)$$

where  $fx$ ,  $fy$ , and  $fz$  represent the gradient quantization values on the X-, Y- and Z-axes of the image, respectively. The term  $\gamma$  can be obtained by the following formula:

$$\begin{aligned} \gamma = & fy^2 * fxx + fz^2 * fxx + fx^2 * fyy + fz^2 * fyy \\ & + fx^2 * fzz + fy^2 * fzz - 2 * fx * fy * fyz \\ & - 2 * fx * fy * fxy - 2 * fx * fz * fxz. \end{aligned} \quad (4)$$

In addition, the third part of  $(1) - \lambda_1(\mu_0(x, y, z) - c_1)^2 - \lambda_2(\mu_0(x, y, z) - c_2)^2$ —is the key to moving the contour gradually to the edge of the object in the updating process. Through observation, it was discovered that increasing  $\lambda_1$  could cause the internal broken areas of the contour surface to be connected. This usually makes it easier to identify a target with lower grayscale values. Increasing  $\lambda_2$  can cause the external broken areas of the contour surface to be connected, which helps identify a target with higher grayscale values (e.g., the tumor area in breast MRI). The value of  $v$  in the fourth term of (1) is the overall height adjustment. The higher this value is, the lower the overall  $\varphi$  value will be. Meanwhile, the volume enclosed by the entire contour surface will also be reduced, so the adjustment of  $v$  will affect the volume inside the contour surface.

**2.2. Three-Dimensional Surface Rendering (3DSR).** As the results obtained from 3D image segmentation are only 3D point

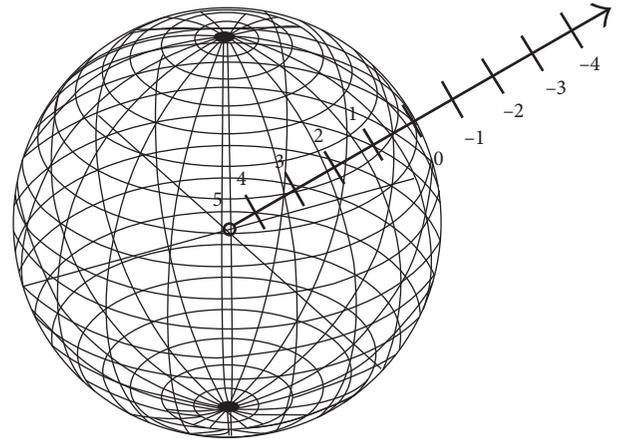


FIGURE 2: Three-dimensional shape-weighted value diagram.

matrices, which are not directly observable on the device, after segmentation using various methods was completed, there is a need to employ 3DSR (three-dimensional surface rendering) techniques to convert the 3D point matrix into a 3D model for viewing by users. While the 3DSR technique is not considered a part of 3D image segmentation, it is an important technique for employing segmentation algorithms in clinical applications. In general, there are two categories of 3DSR approach. The first is the isosurface approximation method [17], which attaches certain geometric planes to the equivalent surface and then uses the 3D mapping method to image the surface. The other is the light projection method [18], in which light is projected in the equidistant sampling mode for different types of accumulation. The typical techniques of 3DSR are contour-tracing isosurface (CTIS) [19], volume rendering (VR) [20], and marching-cubes isosurface (MCIS) [17, 21]. The main disadvantage of CTIS is that there may be multiple closed contours on each cross section, as well as significant differences between contour lines in two adjacent cross sections. This makes it very difficult to trace and connect the contour lines in adjacent cross sections, which might result in a large number of connection errors when tracing complex structures. The disadvantage of VR is that the projected value needs to be recalculated frequently during the process of 3D rotation, so many calculations are required. However, MCIS treats the small cubes in the 3D space as the basic units, and to find the respective isosurface of each one. The cubes can be categorized as being either inside or outside the object, based on values on the eight vertices; there are only 15 types in total, after eliminating rotationally similar states. The corresponding isosurface can then be generated quickly within the small cubes according to a lookup table. Based on the above analysis, this paper uses the MCIS method to construct a 3D model of a breast tumor. About the calculation process of MCIS, please refer to [21] for details.

**2.3. Three-Dimensional Shape-Weighted Level Set Method (3D-SLSM).** In order to improve the speed and accuracy of 3D segmentation, we now combine 3D-LSM with the 3D shape characteristics to obtain the 3D shape-weighted level set method (3D-SLSM). This adds a shape-weighted value each time LSM updates the  $\varphi$  value, thus controlling the

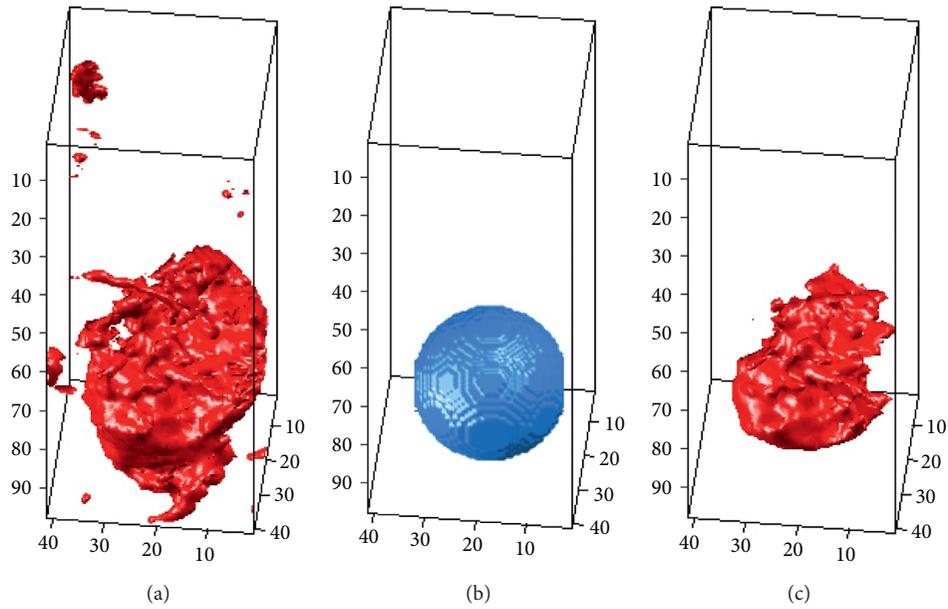


FIGURE 3: Effect of 3D shape-weighted value on a single iteration: (a) segmentation results without adding the 3D shape-weighted value; (b) calculated 3D shape-weighted value from (a); (c) segmentation results with the 3D shape-weighted value added.

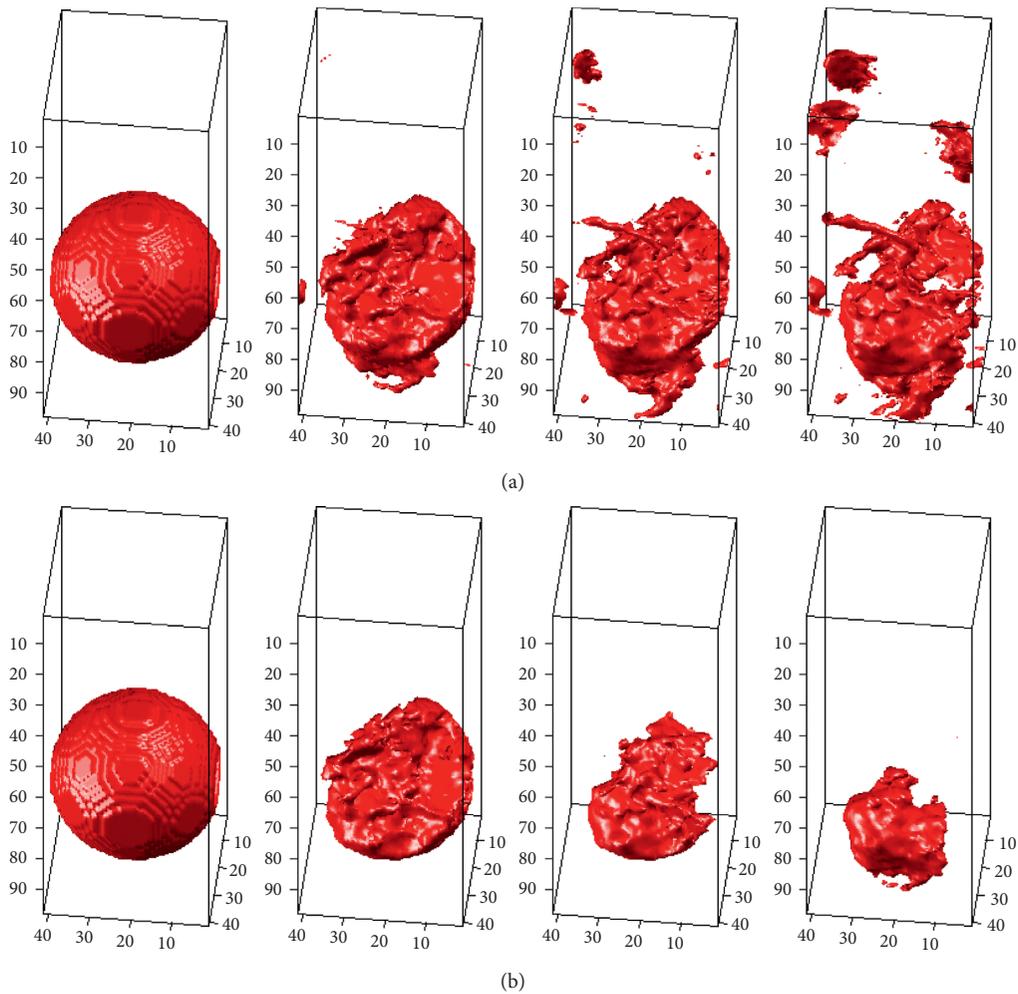


FIGURE 4: Effect of 3D shape-weighted value on the iterative calculation process: (a) partial iteration segmentation results without adding the 3D shape-weighted value; (b) partial iteration segmentation results with the 3D shape-weighted value added. Leftmost sphere is the initial contour surface in both cases.

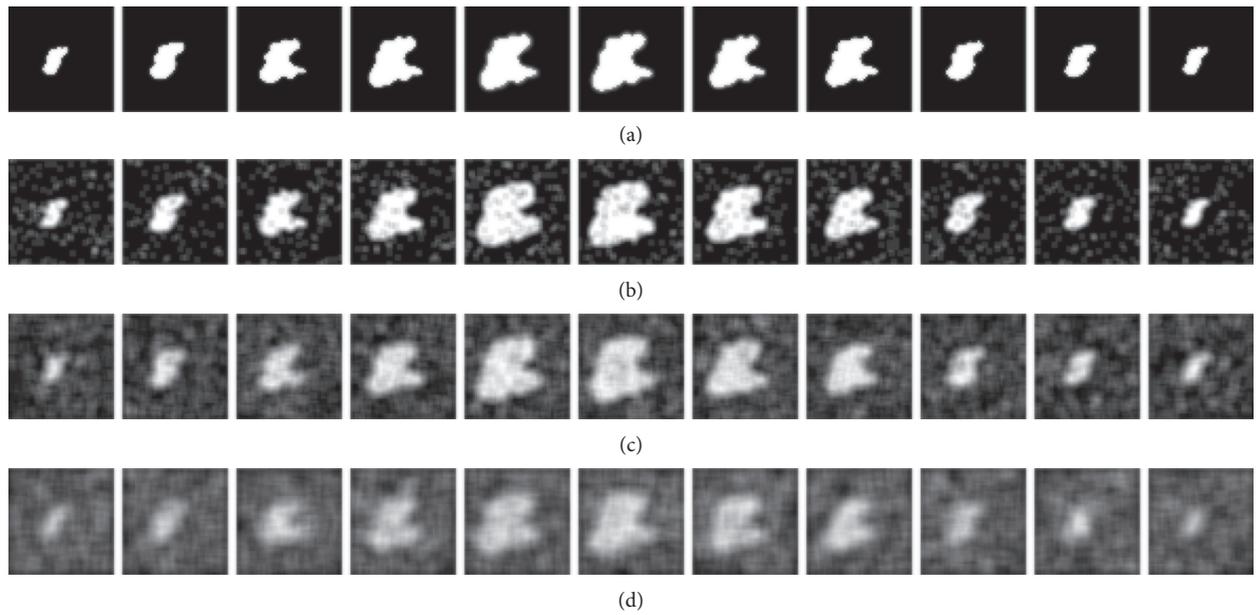


FIGURE 5: The simulated tumor images with three different levels of noise density and blurring: (a) the original image; (b) noise density at 10% with  $3 \times 3$  mask blurring; (c) noise density at 30% with  $5 \times 5$  mask blurring; (d) noise density at 50% with  $7 \times 7$  mask blurring.

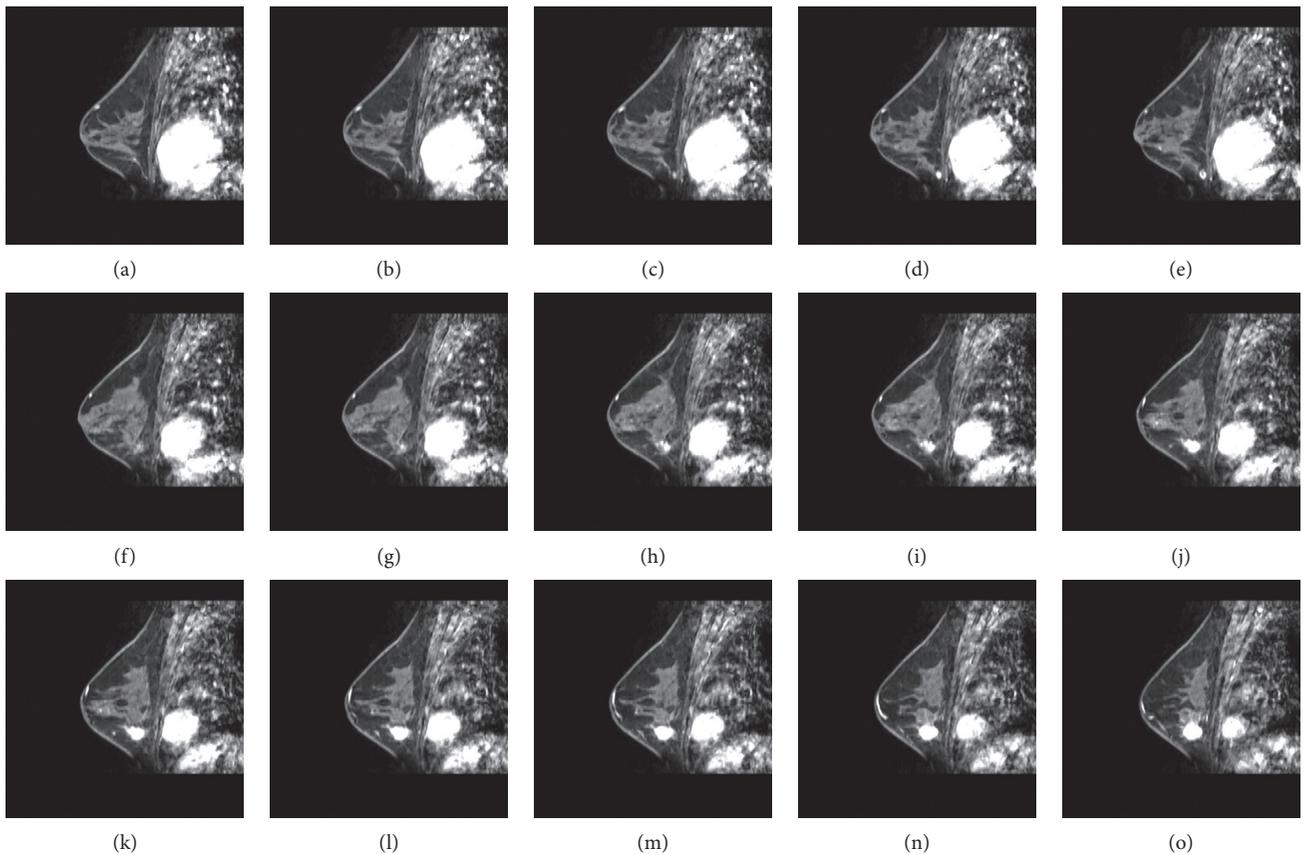


FIGURE 6: A partial image slice of an MRI case with a breast tumor.

convergence of the 3D contour lines. It is important to note that the calculated  $\varphi$  value in the 3D-LSM calculation process is a 3D matrix; thus, we have to use the 3D shape characteristic to combine this  $\varphi$  value effectively with 3D-LSM. Here,

a shape characteristic refers to the desired shape of the object; for different target tissues or organs in medical imaging, the expected shape can require different considerations. This paper uses a breast cancer tumor as a demonstration. In view

TABLE 1: Characteristic list of test cases.

Case number	Breast size	Tumor size	Breast tissue types
Case 1	Medium	Medium	Fatty glandular
Case 2	Small	Small	Fatty
Case 3	Large	Large	Dense glandular

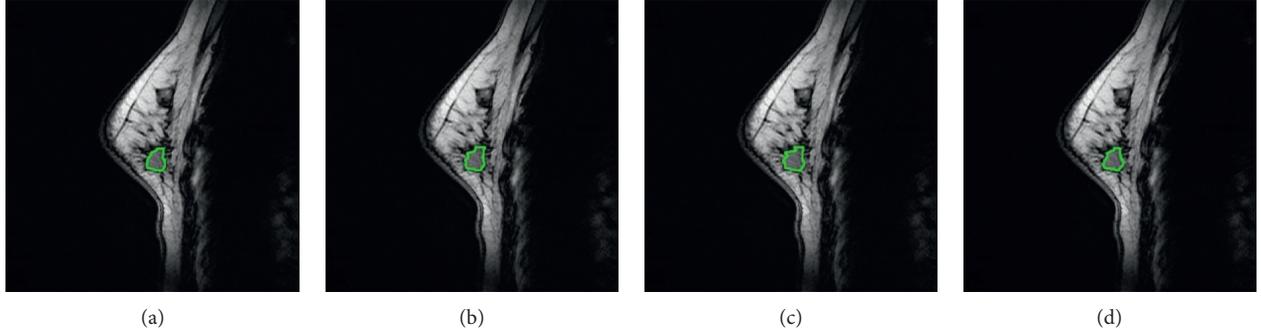


FIGURE 7: Establishment of standard tumor contour in a single slice: (a–c) tumor contour delineated by three experts; (d) contour formed by the intersection area of (a–c).

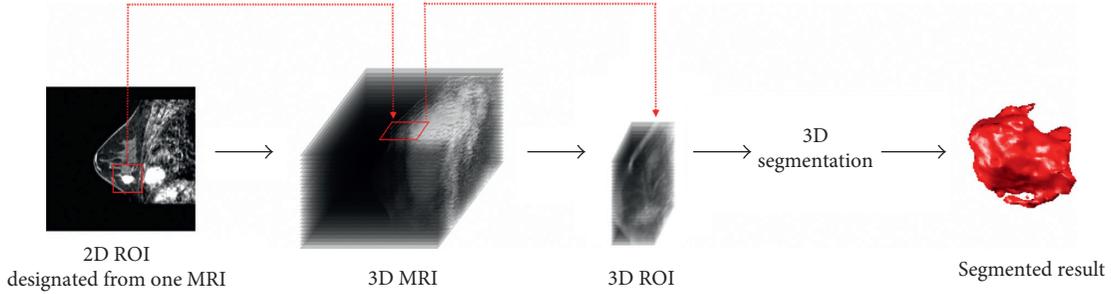


FIGURE 8: A schematic diagram for establishing a 3D ROI model based on actual MRI images.

of the usual presentation of this type of tumor as approximately spherical or ellipsoidal, the shape characteristic in the experiment is computed as a sphere. Accordingly, in each 3D-SLSM iterative operation, we must first initialize a sphere that has the same inner volume as the target. The center coordinates of the sphere locate the center of gravity for all pixels with  $\varphi > 0$ , and the volume of the sphere is equal to the sum of all such pixels. The radius of the sphere can be calculated through the sphere volume equation:

$$V = \frac{4}{3}\pi r^3, \quad (5)$$

where  $V$  represents the number of pixels with  $\varphi > 0$ ,  $\pi$  is the usual mathematical constant, and  $r$  is the radius we wish to calculate. After calculation of (5), we can obtain the radius of the sphere, and in this way, the initial characteristic shape (spherical) matrix can be established via the following equation:

$$\varphi_s(x, y, z) = -\sqrt{(x-c_x)^2 + (y-c_y)^2 + (z-c_z)^2} + r, \quad (6)$$

where  $c_x$ ,  $c_y$ , and  $c_z$  are the coordinates of the center of gravity of all pixels. After establishing the shape characteristic matrix, this is evolved further into a 3D shape-weighted value

that participates in each 3D-SLSM iterative operation. Initially, the weighted value on the characteristic shape surface is defined to be zero. The farther away the interior points are from the shape surface, the greater the weighted value is. The farther away the external points are from the shape surface, the smaller the weighted value is. These two conditions combine to make the contour in the 3D-SLSM calculation process converge towards the expected characteristic shape. The 3D shape-weighted value diagram is shown in Figure 2.

After combining the 3D shape-weighted value, the definition of 3D-SLSM is shown in the following equation:

$$\Delta\varphi = \delta(\varphi) \left( \mu \cdot \operatorname{div} \left( \frac{\nabla\varphi}{|\nabla\varphi|} \right) - \lambda_1 (\mu_0(x, y, z) - c_1)^2 - \lambda_2 (\mu_0(x, y, z) - c_2)^2 - \nu + \tau\varphi_s \right), \quad (7)$$

where  $\varphi_s$  is the 3D characteristic shape matrix produced by the characteristic shape,  $\tau$  is the weighted matrix, and a combination of the two becomes the 3D shape-weighted value that controls the degree of “force” that pushes the contour towards the characteristic shape. We use an example to demonstrate the effect of the 3D shape-weighted value. Figure 3(a) represents a 3D-LSM iterative operation. After calculating the center of

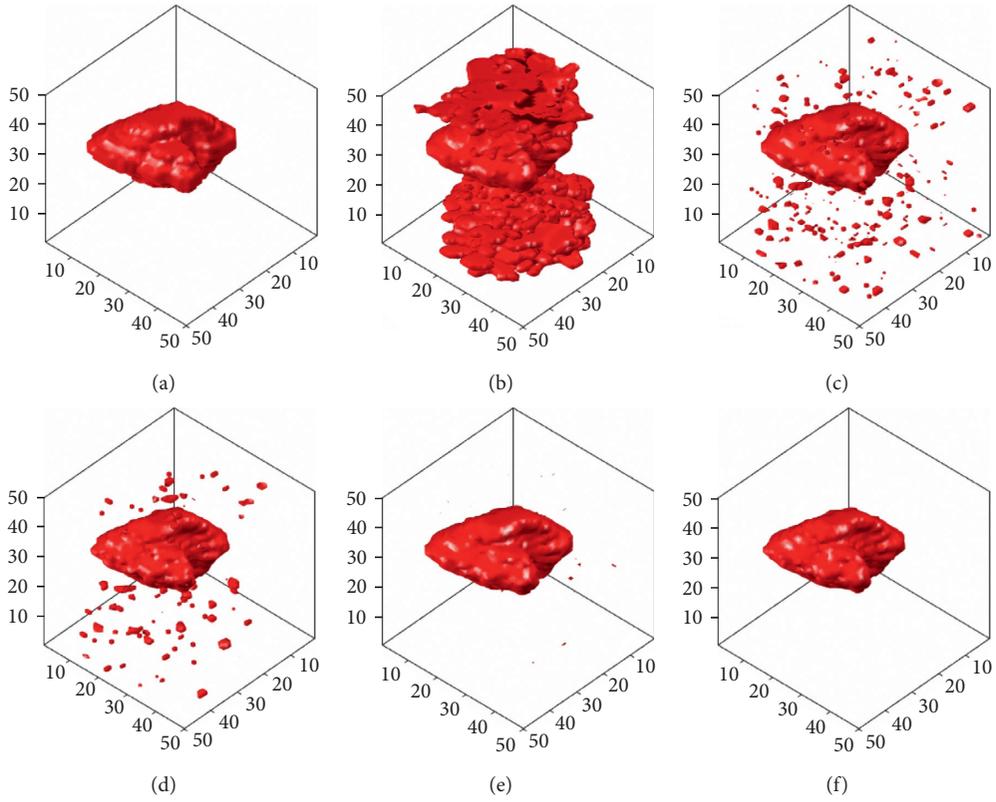


FIGURE 9: The segmentation results of computer-simulated images with 10% noise density and  $3 \times 3$  masked blurring: (a) standard; (b) ACM; (c) LSM; (d) SLSM; (e) 3D-LSM; (f) 3D-SLSM.

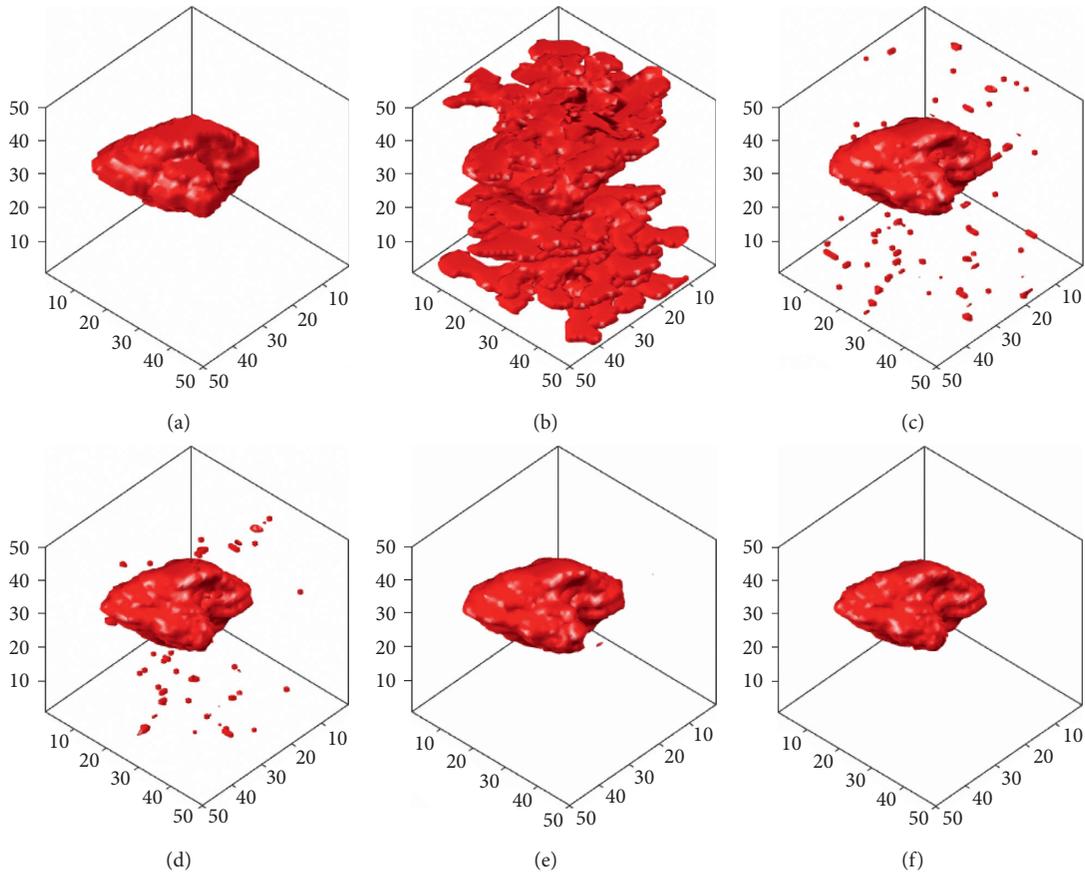


FIGURE 10: The segmentation results of computer-simulated images with 30% noise density and  $5 \times 5$  masked blurring: (a) standard; (b) ACM; (c) LSM; (d) SLSM; (e) 3D-LSM; (f) 3D-SLSM.

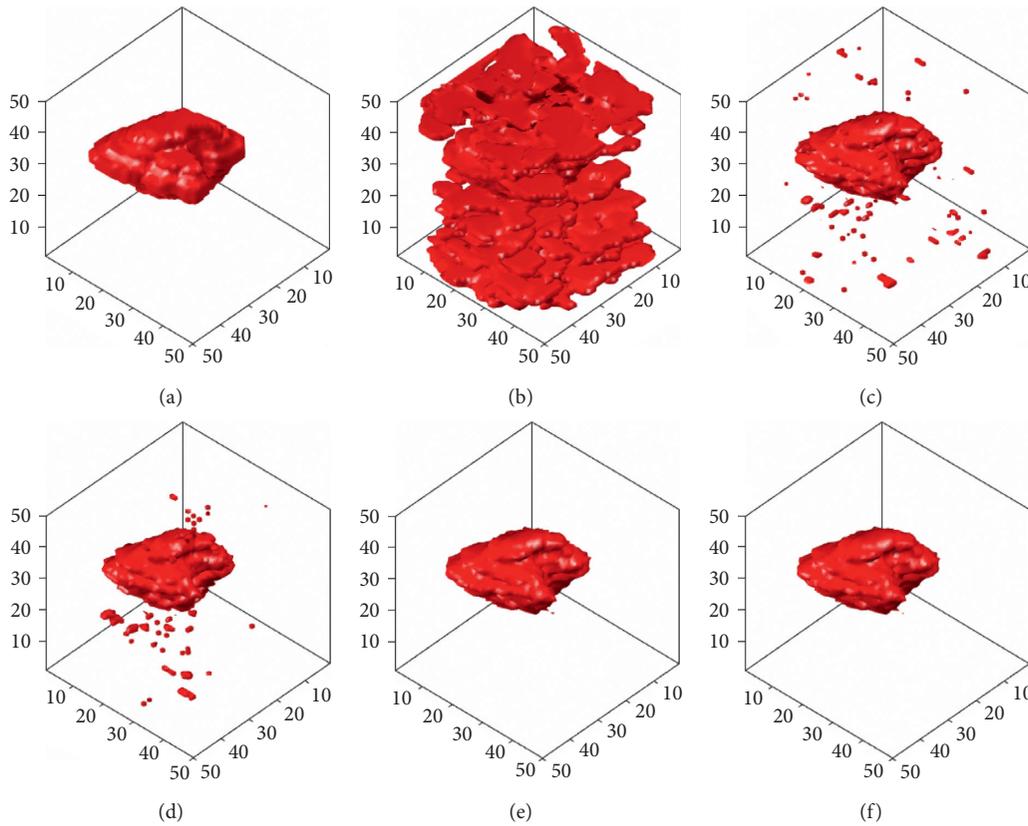


FIGURE 11: The segmentation results of computer-simulated images with 50% noise density and  $7 \times 7$  masked blurring: (a) standard; (b) ACM; (c) LSM; (d) SLSM; (e) 3D-LSM; (f) 3D-SLSM.

gravity and the volume of Figure 3(a), we can obtain the 3D shape-weighted value diagram in Figure 3(b). Based on the 3D shape-weighted design, the closer the points are to the center, the bigger the (positive) weight will be; the farther away the points are from the center, the smaller the (negative) weight will be. The calculated 3D-SLSM results combined with the 3D shape-weighted value are shown in Figure 3(c). Comparing Figures 3(a) and 3(c), we can see clearly in Figure 3(a) that there is a lot of surrounding noise and incomplete block fragments around the periphery of the tumor, while Figure 3(c) provides a closer look at the external appearance of the actual tumor. Figure 4 shows the effect of the 3D shape-weighted value in the iterative calculation process. The leftmost sphere in Figure 4(a) is the initial contour surface, followed by the segmentation models of partial iterative calculation without the 3D shape-weighted value. In Figure 4(b), the leftmost sphere is the initial contour surface, followed by the segmentation models of partial iterative calculation with the 3D shape-weighted value. Comparing Figures 4(a) and 4(b), it can be observed that the 3D shape-weighted value pulls the contour towards the inside of the characteristic shape, and eliminates the impact of environmental noise and fragmented blocks around the tumor.

### 3. Experimental Data and Evaluation Methods

This paper uses breast MRI images and computer-simulated images to demonstrate three-dimensional tumor segmentation and the evaluation of its efficacy. The computer simulation is

used for accurate quantitative analysis, and the actual cases aid in observing feasibility for various methods in clinical applications.

*3.1. Establishment of Experimental Data and Evaluation Criteria for Computer-Simulated Images.* For the computer-simulated images to closely replicate the actual MRI images of patients with tumors, in addition to the definition of the tumor regions in the simulated images, we also added noises of varying densities (10%–50%) and blurring using masks of different sizes ( $3 \times 3$  to  $7 \times 7$ ). Figure 5 shows the original image and the 2D computer-simulated tumor images with different noise densities and levels of blurring.

*3.2. Establishment of Experimental Data and Evaluation Criteria for Breast MRI Cases.* This paper uses breast MRI as a case study for the 3D tumor segmentation experiment and performance evaluation. The imaging sources are from Taiwan Tri-Service General Hospital. The collected cases all have actual tumors, and the dynamic contrast-enhanced MRI (DCE-MRI) was performed three minutes after the injection of the developer. The resolution of each image is  $512 \times 512$ , each case has 98 image slices with a slice spacing of 2 mm, and the scope of these slices contain the tumor site. Figure 6 shows a partial image slice of an MRI case with a breast tumor.

The selection of test cases took into account the different breast sizes, tumor sizes, and breast tissue types (as shown in

TABLE 2: Accuracy, specificity, and false alarm rate (%) of the different algorithms and noise densities in  $3 \times 3$  mask blurred computer-simulated images.

Noise density		Methods				
		ACM	LSM	SLSM	3D-LSM	3D-SLSM
5%	CCR	90.43	99.24	99.34	99.7	<b>99.76</b>
	SP	89.93	99.2	99.76	99.69	<b>99.96</b>
	FAR	10.13	0.8	0.7	0.31	<b>0.26</b>
10%	CCR	90.07	99.13	99.19	99.6	<b>99.67</b>
	SP	89.59	99.09	99.65	99.51	<b>99.95</b>
	FAR	10.52	0.92	0.85	0.42	<b>0.35</b>
15%	CCR	91.99	98.35	99.12	99.45	<b>99.61</b>
	SP	91.63	98.27	99.63	99.44	<b>99.94</b>
	FAR	8.48	1.75	0.93	0.58	<b>0.42</b>
20%	CCR	91.02	98.52	98.93	99.07	<b>99.52</b>
	SP	90.64	98.48	99.54	99.06	<b>99.93</b>
	FAR	9.51	1.56	1.13	0.98	<b>0.51</b>
25%	CCR	90.3	98.02	98.74	98.6	<b>99.41</b>
	SP	89.89	97.95	99.41	98.56	<b>99.92</b>
	FAR	10.27	2.1	1.33	1.49	<b>0.62</b>
30%	CCR	90.95	96.83	98.61	97.82	<b>99.32</b>
	SP	90.63	96.7	99.37	97.74	<b>99.93</b>
	FAR	9.58	3.35	1.47	2.31	<b>0.72</b>
35%	CCR	90.02	95.22	98.34	98.48	<b>99.16</b>
	SP	89.65	95.02	99.28	98.54	<b>99.91</b>
	FAR	10.57	5.06	1.76	1.61	<b>0.89</b>
40%	CCR	90.52	94.18	98	97.75	<b>98.96</b>
	SP	90.22	94.05	99.22	98.01	<b>99.91</b>
	FAR	10.04	6.17	2.1	2.38	<b>1.11</b>
45%	CCR	89.69	95.9	97.73	96.59	<b>98.7</b>
	SP	89.32	96.05	99.16	96.84	<b>99.86</b>
	FAR	10.92	4.35	2.4	3.61	<b>1.38</b>
50%	CCR	89.39	95.28	97.55	95.08	<b>98.42</b>
	SP	89.09	95.52	99.11	95.32	<b>99.85</b>
	FAR	11.23	5	2.6	5.22	<b>1.67</b>

Table 1). In order to perform systematic quantitative evaluation, the evaluation criteria first had to be established. For slices with the tumor image in each case, three experts delineated the tumor outline. The intersection area was taken as the standard contour, and a standard 3D tumor contour was further established in combination with the standard contour in each slice. In the subsequent experiments, systematic performance evaluation and quantitative analysis was conducted for each case based on its standard 3D tumor contour. Figure 7 shows the method of establishing a standard tumor contour in a single slice. Figures 7(a)–7(c) are the tumor outlines delineated by three experts, and Figure 7(d) is the contour formed by the intersection area of the preceding three figures, that is, the standard tumor contour of that slice. The standard tumor contour in each section will be stacked in the experiment and used to construct a standard 3D tumor model. This will be used to evaluate the efficacy of various segmentation methods. In addition, the proposed system is targeted against the 3D ROI from actual cases for segmentation. The ROI region can be first delineated by the user using any 2D MRI image before the construction of a 3D ROI model. This approach facilitates the calculation of subsequent

TABLE 3: Accuracy, specificity, and false alarm rate (%) of the different algorithms and noise densities in  $5 \times 5$  mask blurred computer-simulated images.

Noise density		Methods				
		ACM	LSM	SLSM	3D-LSM	3D-SLSM
5%	CCR	87.65	98.88	98.94	99.18	<b>99.43</b>
	SP	87.01	98.81	98.88	99.13	<b>99.9</b>
	FAR	13.08	1.19	1.13	0.87	<b>0.6</b>
10%	CCR	89.85	99.01	99.39	99.36	<b>99.38</b>
	SP	89.37	98.96	99.72	99.34	<b>99.9</b>
	FAR	10.74	1.04	0.65	0.68	<b>0.66</b>
15%	CCR	90.57	99.07	99.34	99.28	<b>99.35</b>
	SP	90.13	99.02	99.7	99.25	<b>99.92</b>
	FAR	9.98	0.99	0.7	0.77	<b>0.68</b>
20%	CCR	90.37	99.11	99.25	99.29	<b>99.39</b>
	SP	89.93	99.07	99.69	99.28	<b>99.86</b>
	FAR	10.2	0.94	0.8	0.75	<b>0.65</b>
25%	CCR	89.71	99.06	99.16	99.38	<b>99.29</b>
	SP	89.26	99.05	99.7	99.42	<b>99.91</b>
	FAR	10.9	0.99	0.89	0.65	<b>0.75</b>
30%	CCR	88.68	99.05	99.14	99.26	<b>99.15</b>
	SP	88.17	99.07	99.7	99.3	<b>99.91</b>
	FAR	11.98	1.01	0.92	0.79	<b>0.9</b>
35%	CCR	88.08	99.12	99.01	99.34	<b>98.94</b>
	SP	87.62	99.22	99.73	99.56	<b>99.96</b>
	FAR	12.62	0.93	1.04	0.7	<b>1.12</b>
40%	CCR	89.18	98.96	98.89	99.25	<b>99.31</b>
	SP	88.77	99.14	99.68	99.45	<b>99.81</b>
	FAR	11.46	1.1	1.17	0.8	<b>0.73</b>
45%	CCR	89.34	98.91	98.8	99.24	<b>99.27</b>
	SP	88.94	99.14	99.72	99.63	<b>99.74</b>
	FAR	11.29	1.16	1.27	0.8	<b>0.77</b>
50%	CCR	88	98.8	98.55	99.01	<b>99.08</b>
	SP	87.55	99.21	99.68	99.46	<b>99.76</b>
	FAR	12.71	1.27	1.53	1.05	<b>0.97</b>

segmentation algorithms. Refer to Figure 8 for the method and process of establishing the 3D ROI.

3.3. *Systematic Evaluation Methods.* The correct classification rate (CCR), specificity (SP), and false alarm rate (FAR) are commonly used evaluation indices in a variety of medical-aided systems. The closer the values of CCR and SP to 100%, the more accurate will be the detection results of the system. However, FAR is a marker of errors that are detected from the system, and a lower percentage indicates a better performance. This work calculates the above evaluation indices based on the standard 3D tumor contour. In addition to evaluating the 3D-SLSM performance, it also compares different algorithms. Each evaluation index can be calculated by the following equations:

$$\text{correct classification rate (CCR)} = \frac{\text{TNP} + \text{TNN}}{N}, \quad (8)$$

$$\text{specificity (SP)} = \frac{\text{TNN}}{\text{TNN} + \text{FPN}}, \quad (9)$$

$$\text{false alarm rate (FAR)} = \frac{\text{FPN}}{N_n}, \quad (10)$$

where  $N$  represents the total number of pixels in the 3D ROI and  $N_n$  represents the total number of pixels outside the standard tumor. TPN (true positive number) represents the number of pixels within the standard tumor and still remained inside the tumor after segmentation, FPN (false positive number) represents the number of pixels that were outside the standard tumor, but which were resulted as inside the tumor after segmentation, TNN (true negative number) represents the number of pixels that were initially outside the standard tumor and still remained outside the tumor after segmentation, and FNN (false negative number) represents the number of pixels that were initially inside the standard tumor, but which were resulted outside the tumor after segmentation. Generally, TPN and TNN represent the numbers of pixels segmented correctly, while FPN and FNN are the numbers of pixels wrongly segmented.

## 4. Experimental Results

*4.1. Establishment and Comparison of 3D Tumor Segmentation Models for Computer-Simulated Images.* In this section, the computer-simulated images that were generated in Figure 5 will be used to carry out different 3D tumor segmentation methods, and the results will be compared after the segmentation. Figures 9–11 demonstrate the segmentation results of computer-simulated images with three different levels of blurring and noise. From the results shown in Figures 9–11, the segmentation of 3D-SLSM from computer-simulated images, which were generated with three different levels of blurring and noise, is the nearest to standard tumors when compared with that of the existing methods (ACM, LSM, SLSM, and 3D-LSM).

*4.2. Quantitative Evaluation and Comparison for Computer-Simulated Images.* The use of computer-simulated tumor images aided in carrying out accurate quantitative analysis as the tumor region is correctly defined. Tables 2–4 show the accuracy, specificity, and the false alarm rates when a 3D tumor segmentation is carried out on computer-simulated images that have undergone three different levels of blurring ( $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$ ) and noise densities. From the quantitative data, we see that 3D-SLSM has the best performance in accuracy, specificity, and false alarm rates in the experimental results (Tables 2 and 3), where image blurring was carried out using different masks ( $3 \times 3$  and  $5 \times 5$ ). However, when the image underwent the maximum blurring ( $7 \times 7$ ), 2D-SLSM and 3D-SLSM demonstrated similar performances in specificity. The overly blurred images benefitted 3D-SLSM in considering the correlation between sections to be weakened. In the same way, with maximum blurring ( $7 \times 7$ ) and a noise density of above 45%, 3D-LSM and 3D-SLSM delivered similar performances with regard to accuracy (as shown in Table 4). This was due to the 3D shape-weighted value not being able to carry out its function in calculation when the image is overly blurred and has too much noise.

TABLE 4: Accuracy, specificity, and false alarm rate (%) of the different algorithms and noise densities in  $7 \times 7$  mask blurred computer-simulated images.

Noise density		Methods				
		ACM	LSM	SLSM	3D-LSM	3D-SLSM
5%	CCR	88.22	98.5	99.16	98.87	<b>99.21</b>
	SP	87.65	98.42	<b>99.63</b>	98.82	99.25
	FAR	12.48	1.59	0.89	1.2	<b>0.84</b>
10%	CCR	88.67	98.67	99.12	98.96	<b>99.23</b>
	SP	88.14	98.62	<b>99.66</b>	98.93	99.29
	FAR	11.99	1.4	0.93	1.11	<b>0.81</b>
15%	CCR	87.42	98.74	99.09	98.99	<b>99.19</b>
	SP	86.82	98.69	<b>99.65</b>	98.97	99.27
	FAR	13.32	1.33	0.96	1.06	<b>0.85</b>
20%	CCR	87.72	98.88	99.04	99.13	<b>99.3</b>
	SP	87.15	98.85	<b>99.68</b>	99.15	99.43
	FAR	13	1.19	1.02	0.92	<b>0.74</b>
25%	CCR	87.89	98.86	98.99	99.11	<b>99.23</b>
	SP	87.35	98.87	<b>99.69</b>	99.17	99.44
	FAR	12.83	1.2	1.07	0.94	<b>0.81</b>
30%	CCR	86.1	98.91	98.96	99.15	<b>99.23</b>
	SP	85.43	98.95	<b>99.71</b>	99.23	99.46
	FAR	14.72	1.16	1.11	0.9	<b>0.82</b>
35%	CCR	86.27	98.91	98.87	99.19	<b>99.21</b>
	SP	85.63	99.03	99.72	99.36	99.54
	FAR	14.54	1.15	1.2	0.86	<b>0.84</b>
40%	CCR	86.43	99	98.74	99.2	<b>99.21</b>
	SP	85.79	99.18	<b>99.73</b>	99.44	99.64
	FAR	14.37	1.05	1.34	<b>0.84</b>	<b>0.84</b>
45%	CCR	84.41	98.94	98.63	<b>99.16</b>	99.13
	SP	83.7	99.24	<b>99.75</b>	99.57	99.73
	FAR	16.51	1.12	1.45	<b>0.89</b>	0.92
50%	CCR	84.29	98.94	98.52	99.1	<b>99.12</b>
	SP	83.55	99.37	<b>99.78</b>	99.66	99.66
	FAR	16.63	1.12	1.57	0.95	<b>0.93</b>

*4.3. Establishment and Comparison of 3D Tumor Segmentation Models for Breast MRI Cases.* In this section, we use ACM, traditional LSM, shape-based LSM (SLSM), 3D-LSM, and the 3D-SLSM proposed in this paper to perform 3D tumor segmentation on the experimental cases. The resulting 3D tumor contour matrix is converted to a 3D surface (3D tumor segmentation model) via the 3DSR process of MCIS in order to observe and compare the results. The first case has a tissue type of both fat and glandular, and both the breast and tumor sizes are medium. The 3D tumor segmentation results of this case that were obtained with the various methods are shown in Figure 12. The second case has a tissue type of relatively more glandular, and both the breast and tumor sizes are relatively small. The segmentation results of this case are shown in Figure 13. The third case has a tissue type of relatively more fat, and both the breast and tumor sizes are relatively large. The segmentation results of this case are shown in Figure 14. In the experimental results of the three cases (Figures 12–14), figure (a) is the standard 3D tumor model, the establishment process of which is

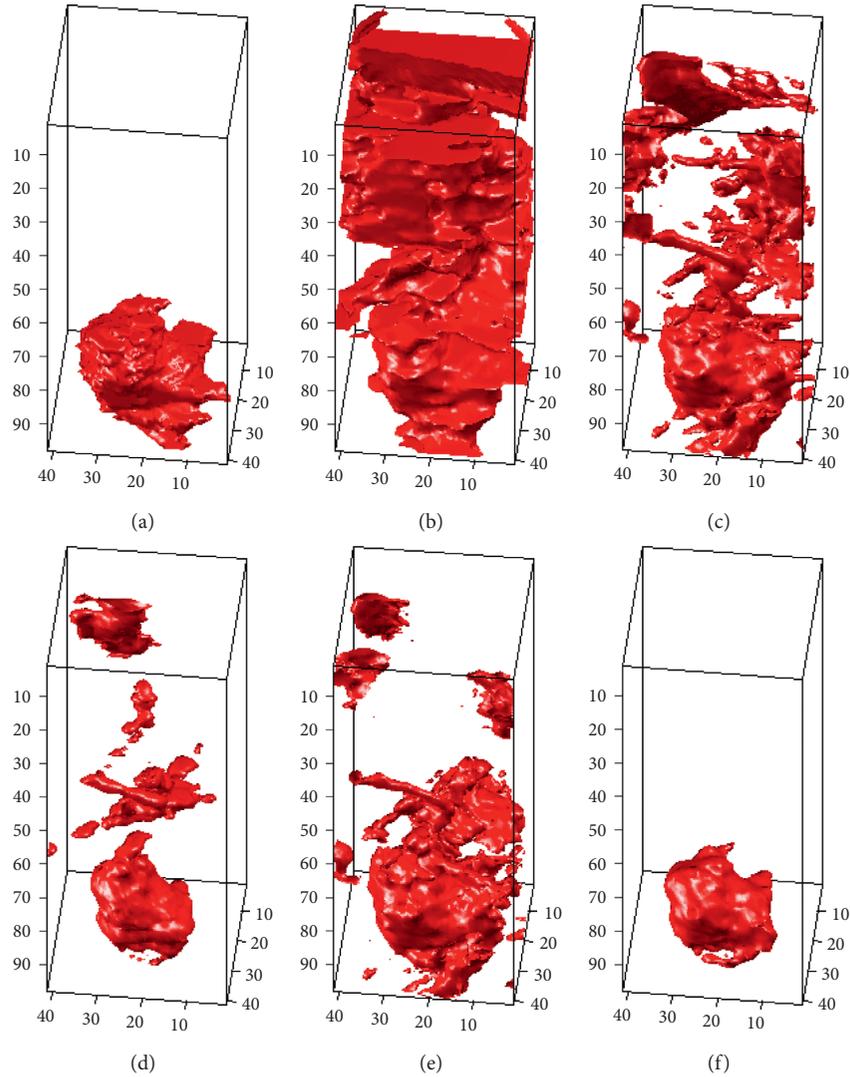


FIGURE 12: 3D tumor segmentation results of Case 1 obtained with different algorithms: (a) standard; (b) ACM; (c) LSM; (d) SLSM; (e) 3D-LSM; (f) 3D-SLSM.

described in Section 3.1; figures (b–f) are the 3D tumor segmentation models obtained by ACM, LSM, SLSM, 3D-LSM, and 3D-SLSM, respectively. The following conclusions can be drawn from these 3D tumor segmentation results. (1) When 2D segmentation techniques evolve into 3D techniques, the application of 3D segmentation gives good performance because the upper and lower slices are connected. (2) Whether it is 2D or 3D segmentation technology, combining shape characteristics can result in better performance. (3) The results in Figures 12–14 show that the 3D tumor segmentation model generated by the 3D-SLSM proposed in this paper is the one that is most consistent with the standard 3D tumor model.

**4.4. Quantitative Evaluation and Comparison for Breast MRI Cases.** This section describes the quantitative analysis based on standard 3D tumor contours obtained from actual MRI cases (Figures 12(a), 13(a), and 14(a)). We compared the

performance of 3D-SLSM with some competitive methods such as ACM, LSM, SLSM, and 3D-LSM. The construction methods employed for standard 3D tumor contours were described in Section 3.2. In addition to using the numbers of TPN, FPN, TNN, FNN,  $N_p$ ,  $N_n$ , and  $N$  to show the analysis results, three evaluation markers (CCR, SP, and FAR) were calculated to facilitate the comparison of the system performance. Here, the newly added  $N_p$  represents the total number of pixels inside standard tumors, and represents the actual size of the tumor. Table 5 shows the various pixel quantities when different methods were employed to perform segmentation in different cases. Equations (8)–(10) were employed to perform the further calculation of three evaluation indicators that are commonly used in medical auxiliary systems, CCR, SP, and FAR (as shown in Table 6). Combining the observation in Tables 5 and 6, we can conclude that (1) LSM has a better performance than ACM in 3D tumor segmentation, and this confirms the appropriateness when LSM was chosen as a foundation in the proposed method.

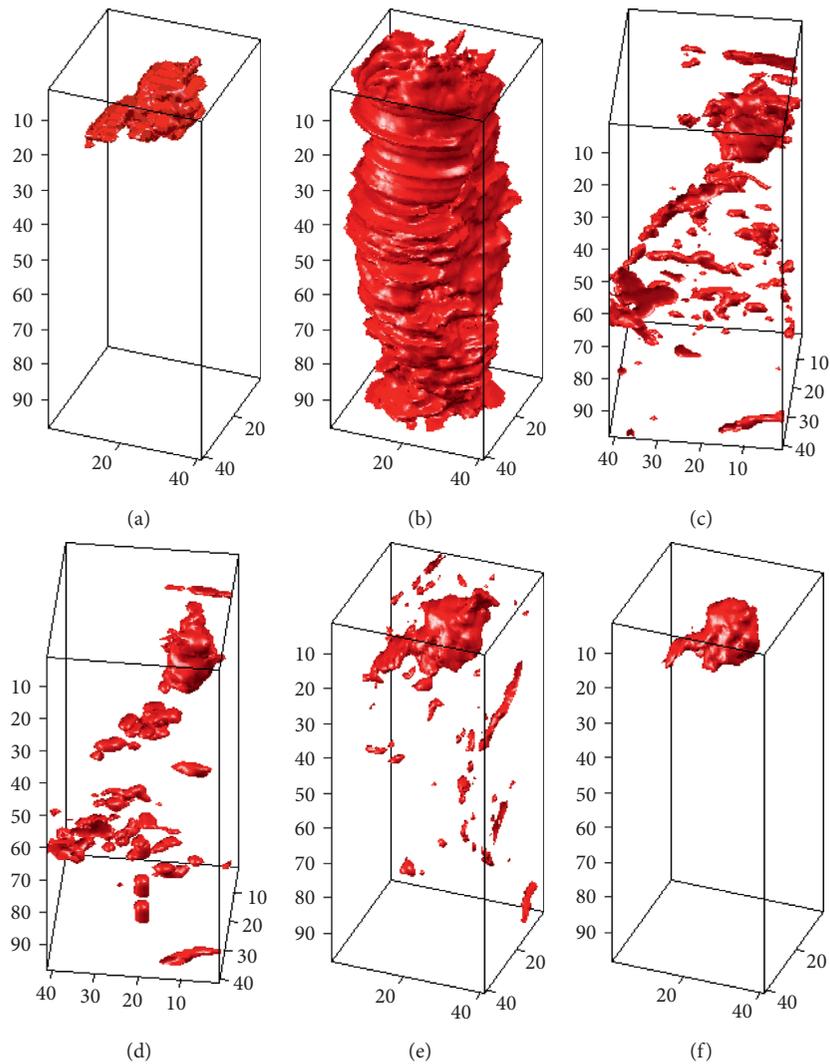


FIGURE 13: 3D tumor segmentation results of Case 2 obtained with different algorithms: (a) standard; (b) ACM; (c) LSM; (d) SLSM; (e) 3D-LSM; (f) 3D-SLSM.

(2) The conversion of traditional 2D algorithms into 3D algorithms can increase the accuracy of 3D segmentation. (3) A combination of shape characteristics can similarly increase segmentation accuracy. (4) Converting traditional 2D-LSM into 3D algorithms and performing a simultaneous combination with shape characteristics into 3D-SLSM can obtain higher CCR and SP, as well as a lower FAR.

These quantitative results verified that 3D-SLSM can effectively eliminate background noise, so contours that are obtained from segmentation are closer to actual tumor boundaries, and demonstrate the feasibility of 3D-SLSM in clinical applications.

## 5. Conclusions

3D medical image segmentation can provide 3D information of lesions or organs, and therefore is an important auxiliary tool for medical diagnosis. However, existing 3D image segmentation techniques still have some shortcomings. Although many 2D image segmentation methods have been

proven to have good results, the overlaying of results from 2D segmentation to carry out 3D image segmentation will not only result in computational burden but also poor results owing to the lack of association between adjacent sections. In order to obtain more precise 3D segmentation results and improve computational efficiency, this paper proposed an innovative 3D medical image segmentation method, which we call the 3D shape-weighted level set method (3D-SLSM). The proposed method can carry out the precise segmentation of tumors from 3D medical images. During the development of 3D-SLSM, the 2D segmentation technology was first evolved to 3D in order to facilitate the simultaneous operation of 3D MRI, with pixels associated with each other in the three coordinate axes ( $X$ ,  $Y$ , and  $Z$ ). Furthermore, since medical images often contain multiple different tissues types, the segmentation results of most of the 3D segmentation algorithms will be affected largely by errors and/or noise. Therefore, 3D-SLSM adds the 3D shape-weighted value in each iterative process according to the change in volume, which not only accelerates convergence

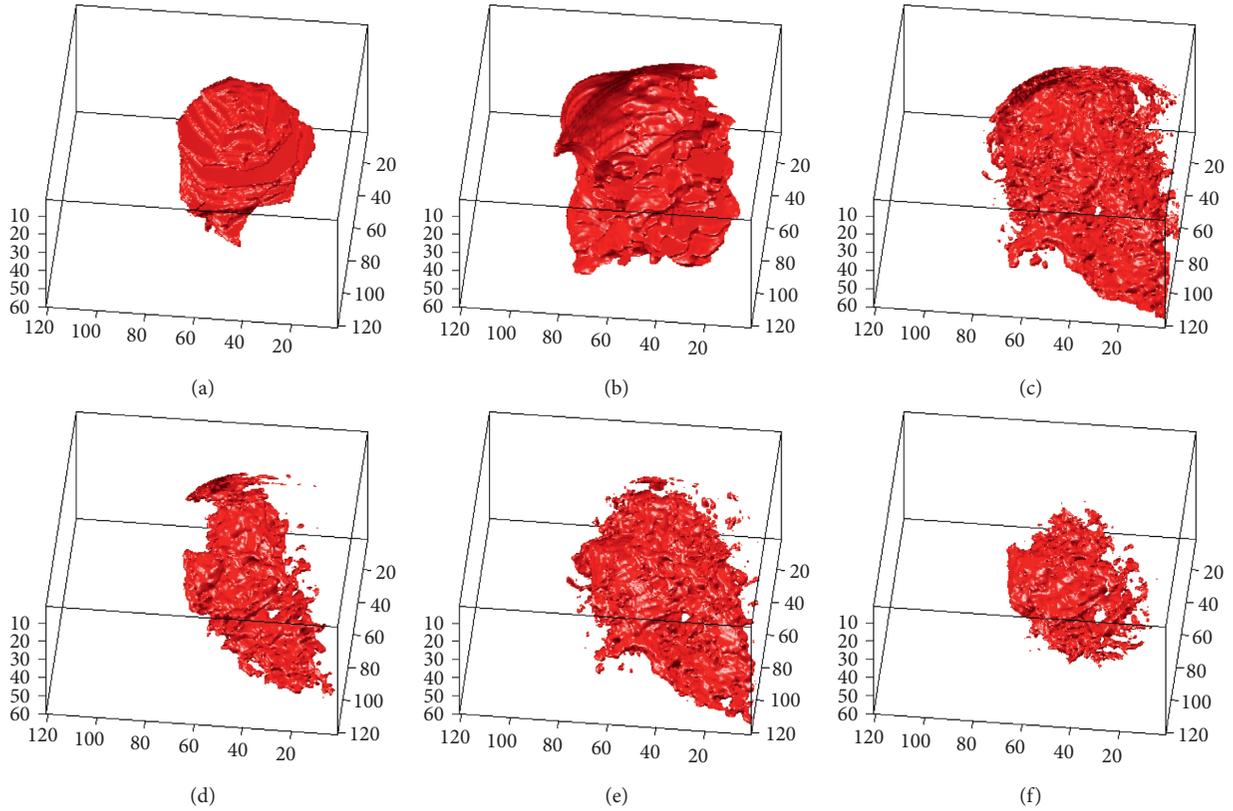


FIGURE 14: 3D tumor segmentation results of Case 3 obtained with different algorithms: (a) standard; (b) ACM; (c) LSM; (d) SLSM; (e) 3D-LSM; (f) 3D-SLSM.

TABLE 5: TPN, FPN, TNN, FNN,  $N_p$ ,  $N_n$ , and  $N$  numbers of the different algorithms for each case in units of pixel.

	TPN	FPN	TNN	FNN	$N_p$	$N_n$	$N$
<i>Case 1</i>							
ACM	6768	44035	113364	571	7339	157399	164738
LSM	6695	15600	141799	644	7339	157399	164738
SLSM	5004	3074	154325	2335	7339	157399	164738
3D-LSM	6926	11959	145440	413	7339	157399	164738
3D-SLSM	6006	518	156881	1333	7339	157399	164738
<i>Case 2</i>							
ACM	5648	64123	93276	1691	2784	161954	164738
LSM	2601	4461	157493	183	2784	161954	164738
SLSM	2122	2156	159798	662	2784	161954	164738
3D-LSM	2530	2421	159533	254	2784	161954	164738
3D-SLSM	2173	253	161701	611	2784	161954	164738
<i>Case 3</i>							
ACM	57877	102006	718123	454	58331	820129	878460
LSM	55432	60356	759773	2899	58331	820129	878460
SLSM	37814	18512	801617	20517	58331	820129	878460
3D-LSM	48647	37622	782507	9684	58331	820129	878460
3D-SLSM	35338	11217	808912	22993	58331	820129	878460

TABLE 6: Accuracy, specificity, and false alarm rate (%) of the different algorithms in each case.

	CCR	SP	FAR
<i>Case 1</i>			
ACM	72.92	72.02	27.98
LSM	90.14	90.09	9.91
SLSM	96.72	98.05	1.95
3D-LSM	92.49	92.40	7.60
3D-SLSM	<b>98.88</b>	<b>99.67</b>	<b>0.33</b>
<i>Case 2</i>			
ACM	60.05	59.26	39.59
LSM	97.18	97.25	2.75
SLSM	98.29	98.67	1.33
3D-LSM	98.38	98.51	1.49
3D-SLSM	<b>99.48</b>	<b>99.84</b>	<b>0.16</b>
<i>Case 3</i>			
ACM	88.34	87.56	12.44
LSM	92.80	92.64	7.36
SLSM	95.56	97.74	2.26
3D-LSM	94.61	95.41	4.59
3D-SLSM	<b>96.11</b>	<b>98.63</b>	<b>1.37</b>

and eliminates background noise but also brings the segmented contour closer to the actual outline of the tumor margin.

To evaluate the accuracy of 3D-SLSM, we use breast MRI cases and computer-simulated images to demonstrate 3D

tumor segmentation results. The actual cases aid in observing feasibility for various methods in clinical applications and the use of computer-simulated tumor images aided in carrying out accurate quantitative analysis as the tumor region is correctly defined. For the quantitative analysis to be fairer and complete, we considered three important influencing factors (the breast volume, the tumor size, and the breast tissue type) in image segmentation for actual cases, and made serial rational arrangements for the level of noise density and blurring intensity in computer-simulated cases. As the raw results from 3D segmentation are in the form of a 3D point matrix, in order to facilitate visual observation and comparison, the 3D point matrix after segmentation was first processed by MCIS to construct a 3D tumor image model. Subsequently, the standard tumor model was used for quantitative evaluation in order to validate the performance of various algorithms. Finally, the accuracy and error rates of conventional ACM, 2D, and 3D-LSM, as well as other methods were compared. The experimental results demonstrate that the 3D-SLSM developed in this study is not only more accurate than existing methods and has less noise after segmentation, but it also has the highest accuracy and lowest false alarm rate when compared with the standard tumor model. However, it is worth noting that 3D-SLSM benefits as the overall 3D calculations decrease when the level of blurring is high, resulting in its accuracy being comparable to 2D-SLSM. When the level of blurring and noise density is simultaneously high, the 3D shape-weighted value cannot carry out its function, causing a decrease in specificity. The results of this paper may be used in the future to aid clinical diagnosis, tracking of lesions, surgical guidance, 3D shape-feature extraction, and other research.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The work was supported by the Ministry of Science and Technology, Taiwan, under Grant no. MOST 103-2221-E-167-025. The authors would like to thank Dr. Hsian-He Hsu, Department of Radiology, Tri-Service General Hospital in Taiwan, for providing cases based on the research collaboration project and relevant medical specifications.

## References

- [1] Q. Ge, L. Xiao, and Z. H. Wei, "Active contour model for simultaneous MR image segmentation and denoising," *Digital Signal Processing*, vol. 23, no. 4, pp. 1186–1196, 2013.
- [2] D. A. Coppa and R. G. Sanfelice, "A zero-crossing detection algorithm for robust simulation of hybrid systems jumping on surfaces," *Simulation Modelling Practice and Theory*, vol. 68, pp. 1–17, 2016.
- [3] S. Sarkar and S. Das, "Multilevel image thresholding based on 2D histogram and maximum tsallis entropy—a differential evolution approach," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4788–4797, 2013.
- [4] D. Zhang, Y. Liu, Y. Yang, M. Xu, Y. Yan, and Q. Qin, "A region-based segmentation method for ultrasound images in HIFU therapy," *Medical Physics*, vol. 43, no. 6, pp. 2975–2989, 2016.
- [5] G. Hamarneh and X. Li, "Watershed segmentation using prior shape and appearance knowledge," *Image and Vision Computing*, vol. 27, no. 1-2, pp. 59–68, 2009.
- [6] T. Brox and J. Weickert, "Level set segmentation with multiple regions," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 3213–3218, 2006.
- [7] J. Shi, B. Sahiner, H.-P. Chan et al., "Characterization of mammographic masses based on level set segmentation with new image features and patient information," *Medical Physics*, vol. 35, no. 1, pp. 280–290, 2008.
- [8] D. A. Rajon and W. E. Bolch, "Marchingcube algorithm, review and trilinear interpolation adaptation for image-based dosimetric models," *Computerized Medical Imaging and Graphics*, vol. 27, no. 5, pp. 411–435, 2003.
- [9] S. C. Yang, C. Y. Yu, C. J. Lin, H. Y. Lin, and C. Y. Lin, "Reconstruction of three-dimensional breast-tumor model using multispectral gradient vector flow snake method," *Journal of Applied Research and Technology*, vol. 13, no. 2, pp. 279–290, 2015.
- [10] P. Gu, W. M. Lee, M. A. Roubidoux, J. Yuan, X. Wang, and P. L. Carson, "Automated 3D ultrasound image segmentation to aid breast cancer image interpretation," *Ultrasonics*, vol. 65, pp. 51–58, 2016.
- [11] L. E. Gangsei and J. Kongsro, "Automatic segmentation of computed tomography (CT) images of domestic pig skeleton using a 3D expansion of Dijkstra's algorithm," *Computers and Electronics in Agriculture*, vol. 121, pp. 191–194, 2016.
- [12] E. S. Rebouças, R. M. Sarmento, and P. P. Rebouças Filho, "3D adaptive balloon active contour: method of segmentation of structures in three dimensions," *IEEE Latin America Transactions*, vol. 13, no. 1, pp. 195–203, 2015.
- [13] W. Liao, K. Rohr, C.-K. Kang, Z.-H. Cho, and S. Wörz, "Automatic 3D segmentation and quantification of lenticulostriate arteries from high-resolution 7 Tesla MRA images," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 400–413, 2016.
- [14] H. Hemmati, A. Kamli-Asl, A. Talebpour, and S. Shirani, "Semi-automatic 3D segmentation of carotid lumen in contrast-enhanced computed tomography angiography images," *Physica Medica*, vol. 31, no. 8, pp. 1098–1104, 2015.
- [15] A. Bartesaghi, G. Sapiro, and S. Subramaniam, "An energy-based three-dimensional segmentation approach for the quantitative interpretation of electron tomograms," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1314–1323, 2005.
- [16] S. Osher and J. A. Sethian, "Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.
- [17] T. S. Newman and H. Yi, "A survey of the marching cubes algorithm," *Computers & Graphics*, vol. 30, no. 5, pp. 854–879, 2006.
- [18] J. Kim, W. Cai, S. Eberl, and D. Feng, "Real-time volume rendering visualization of dual-modality PET/CT images with interactive fuzzy thresholding segmentation," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 2, pp. 161–169, 2007.
- [19] F. Chang, C.-J. Chen, and C.-J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Computer*

*Vision and Image Understanding*, vol. 93, no. 2, pp. 206–220, 2004.

- [20] Y. Dai, J. Zheng, Y. Yang, D. Kuai, and X. Yang, “Volume-rendering-based interactive 3D measurement for quantitative analysis of 3D medical images,” *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 804573, 9 pages, 2013.
- [21] S. C. Yang and C. J. Hsu, “3D tumor segmentation in breast MRIs using 3D modified active contour method,” *International Journal of Computer & Software Engineering*, vol. 2, no. 119, p. 8, 2017.

## Research Article

# Strabismus Recognition Using Eye-Tracking Data and Convolutional Neural Networks

Zenghai Chen,<sup>1</sup> Hong Fu ,<sup>1</sup> Wai-Lun Lo,<sup>1</sup> and Zheru Chi<sup>2</sup>

<sup>1</sup>Department of Computer Science, Chu Hai College of Higher Education, 80 Castle Peak Road, Castle Peak Bay, Tuen Mun, NT, Hong Kong

<sup>2</sup>Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Correspondence should be addressed to Hong Fu; hongfu@chuhai.edu.hk

Received 25 October 2017; Revised 7 January 2018; Accepted 5 February 2018; Published 26 April 2018

Academic Editor: Weide Chang

Copyright © 2018 Zenghai Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Strabismus is one of the most common vision diseases that would cause amblyopia and even permanent vision loss. Timely diagnosis is crucial for well treating strabismus. In contrast to manual diagnosis, automatic recognition can significantly reduce labor cost and increase diagnosis efficiency. In this paper, we propose to recognize strabismus using eye-tracking data and convolutional neural networks. In particular, an eye tracker is first exploited to record a subject's eye movements. A gaze deviation (GaDe) image is then proposed to characterize the subject's eye-tracking data according to the accuracies of gaze points. The GaDe image is fed to a convolutional neural network (CNN) that has been trained on a large image database called ImageNet. The outputs of the full connection layers of the CNN are used as the GaDe image's features for strabismus recognition. A dataset containing eye-tracking data of both strabismic subjects and normal subjects is established for experiments. Experimental results demonstrate that the natural image features can be well transferred to represent eye-tracking data, and strabismus can be effectively recognized by our proposed method.

## 1. Introduction

Strabismus is a common ophthalmic disease that can lead to weak 3D perception, amblyopia (termed lazy eye as well), or even blindness if it is not timely diagnosed and well treated [1, 2]. More importantly, it has been shown that strabismus would cause serious psychosocial consequences in both children and adults [3–12]. These adverse consequences include education [5], employment [6], and dating [8]. Many young strabismic patients could be well treated if diagnosis and treatment were taken at their early ages. A preschool child's strabismus has a much larger chance to be cured than that of an adult. Timely diagnosis is thus essential. Traditional strabismus diagnosis methods, for example, cover test, Hirschberg test, and Maddox rod, are manually conducted by professional ophthalmologists. This would make the diagnosis expensive and drive people out of professional examinations consequently. Furthermore, ophthalmologists make decisions according to their experiences, and thus the diagnosis results are subjective. In view of that, we propose

automatic recognition of strabismus in this paper. Automatic recognition of strabismus, which can be termed strabismus recognition as well, would perform strabismus diagnosis without ophthalmologists. As a result, the diagnosis results would be objective, and the diagnosis cost can be significantly reduced. We realize strabismus recognition by exploiting eye-tracking data, which are acquired using an eye tracker. The proposed eye-tracking-based strabismus recognition method allows us to build an objective, noninvasive, and automatic diagnosis system that could be used to carry out strabismus examination in large communities. For instance, we can place the system in a primary school such that the students can take their examinations at any time.

An eye-tracking technique has been successfully applied to solve various problems, for example, object recognition [13], content-based image retrieval [14], attention modeling [15], and image quality assessment [16]. But very little research on the eye-tracking technique for strabismus diagnosis has been reported. People have also proposed to leverage eye-tracking methodology for strabismus examination

[17–20]. Pulido [17] employed the Tobii eye tracker to acquire gaze data to conduct ophthalmic examination including strabismus by calculating the deviation of gaze data. But Pulido proposed a method prototype only in [17]. The author had no real strabismic gaze data to demonstrate the prototype’s performance. Model and Eizenman [18] proposed an eye-tracking-based approach for performing the Hirschberg test, a classical method to measure binocular ocular misalignment. But the performance of their approach was studied with five healthy infants only. The method’s effectiveness for strabismus examination had not been tested. Bakker et al. [19] developed a gaze direction measurement instrument to estimate the strabismus angle. The instrument allows for unrestrained head movements. But only three subjects participated in the experiment. The number of subjects is relatively too small. Furthermore, there is no ground truth available for strabismic subjects. It is hence impossible to comprehensively evaluate the instrument’s performance. In our previous work [20], we developed a system based on the eye-tracking technique to acquire gaze data for strabismus diagnosis. The diagnosis is performed by intuitively analyzing gaze deviations. But the system’s effectiveness is verified by a strabismic subject and a normal subject only. In this paper, we developed a more effective eye-tracking system than that of [20] to acquire gaze data for strabismus classification. Instead of examining strabismus by directly analyzing gaze deviations in previous methods, we explore a machine learning method to realize strabismus classification. One big disadvantage of previous methods is that their accuracy is dramatically affected by every single gaze point. A noisy gaze point would cause an inaccurate examination result. By contrast, a learning method can eliminate the effect of a small number of noisy gaze points by using a large amount of data, so as to generate a more accurate result. Particularly, we leverage convolutional neural networks (CNNs), a powerful deep learning algorithm, to extract features from gaze data for strabismus recognition.

With the rapid developments of deep learning in recent years, the CNN has achieved numerous successes in computer vision and pattern recognition, for example, image classification [21], scene labeling [22], action recognition [23], and speech recognition [24]. With a hierarchical structure of multiple convolution-pooling layers, CNNs can encode abstract features from raw multimedia data. Especially for learning image features, CNNs have shown impressive performances. In our work, CNNs are exploited to generate useful features to characterize eye-tracking data for strabismus recognition. Concretely, a subject is asked to successively fixate on nine points. Meanwhile, the subject’s eye movements are captured by an eye tracker. The eye-tracking data are then represented by a gaze deviation (GaDe) image which is produced according to the fixation accuracies of the subject’s gaze points. After that, the GaDe image is fed to a CNN that has been trained on a large image database called ImageNet [25]. The output vectors of the full connection (FC) layers of a CNN are used as features for representing the GaDe image. Finally, the features are input to a support vector machine (SVM) for strabismus classification. It is expected that the image features of ImageNet learnt

by CNNs would be well transferred to represent eye-tracking data for strabismus recognition. We build a gaze dataset using our eye-tracking system to demonstrate the proposed method’s performance. The dataset is much larger than previously published strabismus datasets.

The rest of this paper is organized as follows. Section 2 describes the methods exploited for strabismus recognition. Section 3 introduces the dataset for experiments and reports experimental results. Section 4 concludes this paper with final remarks. Before ending this introductory section, it is worth mentioning the contributions of this paper as follows:

- (i) We develop an effective eye-tracking system to acquire gaze data for strabismus recognition.
- (ii) We propose a gaze deviation image to characterize eye-tracking data.
- (iii) We exploit convolutional neural networks to generate features for gaze deviation image representation.
- (iv) We demonstrate that natural image features learnt by convolutional neural networks can be well transferred to represent eye-tracking data, and strabismus can be effectively recognized by our method.

## 2. Methodology

*2.1. The Proposed Strabismus Recognition Framework.* Figure 1 shows our proposed framework for strabismus recognition. The recognition procedure is conducted as follows. First of all, the subject is asked to look at nine points respectively shown at different positions on a screen. Meanwhile, an eye tracker mounted below the screen detects the subject’s eye movements and records his or her gaze points. The gaze data recorded by an eye tracker are then exploited to generate three gaze deviation (GaDe) maps, based on the fixation accuracies of left-eye gaze points, right-eye gaze points, and center points of two eyes, respectively. The three maps are combined to form a GaDe image with three maps denoting R, G, and B channels of the image. After that, the GaDe image is fed to a CNN which has been trained on ImageNet, so as to produce a feature vector for representing the GaDe image. Finally, the feature vector is fed to a SVM for classification, and the subject will be classified as strabismic or normal.

It is worth to present the motivations for the use of a CNN and GaDe image in our method before digging into the implementation details. We use the CNN to tackle our problems for two reasons. Firstly, eye-tracking gaze data are difficult to characterize. Up to now, there is still no standard feature for eye-tracking data representation. People have proposed some features such as fixation time and saccade path. But these features are designed for specific tasks. They are not suited for our strabismus recognition problem. Secondly, the CNN is powerful for learning discriminative features from raw images. It has shown state-of-the-art performance for various pattern recognition and image classification problems. We thus expect that the CNN can extract effective features for eye-tracking data representation. Since the CNN is good at extracting image features, we need to convert the

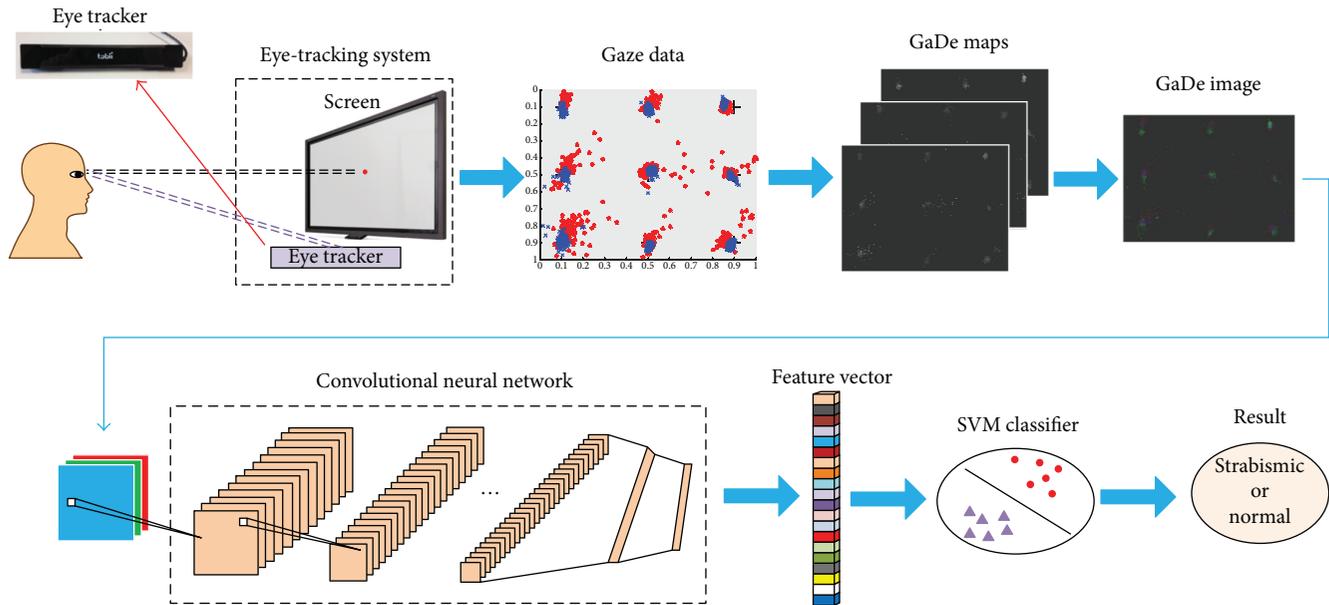


FIGURE 1: The proposed strabismus recognition framework.

raw gaze data to images before feature extraction by the CNN. That is why we propose GaDe images to represent the gaze data. The principle for us to design GaDe images is that the images should be able to well describe the difference between normal data and strabismic data. The details of eye-tracking data acquisition, GaDe image generation, and CNN models used will be presented in the following subsections.

**2.2. Eye-Tracking Data Acquisition.** We use the eye tracker Tobii X2-60 (shown in Figure 1) to acquire eye-tracking gaze data. Tobii X2-60 has a sampling rate of 60 Hz and tracking accuracy of 0.4 degree. Both of the sampling rate and tracking accuracy are high enough to precisely capture strabismic gaze data in our experiments. The eye tracker is adhered below the monitor of a laptop to build our eye-tracking system. The laptop is Lenovo ThinkPad T540p with a  $1920 \times 1080$  screen resolution. The main reason for us to choose a laptop rather than a desktop for building the system is that it is convenient to carry the system in different environments for data acquisition. In order to position gaze points on the screen digitally, we need to define a coordinate system for the screen. The upper-left corner of the screen is set as the origin, the position value of which is  $(0,0)$ , with horizontal line denoting  $x$ -coordinate and vertical line denoting  $y$ -coordinate. The values of the lower-right corner, upper-right corner, and lower-left corner are  $(1,1)$ ,  $(1,0)$ , and  $(0,1)$ , respectively. In other words, both  $x$  and  $y$  lie in interval  $(0,1)$  on the screen. We exploit Tobii Matlab SDK to develop our data acquisition interface.

Calibration needs to be performed before using the eye tracker to acquire gaze data. The purpose of calibration is to teach the eye tracking system the characteristics of the subject, such that the eye tracker can precisely detect the subject's eye movements. During the calibration, the subject is asked to fixate on a number of points displayed on the screen. In terms of the number of points used, we can have

different calibration schemes, for example, one-point, three-point, or nine-point. We adopt a nine-point calibration scheme, as it can provide a high-tracking accuracy. The positions of the nine points on the screen are  $(0.1,0.1)$ ,  $(0.5,0.1)$ ,  $(0.9,0.1)$ ,  $(0.1,0.5)$ ,  $(0.5,0.5)$ ,  $(0.9,0.5)$ ,  $(0.1,0.9)$ ,  $(0.5,0.9)$ , and  $(0.9,0.9)$ . The result would be shown after each calibration. We can start real tracking tests if the calibration accuracy is acceptable. Otherwise, we should recalibrate.

A traditional method for ophthalmologists to examine strabismus is a nine-point method. The nine-point method is to ask the patient to fixate on nine target points at a certain distance in front sequentially. Meanwhile, the ophthalmologist observes the patient's eye movements. This method is able to comprehensively examine the patient's eye movements with rotations at different angles. Therefore, we adopt the same method to develop a gaze data acquisition interface. The nine points' positions are the same to the nine calibration points. Figure 2 shows the nine-point interface. We use a black background, as it helps the subject to concentrate on the target points. A point is comprised by a red inner circle and a white outer circle. The radiuses of the inner circle and outer circle are 15 and 30 pixels, respectively. The points are displayed one by one orderly. The white arrows point out the display order. In a real test, the subject's position is adjusted to make sure that the subject is at a fixed distance (50 cm in our test) from the screen, and the subject's eye level and the screen center are in the same horizontal line. A distance of 50 cm is an optimal distance for the eye tracker Tobii X2-60 to track the subject's eye movements.

Figure 3 shows the procedure of gaze data acquisition. Each time one target point is displayed, the eye tracker records the subject's gaze points of both eyes simultaneously. The next target point would be displayed, if the number of effective gaze pairs acquired exceeds 100, where a gaze pair is defined as the two gaze points of two eyes captured by an eye tracker at one sampling moment and "effective" indicates

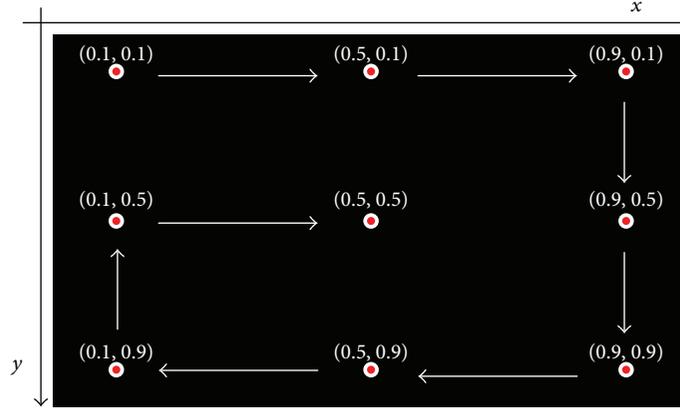


FIGURE 2: The nine-point gaze data acquisition interface.

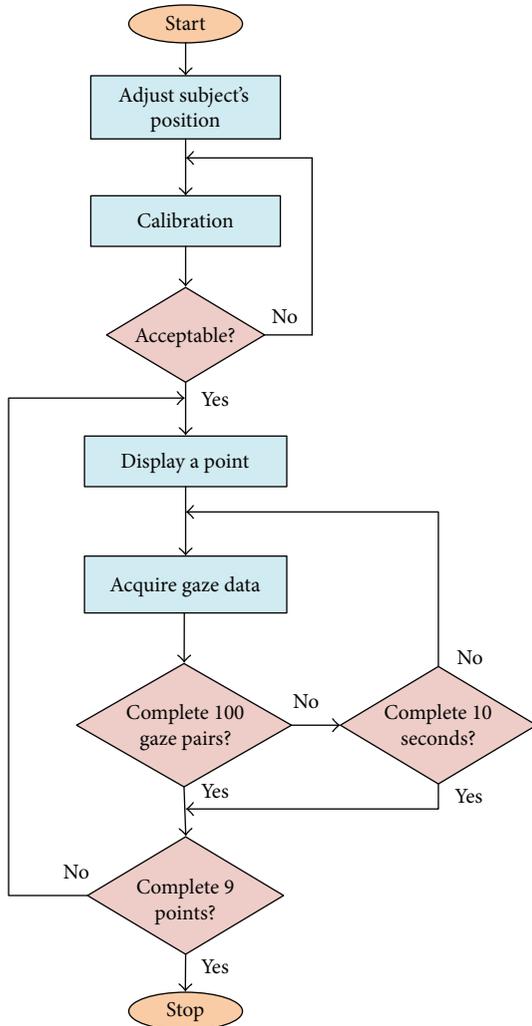


FIGURE 3: The gaze data acquisition procedure.

that at least one gaze point of a gaze pair is located close enough to the target point. That is, the distance between the target point and either gaze point of a gaze pair must be smaller than a threshold (0.05 in this paper) predefined empirically. The distance is defined as Euclidean distance

shown in Section 2.3. It is worth mentioning that for some serious strabismic subjects, it would be sometimes difficult to capture effective gaze points at some target points, in particular the points located at the four corners of the screen, because the strabismic subjects need to rotate the eyeballs to their extreme. In view of that, we let each target point display for at most 10 seconds for gaze data acquisition. The next target point would be displayed after 10 seconds no matter whether or not the system has collected 100 pairs of effective gaze points. Since the sampling rate of our eye tracker is 60 Hz, it would take only two seconds for collecting 100 gaze pairs from normal subjects. Hence, 10 seconds are long enough to capture gaze data for each point.

**2.3. Gaze Deviation Image.** The next step after gaze data acquisition is to generate a GaDe image to characterize the gaze data. To realize that, we need to first calculate three GaDe maps, which will serve as R, G, and B channels of the GaDe image, based on the fixation accuracies of two eyes' gaze points. Let  $g_{ij}$  denote the  $i$ th gaze pair for the  $j$ th target point and  $p_{ij}^l = (x_{ij}^l, y_{ij}^l)$  and  $p_{ij}^r = (x_{ij}^r, y_{ij}^r)$  denote the values of the left-eye gaze point and right-eye gaze point of gaze pair  $g_{ij}$ , where  $1 \leq j \leq 9$  and  $1 \leq i \leq 100$ , and superscripts l and r indicate left and right. Let  $p_j^t = (x_j^t, y_j^t)$  denote the  $j$ th target point's value. Then, we can have the fixation accuracies in terms of Euclidean distance for left-eye gaze point  $p_{ij}^l$  as

$$d_{ij}^l = \sqrt{(x_j^t - x_{ij}^l)^2 + (y_j^t - y_{ij}^l)^2}, \quad (1)$$

and for right-eye gaze point  $p_{ij}^r$  as

$$d_{ij}^r = \sqrt{(x_j^t - x_{ij}^r)^2 + (y_j^t - y_{ij}^r)^2}. \quad (2)$$

We defined the center of the gaze pair  $p_{ij}^l$  and  $p_{ij}^r$  as  $p_{ij}^c = (x_{ij}^c, y_{ij}^c)$ , which can be simply formulated as follows:

$$p_{ij}^c = \left( \frac{x_{ij}^l + x_{ij}^r}{2}, \frac{y_{ij}^l + y_{ij}^r}{2} \right). \quad (3)$$

Then, similar to (1) and (2), the fixation accuracy of the center  $p_{ij}^c$  is calculated as

$$d_{ij}^c = \sqrt{(x_j^t - x_{ij}^c)^2 + (y_j^t - y_{ij}^c)^2}. \quad (4)$$

For one subject, we calculate the fixation accuracies  $d_{ij}^l$ ,  $d_{ij}^r$ , and  $d_{ij}^c$  for all of his or her gaze pairs. Based on the three types of fixation accuracies, three GaDe maps  $M^l$ ,  $M^r$ , and  $M^c$  can be computed, respectively. The map size is equivalent to the input size of a CNN. In this paper, two map sizes (224\*224 and 227\*227) are adopted. The element values of the three maps are derived from three fixation accuracies  $d_{ij}^l$ ,  $d_{ij}^r$ , and  $d_{ij}^c$  of all the gaze pairs. One gaze pair represents one element in the three maps. In particular, the values of the gaze pair  $g_{ij}$  in the three GaDe maps  $M^l$ ,  $M^r$ , and  $M^c$  are respectively calculated as

$$\begin{aligned} v_{ij}^l &= \text{round} \left( \frac{d_{ij}^l}{\max_{(i,j)} d_{ij}^l} \right) * 255, \\ v_{ij}^r &= \text{round} \left( \frac{d_{ij}^r}{\max_{(i,j)} d_{ij}^r} \right) * 255, \\ v_{ij}^c &= \text{round} \left( \frac{d_{ij}^c}{\max_{(i,j)} d_{ij}^c} \right) * 255, \end{aligned} \quad (5)$$

where  $\max(\cdot)$  finds out the maximum value over all gaze pairs and  $\text{round}(\cdot)$  rounds a value to its nearest integer. Equation (5) guarantees that the element values of the three GaDe maps are integers lying in interval (0,255), which is also the value interval of a real digital image. The positions of  $v_{ij}^l$ ,  $v_{ij}^r$ , and  $v_{ij}^c$  in maps  $M^l$ ,  $M^r$ , and  $M^c$  are specified by  $(x_{ij}^l, y_{ij}^l)$ ,  $(x_{ij}^r, y_{ij}^r)$ , and  $(x_{ij}^c, y_{ij}^c)$ , respectively. The elements that are not associated with any gaze points in the three maps are assigned a value of 0. In the GaDe maps, big fixation deviations (far from the target point) possess big values. In other words, inaccurate gaze points play more important roles in a GaDe map. This makes sense, as the prominent difference between strabismic people and normal people is that strabismic people's one eye or even two eyes cannot well fixate on target objects. The three GaDe maps can thus effectively characterize the properties of strabismic gaze data. Generally, normal people's GaDe maps would have only a few bright (large intensity) points far from nine target points, and most relatively dark points are around the target points, while strabismic people's GaDe images usually have a large number of bright points located far from the target points. We combine three GaDe maps to form a GaDe image, with each map representing a color channel of the GaDe image. The GaDe image is then fed to a CNN for feature extraction.

**2.4. Convolutional Neural Networks.** A CNN is a hierarchical architecture that consists of a number of convolution layers and pooling layers. CNNs usually receive raw data, for example, image's pixels, as input and extract increasingly abstract features through hierarchical convolution-pooling

layers. Take color image feature extraction as an example. An image's three-color channels are fed to the first convolution layer of the CNN. The convolution results, called convolution feature maps as well, are then downsampled in the pooling (e.g., max-pooling) layer following the first convolution layer, to generate pooling-feature maps. The pooling-feature maps are further passed to the next convolution layer and then to the pooling layer for processing. After a number of convolution and pooling operations, the feature maps are connected to an output layer through one or more FC layers. The FC layers can be used for classification like a multilayer perceptron, with the output vector representing different classes, or we can employ the outputs of FC layers as a feature vector of the input image and then use a classifier, for example, SVM, to perform classification on the feature vector. The hierarchical convolution and pooling operations make the features extracted by a CNN insensitive to transformation, rotation, and scaling.

We adopt six different CNN models that have been trained on ImageNet to generate features for representing eye-tracking gaze data. We use pretrained CNN models as feature extractors and do not train them using eye-tracking data in our work. There are two main reasons for us to do this. Firstly, we do not have enough eye-tracking data to well train a complicated CNN model. A CNN model may have thousands or even hundreds of thousands of weights that need to be trained. A large dataset is hence necessary to effectively tune so many weights. For instance, the CNN models we adopted have been trained on ImageNet, an image database that contains more than one million training images associated with 1000 classes. For strabismus classification problem, it is difficult to build a large dataset, since not so many strabismic people can be found to participate in the experiments. Actually, only 17 strabismic people participate in our experiments. It is, therefore, impractical to train a CNN model using such a few strabismic gaze data. However, it would be a good idea to employ a pretrained CNN as a feature extractor to generate features for gaze data representation. This will be demonstrated in Section 3. Secondly, the weights of CNN models are tuned using natural images rather than eye-tracking gaze data. We would like to investigate whether or not the information extracted from the natural image domain is applicable to the eye-tracking data domain. It would be meaningful if the features of natural images can be well transferred to represent eye-tracking data, since we would be able to make use of large quantities of natural images in the internet to help generate features, rather than to manually design complicated algorithms to extract features, for eye-tracking data representation. The six CNN models we adopted are named AlexNet [21], VGG-F, VGG-M, VGG-S [26], VGG-16, and VGG-19 [27]. All of them have three FC layers but different numbers of convolution layers. Their differences also lie in input size, number of convolution filters in each layer, max-pooling size, and so on. People can refer to [21, 26, 27] for the architecture details of the six models. The six models have the same three FC layers. The first two FC layers use ReLU [28] transfer function, and the final FC layer adopts Softmax transfer function. For each FC layer, we employ the input vector and output vector of the

transfer function as feature vectors of GaDe images. Then, we can extract in total six feature vectors from three FC layers. The six feature vectors are denoted by  $I_1, I_2, I_3, I_4, I_5,$  and  $I_6$ , and their dimension sizes are 4096, 4096, 4096, 4096, 1000, and 1000, respectively. We will compare the performances of six feature vectors for six CNN models in Section 3. Note that the input size of AlexNet is  $227 \times 227$ , while the input sizes of the other five models are all  $224 \times 224$ . Therefore, the GaDe images need to be resized to  $227 \times 227$  for AlexNet and  $224 \times 224$  for the other five models.

**2.5. Baseline Method.** In order to demonstrate the effectiveness of CNNs for extracting features from gaze data for strabismus recognition, we propose a baseline method for a comparison. The baseline method models the normal gaze points of each target point as a multivariate Gaussian distribution. The parameters (mean vector and covariance matrix) of each Gaussian distribution are calculated using the normal training data. To construct the Gaussian distribution, we represent a gaze pair  $g_{ij}$  by the  $x$ -coordinate differences and  $y$ -coordinate differences between the target point  $p_j^t$  and the pair's two gaze points  $p_{ij}^l$  and  $p_{ij}^r$  as follows:

$$u_{ij} = \begin{bmatrix} x_j^t - x_{ij}^l, y_j^t - y_{ij}^l, x_j^t - x_{ij}^r, y_j^t - y_{ij}^r \end{bmatrix}^T. \quad (6)$$

Then, we can have the Gaussian probability density function for the gaze pair  $g_{ij}$  as

$$p(u_{ij}; \mu_j, \Sigma_j) = \frac{1}{(2\pi)^2 \sqrt{|\Sigma_j|}} \exp\left(-\frac{1}{2} (u_{ij} - \mu_j)^T \Sigma_j^{-1} (u_{ij} - \mu_j)\right), \quad (7)$$

where  $\mu_j$  and  $\Sigma_j$  are the mean vector and covariance matrix of the Gaussian distribution for the  $j$ th target point, respectively.  $\mu_j$  and  $\Sigma_j$  are calculated using the normal training gaze pairs that belong to the  $j$ th target point.  $|\Sigma_j|$  computes the determinant of  $\Sigma_j$ .

The baseline method performs classification as follows. Given the gaze pair  $g_{ij}$ , if its density value in (7) is larger than the threshold  $\alpha_j$ , the gaze pair is classified as normal. If the proportion of normal gaze pairs is larger than the threshold  $\beta_j$ , then the target point is classified as normal for the subject. Otherwise, the target point is classified as strabismic for the subject. If one of the right target points is classified as strabismic, the subject will be finally classified as strabismic. In other words, a normal subject should possess normal fixations on all nine different directions. Once the fixation on one direction is abnormal, the subject will be diagnosed as strabismic. This is reasonable, since some types of strabismus such as incomitant strabismus may fixate poorly on a specific direction only. In medical examination, a subject may be also diagnosed to have strabismus once the ophthalmologist observes that the subject's two eyes do not align at a specific direction. Thresholds  $\alpha_j$  and  $\beta_j$  are learnt using grid search, such that the classification accuracy on the training data is maximized.

### 3. Experiments

**3.1. Eye-Tracking Gaze Data.** We cooperated with Hong Kong Association of Squint and Double Vision Sufferers to collect strabismic data. In total, 17 members of the association suffering from strabismus consented to participate in our experiments. In addition to the 17 strabismic subjects, we invited 25 normal subjects to join in our study. All subjects are adults, with age ranging from 25 to 63, including both male and female. They have been diagnosed by a professional ophthalmologist, and the diagnosis results are used as ground truth in this paper. After ethics approval and informed consent, the 42 subjects followed the data acquisition procedure introduced in Section 2.2 to participate in our experiments, and finally, we collected 42 eye-tracking samples. The 17 strabismic subjects suffer from different types of strabismus (e.g., recessive, intermittent, and manifest) in various severities (e.g., mild, moderate, and severe). Recessive strabismus is only present when binocular vision has been interrupted, such as covering one eye. This type of patients can still maintain fusion. The patients are usually aware of having recessive strabismus after taking examination by an ophthalmologist. Manifest strabismus can be observed while a patient looks at an object binocularly. Intermittent strabismus is a combination of recessive strabismus and manifest strabismus. People suffering from recessive strabismus, intermittent strabismus, and mild manifest strabismus are sometimes difficult to be distinguished from normal people apparently, as their fixation deviations are small, especially for recessive strabismus and intermittent strabismus.

Figure 4 shows some examples of gaze data and GaDe images. The first row displays all gaze points for nine target points in a map, with red \* denoting left gaze points and blue x denoting right gaze points. The second row shows the corresponding GaDe images for the gaze data of the first row. For a better visualization, the GaDe images have been brightened by adding a number (50) to the gaze points' values in the images. The first two columns are two normal samples, one with good fixation (small deviation) and one with relatively poor fixation (large deviation). The other three columns from left to right represent strabismic samples of recessive strabismus, intermittent strabismus, and manifest strabismus, respectively. Note that the colors in the first row represent the left and right gaze points, and the colors in the second row represent the R, G, and B channels of GaDe images. The two main observations can be drawn from Figure 4. Firstly, gaze points with large deviations shown in the first row are highlighted in the corresponding GaDe images, and those with small deviations are inhibited. Therefore, inaccurate gaze points would contribute more in recognizing strabismus using GaDe images. Secondly, the data distributions of normal sample with small deviation and manifest sample are distinctive. They are easy to distinguish. By contrast, the data distributions of normal sample with large deviation and recessive or intermittent sample look similar. It is difficult to distinguish them intuitively. That is why we exploit CNNs to solve the problem. We expect that CNNs as a powerful abstract feature extractor can extract

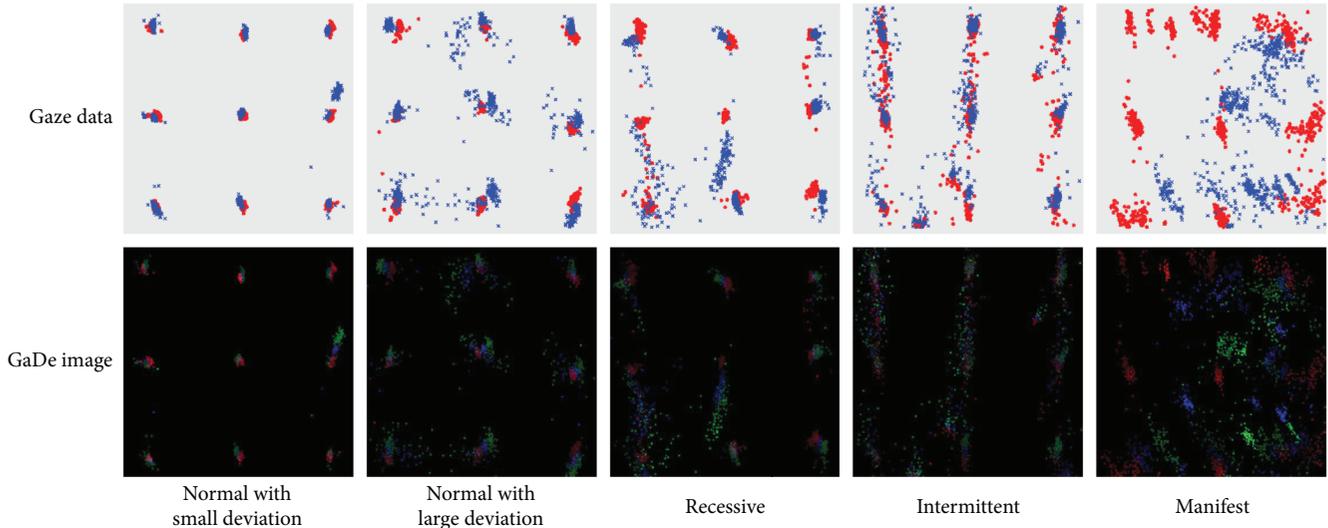


FIGURE 4: Examples of gaze data and corresponding GaDe images, where the first two columns represent normal data with small deviation and large deviation and the third, fourth, and fifth columns represent data of recessive strabismus, intermittent strabismus, and manifest strabismus, respectively.

distinctive features from different samples, so as to effectively classify normal data and strabismic data. It is worth mentioning that we focus on binary strabismus classification rather than recognizing different types of strabismus in this paper. The major reason is that we do not have enough data for each strabismus type at this moment. However, we consider that CNNs could extract useful features from GaDe images of different strabismus types in case that sufficient data are provided, and then the proposed method would be well applied to recognizing strabismus types. We leave this task for future work when we acquire sufficient data for different strabismus types.

**3.2. Experimental Results.** We have in total 42 samples, with 25 normal samples and 17 strabismic samples. A leave-one-out evaluation scheme is adopted. That is, each time one sample is used for testing, and the rest 41 samples are for training. We can thus have 42 different results. The 42 results are averaged to get the final performance. LIBSVM [29] is employed to implement SVM classification. A linear kernel of a SVM is used for both the CNN method and baseline method, and the SVM classifiers are trained using a two-fold cross validation scheme. Table 1 tabulates the classification accuracies of six CNN models (by row) using different feature vectors (by column) extracted from three FC layers. The final column represents the concatenation of all the six feature vectors as a one feature vector. The accuracy of the baseline method is 69.1. As can be seen from Table 1, the features extracted from VGG-S overall perform the best. The highest accuracy (95.2%) is achieved when the feature vector  $I_3$  of VGG-S is used. Feature vectors  $I_1$ ,  $I_2$ ,  $I_3$ , and  $I_4$  outperform feature vectors  $I_5$  and  $I_6$  for most of the cases. One possible reason is that  $I_1$ ,  $I_2$ ,  $I_3$ , and  $I_4$  extracted from the first two FC layers contain richer features than  $I_5$  and  $I_6$ . Note that the concatenation of all feature vectors sometimes obtains lower accuracy than that of some individual feature vectors, as shown in the final column. The most important

TABLE 1: Accuracies (%) of different CNN models. The accuracy of the baseline method is 69.1.

Feature	$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	All
AlexNet	78.6	78.6	76.2	76.2	73.8	76.7	76.2
VGG-F	76.2	76.2	76.2	76.2	78.6	65.1	81.0
VGG-M	88.1	85.7	85.7	85.7	78.6	57.1	78.6
VGG-S	85.7	81.0	78.6	<b>95.2</b>	76.2	79.1	83.3
VGG-16	83.3	81.0	76.2	81.0	76.2	67.4	83.3
VGG-19	81.0	78.6	81.0	81.0	71.4	62.8	83.3

finding from Table 1 is that the features extracted from six CNN models perform much better than the baseline method, except for some cases when the feature vector  $I_6$  is used. This indicates that the CNN features can effectively characterize the GaDe images derived from eye-tracking data. CNN features can be a promising representation of eye-tracking data.

Specificity and sensitivity are two important metrics to measure the performance of a medical classification method. A good method should have high values for both specificity and sensitivity. For our strabismus recognition problem, specificity is defined as the percentage of normal subjects who are correctly classified as normal and sensitivity is defined as the percentage of strabismic subjects who are correctly classified as strabismic. In order to study the specificity and sensitivity of our method, for each CNN model, we select the result with the highest classification accuracy. According to Table 1, the highest accuracies for AlexNet, VGG-F, VGG-M, VGG-S, VGG-16, and VGG-19 are 78.6 ( $I_1$ ), 81.0 (column “All”), 88.1 ( $I_1$ ), 95.2 ( $I_4$ ), 83.3 ( $I_1$ ), and 83.3 (column “All”), respectively. We show the specificity and sensitivity of the six CNN results as well as the baseline method in Figure 5. Evidently, VGG-S possesses the best specificity and sensitivity. Only one normal subject and one strabismic subject are misclassified by VGG-S. The

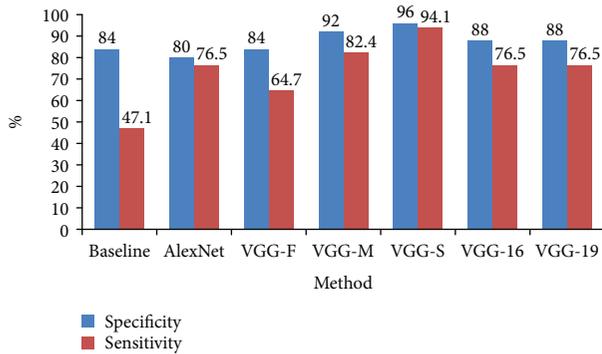


FIGURE 5: Specificity and sensitivity of different methods.

baseline method has a high specificity (84%) but a very low sensitivity (47.1%). This means that the baseline method is insensitive to strabismic data. It tends to classify the data as normal. By contrast, the difference between specificity and sensitivity of CNN features is relatively small, especially for VGG-S. This substantiates two things. Firstly, the proposed GaDe images are able to effectively characterize both normal gaze data and strabismic gaze data. The two types of eye-tracking data can be well separated by GaDe images. Secondly, the natural image features learnt by CNNs can be well transferred to represent GaDe images.

Overall, the experimental results have demonstrated that the proposed method is a promising alternative for strabismus recognition. In the future, the accuracy can be improved in two major ways. One way is to employ more advanced pretrained CNN models for better feature extraction. The other way is to collect more gaze data, especially data of different strabismus types. With sufficient data, we would then be able to fine-tune CNN models, as a result of which CNN models could learn more discriminative features to boost the classification accuracy.

#### 4. Conclusion

In this paper, we first design an eye-tracking system to acquire gaze data from both normal and strabismic people and then propose a GaDe image based on gaze points' fixation deviation to characterize eye-tracking data. Finally, we exploit CNNs that have been trained on a large real image database to extract features from GaDe images for strabismus recognition. Experimental results show that GaDe images are effective for characterizing strabismic gaze data, and CNNs can be a powerful alternative in feature extraction of eye-tracking data. The effectiveness of our proposed method for strabismus recognition has been demonstrated.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### Acknowledgments

This work was supported by a grant from the Hong Kong RGC (Project reference no. UGC/FDS13/E04/14).

#### References

- [1] M. S. Castanes, "Major review: the underutilization of vision screening (for amblyopia, optical anomalies and strabismus) among preschool age children," *Binocular Vision & Strabismus Quarterly*, vol. 18, no. 4, pp. 217–232, 2003.
- [2] S. M. Mojon-Azzi, A. Kunz, and D. S. Mojon, "The perception of strabismus by children and adults," *Graefes' Archive for Clinical and Experimental Ophthalmology*, vol. 249, no. 5, pp. 753–757, 2011.
- [3] J. M. Durnian, C. P. Noonan, and I. B. Marsh, "The psychosocial effects of adult strabismus: a review," *The British Journal of Ophthalmology*, vol. 95, no. 4, pp. 450–453, 2011.
- [4] B. G. Mohny, J. A. McKenzie, J. A. Capo, K. J. Nusz, D. Mrazek, and N. N. Diehl, "Mental illness in young adults who had strabismus as children," *Pediatrics*, vol. 122, no. 5, pp. 1033–1038, 2008.
- [5] O. Uretmen, S. Egrilmez, S. Kose, K. Pamukcu, C. Akkin, and M. Palamar, "Negative social bias against children with strabismus," *Acta Ophthalmologica Scandinavica*, vol. 81, no. 2, pp. 138–142, 2003.
- [6] S. M. Mojon-Azzi and D. S. Mojon, "Strabismus and employment: the opinion of headhunters," *Acta Ophthalmologica*, vol. 87, no. 7, pp. 784–788, 2009.
- [7] M. J. Goff, A. W. Suhr, J. A. Ward, J. K. Croley, and M. A. O'Hara, "Effect of adult strabismus on ratings of official U.S. Army photographs," *Journal of AAPOS*, vol. 10, no. 5, pp. 400–403, 2006.
- [8] S. M. Mojon-Azzi, W. Potnik, and D. S. Mojon, "Opinions of dating agents about strabismic subjects' ability to find a partner," *British Journal of Ophthalmology*, vol. 92, no. 6, pp. 765–769, 2008.
- [9] S. M. Mojon-Azzi and D. S. Mojon, "Opinion of headhunters about the ability of strabismic subjects to obtain employment," *Ophthalmologica*, vol. 221, no. 6, pp. 430–433, 2007.
- [10] S. E. Olitsky, S. Sudesh, A. Graziano, J. Hamblen, S. E. Brooks, and S. H. Shaha, "The negative psychosocial impact of strabismus in adults," *Journal of AAPOS*, vol. 3, no. 4, pp. 209–211, 1999.
- [11] E. A. Paysse, E. A. Steele, K. M. McCreery, K. R. Wilhelmus, and D. K. Coats, "Age of the emergence of negative attitudes toward strabismus," *Journal of AAPOS*, vol. 5, no. 6, pp. 361–366, 2001.
- [12] D. K. Coats, E. A. Paysse, A. J. Towler, and R. L. Dipboye, "Impact of large angle horizontal strabismus on ability to obtain employment," *Ophthalmology*, vol. 107, no. 2, pp. 402–405, 2000.
- [13] T. Toyama, T. Kieninger, F. Shafait, and A. Dengel, "Gaze guided object recognition using a head-mounted eye tracker," in *ETRA '12 Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 91–98, Santa Barbara, CA, USA, March 2012.
- [14] Z. Liang, H. Fu, Y. Zhang, Z. Chi, and D. Feng, "Content-based image retrieval using a combination of visual features and eye tracking data," in *ETRA '10 Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pp. 41–44, Austin, TX, USA, March 2010.
- [15] Z. Liang, H. Fu, Z. Chi, and D. Feng, "Refining a region based attention model using eye tracking data," in *2010 IEEE International Conference on Image Processing*, pp. 1105–1108, Hong Kong, September 2010.

- [16] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971–982, 2011.
- [17] R. A. Pulido, *Ophthalmic Diagnostics Using Eye Tracking Technology*, [M.S. thesis], KTH Royal Institute of Technology, Sweden, 2012.
- [18] D. Model and M. Eizenman, "An automated Hirschberg test for infants," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 103–109, 2011.
- [19] N. M. Bakker, B. A. J. Lenseigne, S. Schutte et al., "Accurate gaze direction measurements with free head movement for strabismus angle estimation," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 11, pp. 3028–3035, 2013.
- [20] Z. Chen, H. Fu, W. L. Lo, and Z. Chi, "Eye-tracking aided digital system for strabismus diagnosis," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2305–2309, Kowloon, October 2015.
- [21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol. 1, pp. 1097–1105, Lake Tahoe, NV, USA, December 2012.
- [22] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1915–1929, 2013.
- [23] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013.
- [24] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1533–1545, 2014.
- [25] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: a large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, Miami, FL, USA, June 2009.
- [26] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: delving deep into convolutional nets," in *Proceedings British Machine Vision Conference 2014*, UK, September 2014.
- [27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, In ICLR.
- [28] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–814, Haifa, Israel, June 2010.
- [29] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

## Research Article

# A Novel Multiscale Gaussian-Matched Filter Using Neural Networks for the Segmentation of X-Ray Coronary Angiograms

Ivan Cruz-Aceves <sup>1</sup>, Fernando Cervantes-Sanchez,<sup>2</sup> and Maria Susana Avila-Garcia <sup>3</sup>

<sup>1</sup>CONACYT - Centro de Investigación en Matemáticas (CIMAT), A.C., Jalisco S/N, Col. Valenciana, 36000 Guanajuato, GTO, Mexico

<sup>2</sup>Centro de Investigación en Matemáticas (CIMAT), A.C., Jalisco S/N, Col. Valenciana, 36000 Guanajuato, GTO, Mexico

<sup>3</sup>Departamento de Estudios Multidisciplinarios, División de Ingenierías, Universidad de Guanajuato, Av. Universidad S/N, Col. Yacatitas, 38944 Yuriria, GTO, Mexico

Correspondence should be addressed to Maria Susana Avila-Garcia; [susana.avila@ugto.mx](mailto:susana.avila@ugto.mx)

Received 2 December 2017; Accepted 4 January 2018; Published 18 April 2018

Academic Editor: Yung-Fu Chen

Copyright © 2018 Ivan Cruz-Aceves et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The accurate and efficient segmentation of coronary arteries in X-ray angiograms represents an essential task for computer-aided diagnosis. This paper presents a new multiscale Gaussian-matched filter (MGMF) based on artificial neural networks. The proposed method consists of two different stages. In the first stage, MGMF is used for detecting vessel-like structures while reducing image noise. The results of MGMF are compared with those obtained using six GMF-based detection methods in terms of the area ( $A_z$ ) under the receiver operating characteristic (ROC) curve. In the second stage, ten thresholding methods of the state of the art are compared in order to classify the magnitude of the multiscale Gaussian response into vessel and nonvessel pixels, respectively. The accuracy measure is used to analyze the segmentation methods, by comparing the results with a set of 100 X-ray coronary angiograms, which were outlined by a specialist to form the ground truth. Finally, the proposed method is compared with seven state-of-the-art vessel segmentation methods. The vessel detection results using the proposed MGMF method achieved an  $A_z = 0.9357$  with a training set of 50 angiograms and  $A_z = 0.9362$  with the test set of 50 images. In addition, the segmentation results using the intraclass variance thresholding method provided a segmentation accuracy of 0.9568 with the test set of coronary angiograms.

## 1. Introduction

Coronary angiography is the standard X-ray imaging procedure used by cardiologists in diagnosing and monitoring vascular abnormalities. In recent years, the development of computational methods to perform image analysis along with computer-aided diagnosis (CAD) has begun to attract more attention. Automatic segmentation of coronary arteries is the main image processing step in cardiology CAD systems and is also a challenging and complex task. The main disadvantages in X-ray angiograms are the uneven illumination and weak contrast between coronary arteries and image background. Given that these two disadvantages generate multimodal histograms, the segmentation task has been commonly addressed in two stages: vessel enhancement also

called detection and binary classification also known as segmentation. The first stage is performed to enhance vessel-like structures from the image background while removing image noise, and the second stage focuses on using a soft classification method to differentiate vessel and nonvessel pixels.

Since the automatic coronary artery segmentation stage is an essential task for a number of CAD systems, different computational methods have been introduced for this purpose. In literature, several techniques for working in the spatial image domain with diverse types of vessel detection strategies have been reported. The most basic strategy is based on mathematical morphology [1–5], where the top-hat operator represents the main idea of these methods, obtaining a low performance to detect small vessels. Another types of vessel detection methods are based on the

eigenvalues of the Hessian matrix to compute a vesselness measure [6–11]. However, because the Hessian matrix is based on the second-order derivative of a Gaussian kernel, the detection performance can be highly sensitive to noise.

On the other hand, based on the idea of a Gaussian kernel, the Gaussian-matched filters (GMF) [12] were introduced and have been successfully applied in different problems such as image registration [13, 14], retinal vessel detection [15–17], and coronary artery detection [18, 19]. The GMF method is based on a Gaussian template matching used for the detection of vessel-like structures at different orientations. GMF works on the assumption that by using a Gaussian curve, the shape of vessel-like structures can be approximated. To form the Gaussian template, four different parameters have to be tuned. The main parameter is the continuous  $\sigma$  value that represents the spread of the intensity profile. The remaining three parameters are defined in the discrete domain. The parameter  $L$  is the length of the vessel segment to be processed,  $T$  is the position where the Gaussian curve trails will cut, and  $\kappa$  represents the number of orientations in the directional filter bank.

The detection performance of the GMF is directly related to the appropriate selection of the parameter values. Chaudhuri et al. [12] establish the following parameters:  $\sigma = 2.0$ ,  $T = 13$ ,  $L = 9$ , and  $\kappa = 12$ . Kang et al. [18–21] modified the number of oriented filters ( $\kappa = 6$ ) along with the  $\sigma$  parameter to adapt the GMF to the coronary artery detection problem. Cinsdikici and Aydin [16] increased the number of oriented filters  $\kappa = 18$ . To avoid empirical values, Al-Rawi et al. [15, 22] proposed a search space to establish a training step by using an exhaustive search and genetic algorithms, respectively. The search space was used to set the  $L$ ,  $T$ , and  $\sigma$  parameters in the retinal vessel detection problem. Cruz et al. [23] introduced a comparative analysis of four nature-inspired algorithms to obtain the optimal set of parameters of the GMF, to be applied in automatic detection of coronary arteries.

In general, the state-of-the-art GMF-based methods mentioned above assume that by using the average width of blood vessels to form a single-scale GMF, all the vessels in the input image can be detected. To overcome this disadvantage, in the present work, a novel multiscale Gaussian-matched filter (MGMF) based on artificial neural networks is introduced. The proposed method consists of detection and segmentation steps of coronary arteries in X-ray angiograms. In the detection step, MGMF is used to improve the contrast of blood vessels, and it is compared with six GMF-based methods in terms of the area ( $A_z$ ) under the ROC curve. In the segmentation step, a comparative analysis of ten thresholding methods of the state of the art is performed using the multiscale Gaussian filter response. Finally, the segmentation results of the proposed method are compared and discussed with those obtained using seven specialized vessel segmentation methods in terms of the accuracy measure.

The remainder of this paper is organized as follows. In Section 2, the fundamentals of the Gaussian-matched filters, artificial neural networks, and the proposed MGMF method are described in detail. The experimental results are presented and discussed in Section 3, and conclusions are given in Section 4.

## 2. Methods

Given the suitable performance of the Gaussian-matched filters for detecting coronary arteries in X-ray angiograms, a new multiscale Gaussian-matched filter based on a multi-layer neural network is proposed in the present work; this method is described in detail in the present section.

**2.1. Gaussian-Matched Filters.** The Gaussian-matched filters (GMF) were originally proposed by Chaudhuri et al. [12] for detecting vessel-like structures in medical imaging. The main idea behind the GMF is to approximate the shape of vessel-like structures in the spatial image domain by applying a Gaussian template. This template is formed by a Gaussian curve, which can be defined as follows:

$$G(x, y) = -\exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), |y| \leq \frac{L}{2}, \quad (1)$$

where the variable  $L$  is used to set the length in pixels of the vessel segment to be detected, and  $\sigma$  represents the average width of the vessel-like structures. To establish the width in pixels of the matching template, a discrete parameter  $T$  must be introduced to define the position where the Gaussian curve trails will cut.

Since the vessel-like structures can appear at different orientations, the Gaussian kernel  $G(x, y)$  can be also rotated by using a geometric transformation at different angles  $\theta$  as follows:

$$\kappa = \begin{bmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{bmatrix}, \quad (2)$$

where  $\kappa$  represents the number of evenly spaced directional filters  $\kappa = 180/\theta$  in the range  $[-\pi/2, (\pi/2)]$ . To obtain the Gaussian filter response, these oriented kernels are convolved with the input image, and the pixels with maximum response over all orientations are preserved.

On the other hand, a tuning step for the four GMF parameters plays an essential role for each application. In Figure 1, an X-ray angiogram along with the ground truth image outlined by a specialist is illustrated. Figure 1(c) presents the Gaussian filter response obtained using the parameter values proposed by Chaudhuri et al. [12] ( $\sigma = 2.0$ ,  $L = 9$ ,  $T = 13$ , and  $\kappa = 12$ ). Figures 1(d)–1(f) present the Gaussian matching templates with the aforementioned values and with  $\theta = 0^\circ$ ,  $\theta = 45^\circ$ , and  $\theta = 90^\circ$ , respectively.

**2.2. Artificial Neural Networks.** Artificial neural networks (ANN) are machine learning techniques inspired by neuron connections in the brain and they are commonly used for classification problems. ANN consists of multiple computing units that resemble to biological neurons connected in a network capable of approximating unknown functions [24]. This network consists of multiple computing units, also called artificial neurons, which perform the weighted sum of their corresponding inputs to be evaluated into an activation function [25]. For each artificial neuron, the evaluation of the activation function is passed as an input for following computing units in the network. This computing units can be

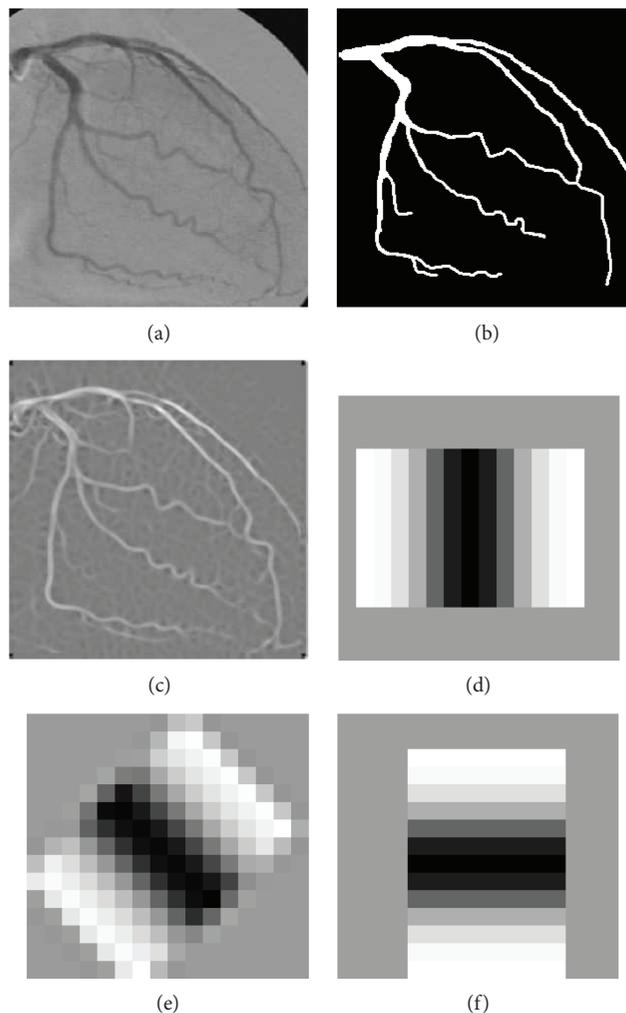


FIGURE 1: (a) Original X-ray coronary angiogram. (b) Ground truth of angiogram in (a). (c) Gaussian filter response applying 12 directional kernels on the angiogram in (a). (d), (e), (f) Gaussian templates with  $\theta = 0^\circ$ ,  $\theta = 45^\circ$ , and  $\theta = 90^\circ$ , respectively, as proposed by Chaudhuri et al. [12].

arranged in layers that receive the same inputs but use different weights. For a single layer, the weights can be arranged into a matrix  $W$  of size  $n \times m$ , where  $n$  is the number of neurons and  $m$  is the number of inputs in the current layer. The computation of one ANN layer can be seen as the evaluation of a matrix-vector product in the continuous activation function  $g(\cdot)$  as follows:

$$F(x) = g(W^T x), \quad (3)$$

where  $W$  is the matrix of weight values and  $x$  is the input vector.

In classification problems, a threshold value for the activation function can be set in order to differentiate between classes. However, the continuous evaluation of the activation function can be used as an universal function approximator.

The performance of the ANN depends on the architecture of the network defined by the number of layers. Each layer contains a number of neurons which are defined by an activation function and an associated weights vector. Commonly, the weights are fitted through a training

process while the architecture and the activation functions remain unchanged. For the training of the network weights, the back-propagation method with gradient descent has been widely employed [26]; however, other optimization schemes inside the back-propagation step such as Levenberg-Marquardt algorithm have proved to be more efficient [27].

Figure 2 illustrates the architecture of an ANN with three different layers. The first layer, also called input layer, consists of  $m$  input values  $x = [x_1, x_2, \dots, x_m]$ , where the input information directly depends on the problem to be solved. In the diagram, the second layer is conformed by  $n$  neurons:  $[1, 2, \dots, n]$  with activation function  $g(\cdot)$ , and it is called a hidden layer because is not relevant to know the result of the function  $g(\cdot)$  for each neuron. The third layer contains one neuron and has an activation function  $h(\cdot)$ , and it is called the output layer because it returns the evaluation value of the input  $x$  in the function  $F(\cdot)$ .

*2.3. Proposed Multiscale Gaussian-Matched Filters.* To overcome the main disadvantage of the single-scale GMF-based

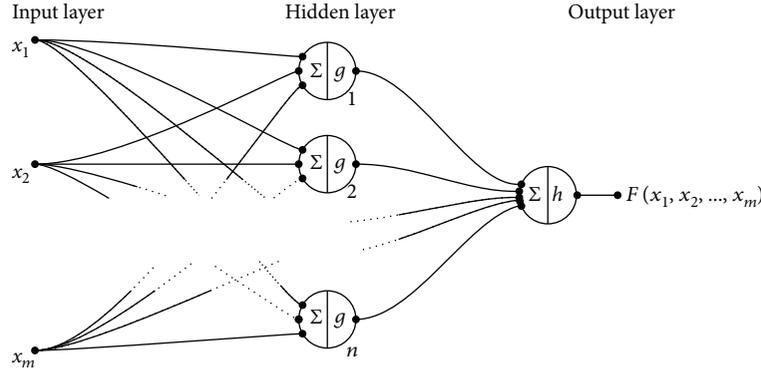


FIGURE 2: Representation of an artificial neural network with one layer of  $m$  input values, a hidden layer with activation function  $g(\cdot)$  and  $n$  neurons, and an output layer with one neuron and activation function  $h(\cdot)$ .

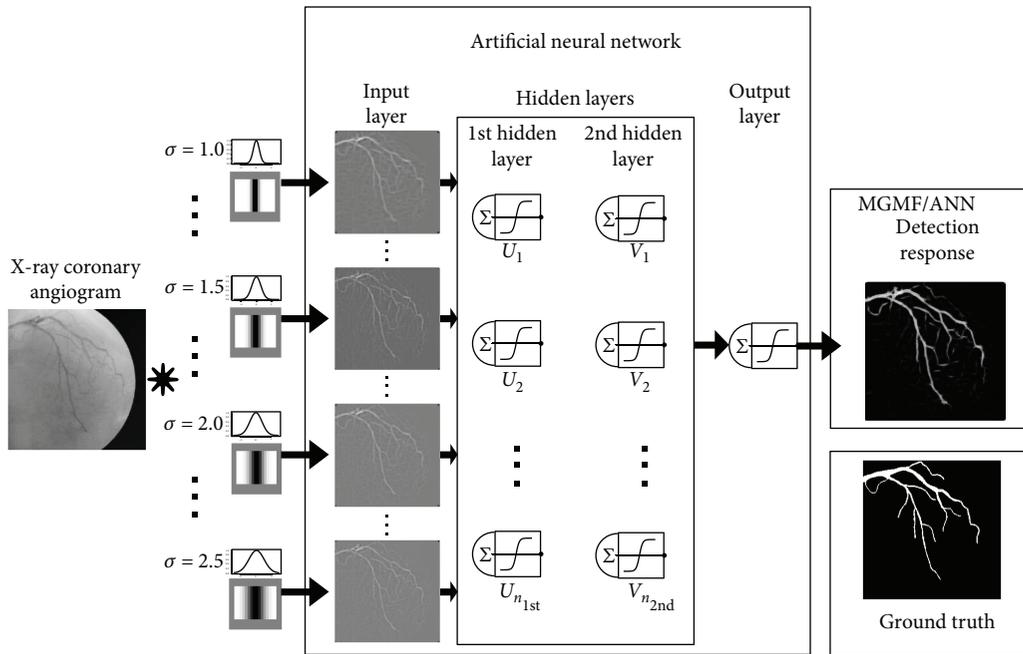


FIGURE 3: Flowchart of the proposed MGMF vessel detection method.

methods in detecting vessels of different calibers, the proposed method takes advantage of Gaussian curves at different scales in a predefined range  $\sigma = \{1, \dots, n\}$ .

The flowchart of the proposed method is illustrated in Figure 3. As it can be observed, the procedure of the proposed MGMF vessel detection method consists of four different steps. Firstly, a number of Gaussian scales  $\{\sigma_i | i = 1, 2, \dots, n\}$  must be defined to form a set of Gaussian matching templates. The set of Gaussian scales was determined by using a global search on the training set of images in terms of the area under the ROC curve. In the second step, the Gaussian templates at different values of the  $\sigma$  parameter are generated in order to be convolved with the X-ray angiogram input images. Each template is formed according to the  $T$  and  $L$  parameters and rotated using the number of directional filters  $\kappa$  with angular resolution  $\theta$ . Since the ANNs represent a supervised machine learning technique, in the

third step, the resulting Gaussian filter responses and the ground truth images can be arranged as an input data matrix of  $n$  columns and label vector, respectively. Finally, in the last step of the proposed method, the ANN is trained by a predefined number of hidden layers (the ANN architecture is discussed in Section 3.1). From the ANN, the resulting image represents the vessel detection response, which can be evaluated using a ground truth image and a metric for binary classification.

In the present work, the area ( $A_2$ ) under the ROC curve is used to select the most suitable set of parameters for the ANN as well as to assess their performance in vessel detection using the training set of angiograms. This measure is explained in the following Section 2.4.

**2.4. Evaluation Metrics.** To assess the performance of the vessel enhancement and vessel segmentation methods,

the area ( $A_z$ ) under the receiver operating characteristic (ROC) curve for gray-scale images and the accuracy metric for binary images have been adopted in this work.

Both evaluation measures are in the range  $[0, 1]$ , where the value 1 is acquired when the vessel pixels and background image obtained from the computational experiments are completely superimposed with the ground truth provided by the specialist; otherwise, the obtained value corresponds to 0.

The ROC curve is a measure that evaluates the performance of a classification method. This measure is a plot of the true-positive fraction (TPF) also called sensitivity and the false-positive fraction (FPF). TPF represents the rate of correctly classified pixels (vessel pixels) and FPF represents the rate of nonvessel pixels incorrectly classified by the computational method. To compute the ROC curve, a sliding threshold over the gray-scale filter response is computed, and the area  $A_z$  under the curve is calculated through the Riemann sum method.

The accuracy measure [28] has been widely used to evaluate the performance of binary classifiers; consequently, it has been adopted to assess the performance of the vessel segmentation results (binary images). This measure is defined as the rate of correctly classified pixels regarding the number of pixels in the input image as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (4)$$

where TP and TN represent the subsets of correctly classified vessel and nonvessel pixels, respectively, and FN and FP represent the subsets of incorrectly classified pixels.

In Section 3, the vessel segmentation results obtained from the proposed MGMF method using a database of 100 X-ray angiographic images are analyzed by the evaluation metrics.

### 3. Results and Discussion

In this section, the vessel enhancement and segmentation results obtained from the proposed MGMF method are presented and analyzed. The computational experiments of the MGMF method based on an artificial neural network were implemented on a computer with an Intel Core i3, 4 GB of RAM, and a 2.13 GHz processor using the Matlab software version 2016a.

The database of gray-scale images used in the present work consists of 100 X-ray coronary angiograms of size  $300 \times 300$  pixels from different patients. Each angiogram was outlined by a specialist to form the ground truth images for evaluation purposes. An ethics approval letter was provided by the Mexican Social Security Institute, UMAE Leon. In the experiments, the whole set of X-ray angiograms was divided into the training and testing sets with 50 images in each one in order to assess the vessel enhancement and segmentation methods.

**3.1. Optimization of the ANN Architecture.** A four layered ANN is proposed to detect coronary arteries using the multiscale Gaussian filter response of the X-ray coronary

angiograms. The first layer is the input layer, which receives the responses of the GMF at 10 different  $\sigma$  values. The second and third layers are hidden layers, which respective number of neurons is defined through an optimization process, described later on this section. The fourth layer is the output layer, responsible for generating the MGMF response.

Considering that the performance of the proposed ANN rely on the two hidden layers, an exhaustive search was used to define the number of neurons inside each of them. The search was performed within the space:  $n_{1st}, n_{2nd} \in [1, 2, \dots, 10]$ , where  $n_{1st}$  and  $n_{2nd}$  are the number of neurons inside the first and second hidden layers, respectively. The objective of the exhaustive search was to design the optimal ANN architecture, which must maximize the area under the ROC curve ( $A_z$ ) using the multiple responses of the MGMF in the training set of angiograms. The parameters of the MGMF were assigned as  $l = 13$  and  $T = 15$ , using  $K = 12$  kernel orientations. Those parameters were defined according to the design of the GMF obtained by Cruz et al. [23], using a nature-inspired algorithm to optimize the  $l$ ,  $T$ , and  $\sigma$  parameters, which reported a high detection performance. The average width of the vessels  $\sigma$  on the proposed MGMF is defined in the range:  $[1.5, 1.6, \dots, 2.5]$ , giving a multiscale approach to the detection method. The search space of the  $A_z$  values with respect to the number of neurons inside the two hidden layers is illustrated in Figure 4.

From the exhaustive search results, the optimal ANN architecture was designed to use 3 neurons in the first hidden layer and 8 in the second hidden layer. This optimal architecture will be referred as ANN(3-8) in the remaining of this article. Table 1 presents a statistical analysis of the proposed MGMF/ANN(3-8) behaviour in the training set of coronary angiograms. The statistical results show a high robustness of the MGMF/ANN(3-8) method according to the low standard deviation of its performance after 30 runs.

**3.2. Results of Vessel Enhancement.** Table 2 presents a comparative analysis between the resulting performance of the proposed MGMF/ANN(3-8) method and six GMF-based methods from the state of the art. In this analysis, the whole set of coronary angiograms was used. The analysis was performed using 7 methods. First, the proposal by Chaudhuri et al. [12] was that the authors defined the GMF parameters experimentally as  $l = 9$ ,  $\sigma = 2.0$ , and  $T = 13$ , with  $\kappa = 12$  orientations. The approach of Cinsdikici and Aydin [16] uses the same set of parameters as proposed by Chaudhuri et al. [12], but changes the number of kernel orientations to  $\kappa = 18$  in order to increase the range of directions of the GMF response. The method of Kang et al. [18, 20, 21] defines the GMF parameters as  $l = 9$ ,  $T = 13$ , and  $\kappa = 6$  and modifies  $\sigma$  to 1.5 according to the experiments performed by the authors. The approach of Al-Rawi et al. [15, 22] proposes the GMF parameters optimization through a full search and later through a genetic algorithm. The search space defined by Al-Rawi et al. [15, 22] was established as  $l = [7, 7.1, \dots, 11]$ ,  $T = [2, 2.25, \dots, 10]$ ,  $\sigma = [1.5, 1.6, \dots, 3]$ , and keeping  $\kappa = 12$  orientations. The method of GMF-Evol [23] compared four algorithms from the evolutionary computation family and defined the parameters for the

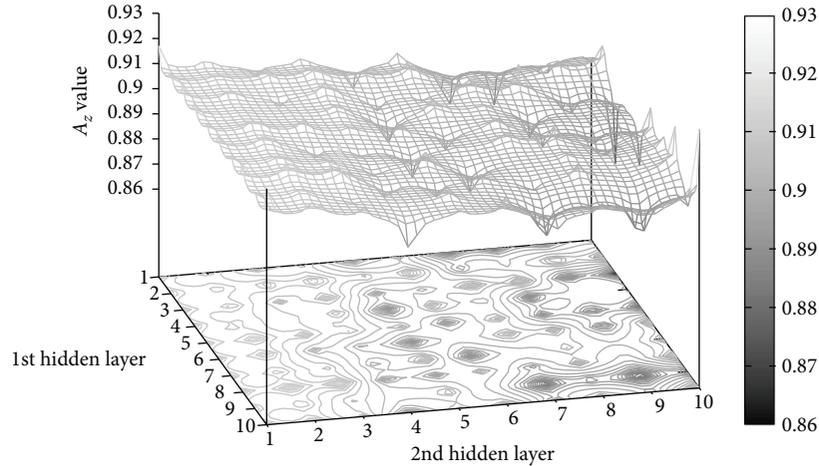


FIGURE 4: Search space to obtain the optimal architecture of the MGMF, including the number of neurons for each hidden layer of the ANN and using the area  $A_z$  under the ROC curve in the training set of angiograms.

TABLE 1: Statistical analysis using 30 runs of the MGMF over the training set of X-ray angiograms.

Method	Maximum	Minimum	Mean	Std. dev.	Median
MGMF/ANN(3-8)	0.9357	0.9155	0.9203	0.002	0.9169

TABLE 2: Comparative analysis of vessel detection performance using the training and testing sets in terms of the area under the ROC curve.

Vessel detection method	Area under ROC curve ( $A_z$ )	
	Training set	Testing set
GMF-Entropy [19]	0.8849	0.8812
Kang et al. [18, 20, 21]	0.8901	0.8852
Chaudhuri et al. [12]	0.9012	0.8963
Cinsdikici and Aydin [16]	0.9087	0.9002
Al-Rawi et al. [15, 22]	0.9104	0.9123
GMF-Evol [23]	0.9142	0.9171
Proposed method (MGMF)	0.9357	0.9362

GMF as the optimal set obtained by the differential evolution algorithm within the search space of  $l = [8, 9, \dots, 15]$ ,  $T = [8, 9, \dots, 15]$ ,  $\sigma = [1, 1.01, \dots, 5]$ , and keeping  $\kappa = 12$  kernel orientations. Finally, the method GMF-Entropy [19] replaced the area ( $A_z$ ) under the ROC curve with a nature-inspired optimization algorithm by using an entropy-based objective function for the GMF parameter optimization.

According to the comparative analysis of the computational experiments, the method GMF-Entropy [19] presented the lowest detection performance in both image sets. Also, it can be noticed that the methods of Kang et al. [18, 20, 21], Chaudhuri et al. [12], and Cinsdikici and Aydin [16] that defined the GMF parameters experimentally presented a similar behaviour. The methods of Al-Rawi et al. [15, 22] and GMF-Evol [23] that searched the optimal parameters for the GMF show a slight increment on the  $A_z$

response. Moreover, the proposed method achieved the best detection with an  $A_z$  rate of 0.9357 in the training set and 0.9362 in the testing set of coronary angiograms.

The vessel detection response of the six GMF-based methods and the proposed MGMF/ANN(3-8) are presented in Figure 5 for a subset of five coronary angiograms from the testing set. The MGMF/ANN(3-8) responded with a greater visible contrast between the detected coronary artery and the background than the comparative GMF methods.

**3.3. Results of Vessel Segmentation.** The extraction of the coronary artery is completed through the classification of the MGMF response into vessel and nonvessel pixels. The classification of the detection response has been commonly carried out by a thresholding algorithm, which defines a limit value with the purpose of separating the vessel pixels from the background image. In order to take advantage of the high detection performance of the MGMF response, ten different thresholding methods from the state of the art were tested with the interest of defining the ideal coronary artery segmentation method.

Methods based on the entropy, such as the method of Kapur et al. [29] and the method of Pal and Pal [30], optimize the location of the value  $t$  which separates the histogram of the gray-scale image into two classes, with the objective of maximize the entropy of the two resulting classes. The moment-preserving method introduced by Tsai [31] assumes that the first three moments of the resulting binary image must be preserved. The threshold value  $t$  is defined as the location in the histogram of the gray-scale image which solves four predefined equations. The method of Rosenfeld and De la Torre [32] referred as the histogram concavity algorithm, which works with the convex hull of the

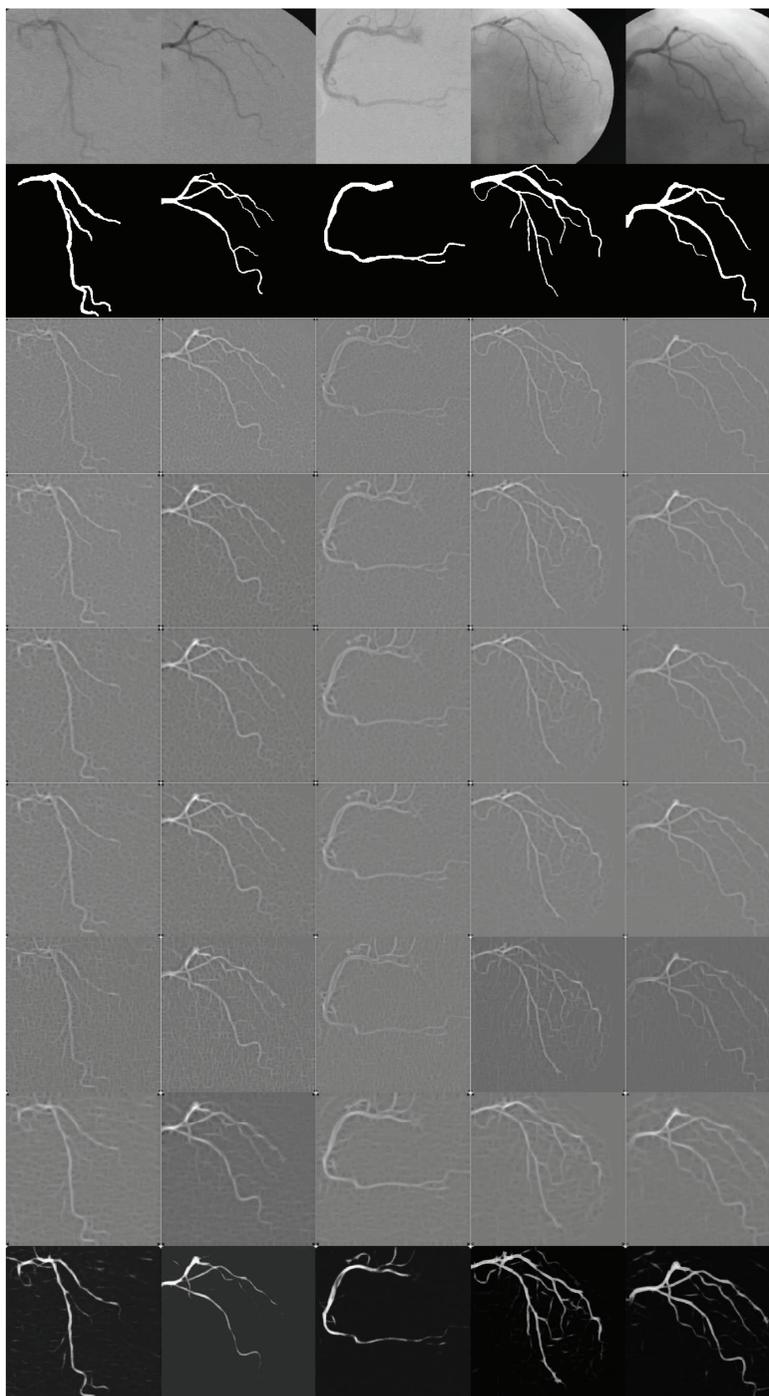


FIGURE 5: First row: subset of angiograms from the test set. Second row: ground truth of images. The remaining seven rows present the GMF response of the methods of GMF-Entropy [19], Kang et al. [18, 20, 21], Chaudhuri et al. [12], Cinsdikici and Aydin [16], Al-Rawi et al. [15, 22], GMF-Evol [23], and the proposed method (MGMF), respectively.

histogram of the gray-scale image. The threshold value  $t$  is defined as the location of the local maxima of the difference between the convex hull and the image histogram. The Rutherford-Appleton threshold selection (RATS) method [33] defines a maximum gradient image from the derivatives in the  $x$  and  $y$  directions of the input image. The threshold value  $t$  is defined as the division of the sum of every element of the dot product between the maximum

gradient image and the input image, divided by the sum of every element of the maximum gradient image. In a probability distribution approach, the method of Otsu [34] proposes the selection of the threshold value  $t$  as the location in the histogram of the gray-scale image which maximizes the between-class variance of the resulting two classes. In a similar approach, Ridler and Calvard [35] model the two classes of the binary image using two

TABLE 3: Comparative analysis of ten thresholding methods of the state of the art using the multiscale Gaussian response over the test set of X-ray angiograms.

Thresholding method	Accuracy
Otsu [34]	0.9568
Moments [31]	0.9561
Ridler and Calvard [35]	0.9560
RATS [33]	0.9533
Kapur et al. [29]	0.9514
Sauvola and Pietikäinen [38]	0.9290
Histogram concavity [32]	0.9254
Pal and Pal [30]	0.8982
Niblack [37]	0.8644
White and Rohrer [36]	0.8304

Gaussian distributions. The resulting classes are separated by a threshold value  $t$ , defined as the average of the location parameters of the two Gaussian distributions, which are estimated through an iterative search. The method of White and Rohrer [36] classify each pixel using a local threshold value, which is defined as the expected intensity value of the corresponding neighborhood. Niblack [37] proposed the use of a similar neighborhood-defined threshold value, which also considers the local standard deviation. Finally, the method of Sauvola and Pietikäinen [38] defines a threshold value for each pixel according to the magnitude of the local standard deviation. This approach increases the threshold value for neighborhoods with low spread of the intensity and decreases the threshold value in neighborhoods with high standard deviation.

Table 3 presents the accuracy of the ten thresholding methods; the MGMF response of the testing set of angiograms was used as input. The comparative analysis shows that the thresholding method proposed by Otsu [34] produces the most accurate segmentation among the ten comparative methods; therefore, the method of Otsu [34] is used to binarize the MGMF response in further analysis.

The proposed MGMF/Otsu method for coronary artery segmentation was compared with seven vessel segmentation methods from the state of the art using the testing set of coronary angiograms. Table 4 presents the comparative analysis of the vessel segmentation accuracy of the proposed MGMF/Otsu method and seven state-of-the-art methods. The experimental results of the comparative analysis show that the MGMF/Otsu outperforms the state-of-the-art methodologies in terms of the vessel extraction accuracy.

Finally, the execution time of the proposed MGMF/Otsu method and the seven state-of-the-art methods is shown in Table 5. The methods of Eiho and Qian [1] and Chanwimaluang et al. [13, 14] present the two lower execution times; however, those methods also produce low accurate segmentation of the coronary arteries from the testing set. The methods of Li et al. [9] and Wang et al. [8] were executed, respectively, in the third and fourth lower times, although those methods performed

TABLE 4: Comparative analysis of the proposed method (MGMF) with respect to seven state-of-the-art vessel segmentation methods using the test set of 50 X-ray images.

Segmentation method	Accuracy
MGMF/Otsu	0.9568
Kang et al. [21]	0.9417
Tsai et al. [10]	0.9402
Li et al. [9]	0.9394
Wang et al. [8]	0.9386
Eiho and Qian [1]	0.9271
Chanwimaluang et al. [13, 14]	0.9150
Kang et al. [18]	0.8843

TABLE 5: Average execution time for the proposed method as compared with to the state-of-the-art segmentation methods using the test set of angiograms.

Segmentation method	Execution time (seconds)
MGMF/Otsu	1.73
Kang et al. [21]	2.51
Tsai et al. [10]	1.91
Li et al. [9]	1.52
Wang et al. [8]	1.63
Eiho and Qian [1]	1.05
Chanwimaluang et al. [13, 14]	1.24
Kang et al. [18]	2.01

similarly and obtained an accuracy below to 94%. The proposed MGMF/Otsu method provides the best trade-off between accuracy and computational time, according to the experimental results, with an execution time of 1.73 seconds.

Figure 6 shows the segmentation responses of the proposed MGMF/Otsu method and the seven state-of-the-art methods. The methods of Tsai et al. [10], Li et al. [9], and Chanwimaluang et al. [13, 14] present a high number of false positives, represented by the white pixels classified as vessels that are absent in the ground truth. Conversely, the methods of Kang et al. [21], Wang et al. [8], and Kang et al. [18] fail to extract thin vessels that are present in the ground truth, producing a high rate of false negative pixels. By visual examination of the Eiho and Qian [1] method responses, it noticed the presence of jagged edges, which reduce the segmentation accuracy. On the other hand, the segmentation responses of the proposed MGMF/Otsu method present smooth edges with an acceptable compromise of true-positive and false-negative pixels.

The comparative results of the performed experiments suggest that the proposed MGMF/Otsu method is robust for the coronary arteries detection and is capable of providing accurate segmentations from X-ray coronary angiograms, within a competitive computational time. The efficiency of the proposed MGMF/Otsu method encourages its usage to aid the decisions making in the medical practice.

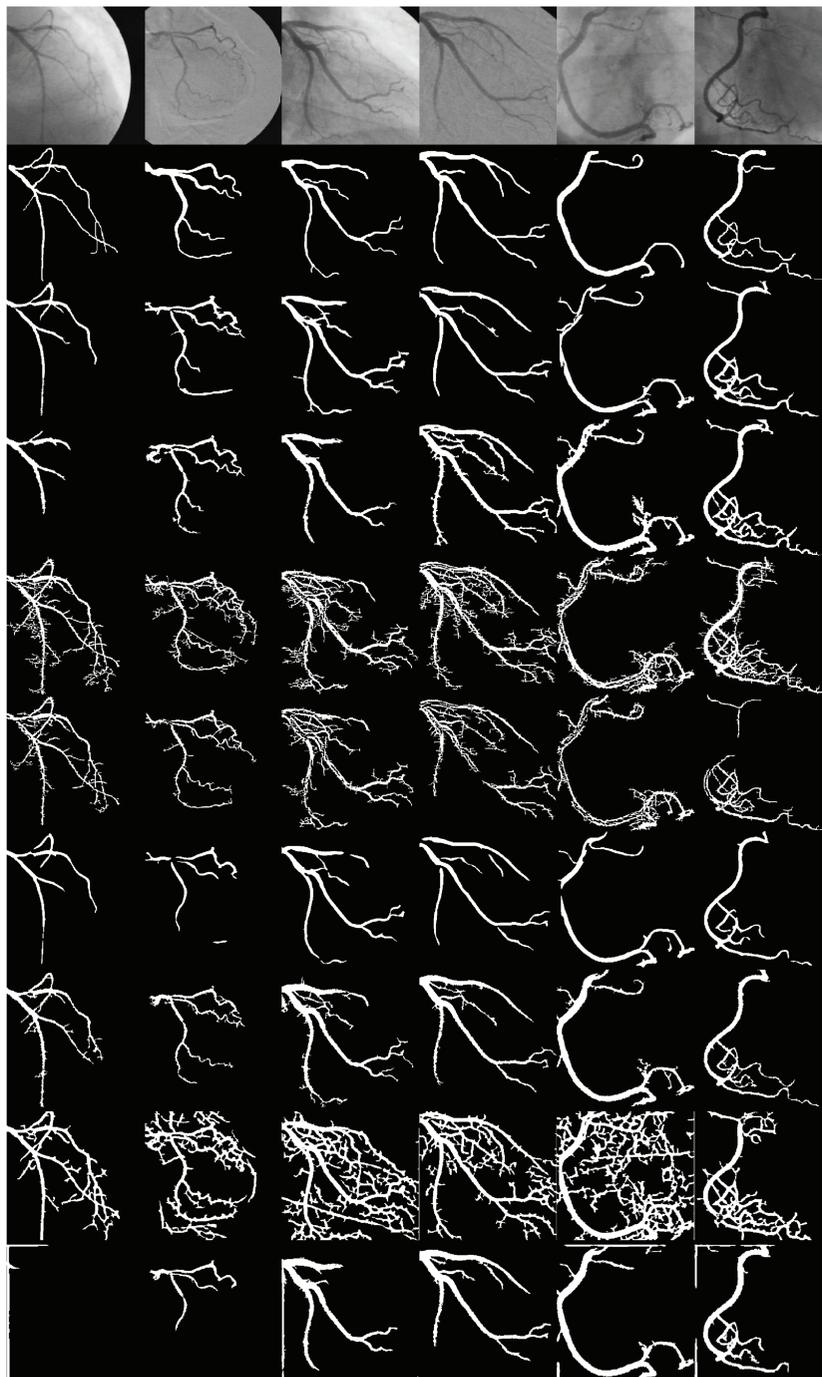


FIGURE 6: First row: subset of angiograms from the test set. Second row: ground truth of images. The remaining eight rows present the segmentation results of the proposed MGF method, Kang et al. [21], Tsai et al. [10], Li et al. [9], Wang et al. [8], Eiho and Qian [1], Chanwimaluang et al. [13, 14], and Kang et al. [18], respectively.

#### 4. Conclusion

In this paper, a novel method based on multiscale Gaussian-matched filters and artificial neural networks approach has been introduced. The statistical results show the robustness of the optimal architecture found for the ANN, which uses 3 neurons in the first hidden layer, and 8 neurons in the second. The optimal design reached a detection rate of

0.9357 in terms of the area ( $A_z$ ) under the ROC curve, using the training set of 50 angiograms. Further analysis showed that the proposed MGF/ANN(3-8) method performs better than six GMF-based methods from the state of the art, presenting an  $A_z = 0.9362$  using the testing set of 50 images. The segmentation of the MGF/ANN(3-8) response carried out by the thresholding method of Otsu [34] has proven to be the most efficient from ten of the

state-of-the-art segmentation techniques. According to the experimental results, the application of an artificial neural network of optimal architecture, over the responses of the multiscale Gaussian-matched filters, and the subsequent threshold of the response by the Otsu's method provide the most accurate segmentation of the coronary artery, with a correspondence of 0.9568, in a competitive execution time of 1.73 seconds.

### Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this article.

### Acknowledgments

This research has been supported by the National Council of Science and Technology of México under the Project Cátedras-CONACYT 3150-3097.

### References

- [1] S. Eiho and Y. Qian, "Detection of coronary artery tree using morphological operator," in *Computers in Cardiology 1997*, vol. 24, pp. 525–528, Lund, Sweden, September 1997.
- [2] Y. Qian, S. Eiho, N. Sugimoto, and M. Fujita, "Automatic extraction of coronary artery tree on coronary angiograms by morphological operators," in *Computers in Cardiology 1998*, vol. 25, pp. 765–768, Cleveland, OH, USA, September 1998.
- [3] K. Q. Sun and N. Sang, "Morphological enhancement of vascular angiogram with multiscale detected by Gabor filters," *Electronic Letters*, vol. 44, no. 2, p. 86, 2008.
- [4] B. Bouraoui, C. Ronse, J. Baruthio, N. Passat, and P. Germain, "Fully automatic 3D segmentation of coronary arteries based on mathematical morphology," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1059–1062, Paris, France, May 2008.
- [5] D. S. D. Lara, A. W. C. Faria, A. Araujo, and D. Menotti, "A semi-automatic method for segmentation of the coronary artery tree from angiography," in *2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*, pp. 194–201, Rio de Janeiro, Brazil, October 2009.
- [6] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, vol. 1496, pp. 130–137, Springer LNCS, 1998.
- [7] N. M. Salem and A. K. Nandi, "Unsupervised segmentation of retinal blood vessels using a single parameter vesselness measure," in *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, vol. 34, pp. 528–534, Bhubaneswar, India, December 2008.
- [8] S. Wang, B. Li, and S. Zhou, "A segmentation method of coronary angiograms based on multi-scale filtering and region-growing," *2012 International Conference on Biomedical Engineering and Biotechnology*, 2012, pp. 678–681, Macao, China, May 2012.
- [9] Y. Li, S. Zhou, J. Wu, X. Ma, and K. Peng, "A novel method of vessel segmentation for X-ray coronary angiography images," in *2012 Fourth International Conference on Computational and Information Sciences*, pp. 468–471, Chongqing, China, August 2012.
- [10] T.-C. Tsai, H.-J. Lee, and M. Y.-C. Chen, "Adaptive segmentation of vessels from coronary angiograms using multi-scale filtering," in *2013 International Conference on Signal-Image Technology & Internet-Based Systems*, pp. 143–147, Kyoto, Japan, December 2013.
- [11] F. M'hiri, L. Duong, C. Desrosiers, and M. Cheriet, "Vessel-walker: coronary arteries segmentation using random walks and hessian-based vesselness filter," in *2013 IEEE 10th International Symposium on Biomedical Imaging*, pp. 918–921, San Francisco, CA, USA, April 2013.
- [12] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on Medical Imaging*, vol. 8, no. 3, pp. 263–269, 1989.
- [13] T. Chanwimaluang and G. Fan, "An efficient blood vessel detection algorithm for retinal images using local entropy thresholding," in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, vol. 5, pp. V-21–V-24, Bangkok, Thailand, May 2003.
- [14] T. Chanwimaluang, G. Fan, and S. R. Fransen, "Hybrid retinal image registration," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 129–142, 2006.
- [15] M. Al-Rawi, M. Qutaishat, and M. Arrar, "An improved matched filter for blood vessel detection of digital retinal images," *Computers in Biology and Medicine*, vol. 37, no. 2, pp. 262–267, 2007.
- [16] M. G. Cinsdikici and D. Aydin, "Detection of blood vessels in ophthalmoscope images using MF/ant (matched filter/ant colony) algorithm," *Computer Methods and Programs in Biomedicine*, vol. 96, no. 2, pp. 85–95, 2009.
- [17] Z. Yavuz and C. Köse, "Blood vessel extraction in color retinal fundus images with enhancement filtering and unsupervised classification," *Journal of Healthcare Engineering*, vol. 2017, Article ID 4897258, 12 pages, 2017.
- [18] W. Kang, K. Wang, W. Chen, and W. Kang, "Segmentation method based on fusion algorithm for coronary angiograms," in *2009 2nd International Congress on Image and Signal Processing*, pp. 1–4, Tianjin, China, October 2009.
- [19] I. Cruz-Aceves, F. Cervantes-Sanchez, A. Hernandez-Aguirre, R. Perez-Rodriguez, and A. Ochoa-Zezzatti, "A novel Gaussian matched filter based on entropy minimization for automatic segmentation of coronary angiograms," *Computers & Electrical Engineering*, vol. 53, pp. 263–275, 2016.
- [20] W. Kang, W. Kang, W. Chen, B. Liu, and W. Wu, "Segmentation method of degree-based transition region extraction for coronary angiograms," in *2010 2nd International Conference on Advanced Computer Control*, vol. 4, pp. 466–470, Shenyang, China, March 2010.
- [21] W. Kang, W. Kang, Y. Li, and Q. Wang, "The segmentation method of degree-based fusion algorithm for coronary angiograms," in *Proceedings of 2013 2nd International Conference on Measurement, Information and Control*, vol. 01, pp. 696–699, Harbin, China, August 2013.
- [22] M. Al-Rawi and H. Karajeh, "Genetic algorithm matched filter optimization for automated detection of blood vessels from digital retinal images," *Computer Methods and Programs in Biomedicine*, vol. 87, no. 3, pp. 248–253, 2007.
- [23] I. Cruz-Aceves, A. Hernandez-Aguirre, and S. I. Valdez, "On the performance of nature inspired algorithms for the automatic segmentation of coronary arteries using Gaussian matched filters," *Applied Soft Computing*, vol. 46, pp. 665–676, 2016.

- [24] J. Feldman and R. Rojas, *Neural Networks: a Systematic Introduction*, Springer, Berlin Heidelberg, 2013.
- [25] W. Pitts and W. S. McCulloch, "How we know universals the perception of auditory and visual forms," *The Bulletin of Mathematical Biophysics*, vol. 9, no. 3, pp. 127–147, 1947.
- [26] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *1960 IRE WESCON Convention Record*, pp. 96–104, Neurocomputing MIT Press, 1988, Cambridge, MA, USA, 1960.
- [27] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.
- [28] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, pp. 271–274, 1998.
- [29] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 3, pp. 273–285, 1985.
- [30] N. R. Pal and S. K. Pal, "Entropic thresholding," *Signal Processing*, vol. 16, no. 2, pp. 97–108, 1989.
- [31] W.-H. Tsai, "Moment-preserving thresholding: a new approach," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 3, pp. 377–393, 1985.
- [32] A. Rosenfeld and P. De la Torre, "Histogram concavity analysis as an aid in threshold selection," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 2, pp. 231–235, 1983.
- [33] J. Kittler, J. Illingworth, and J. Föglein, "Threshold selection based on a simple image statistic," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 2, pp. 125–147, 1985.
- [34] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [35] T. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 8, pp. 630–632, 1978.
- [36] J. M. White and G. D. Rohrer, "Image thresholding for optical character recognition and other applications requiring character image extraction," *IBM Journal of Research and Development*, vol. 27, no. 4, pp. 400–411, 1983.
- [37] W. Niblack, *An Introduction to Digital Image Processing*, Prentice-Hall, New Jersey, USA, 1986.
- [38] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225–236, 2000.

## Research Article

# Effective Pneumothorax Detection for Chest X-Ray Images Using Local Binary Pattern and Support Vector Machine

Yuan-Hao Chan,<sup>1</sup> Yong-Zhi Zeng,<sup>2</sup> Hsien-Chu Wu ,<sup>2,3</sup> Ming-Chi Wu,<sup>4</sup>  
and Hung-Min Sun <sup>1,5</sup>

<sup>1</sup>Institute of Information Systems and Applications, National Tsing Hua University, Hsinchu, Taiwan

<sup>2</sup>Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taichung, Taiwan

<sup>3</sup>Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung, Taiwan

<sup>4</sup>School of Medicine, Chung Shan Medical University, Taichung, Taiwan

<sup>5</sup>Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

Correspondence should be addressed to Hsien-Chu Wu; wuhc@nutc.edu.tw

Received 29 September 2017; Revised 21 January 2018; Accepted 19 February 2018; Published 3 April 2018

Academic Editor: Weide Chang

Copyright © 2018 Yuan-Hao Chan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Automatic image segmentation and feature analysis can assist doctors in the treatment and diagnosis of diseases more accurately. Automatic medical image segmentation is difficult due to the varying image quality among equipment. In this paper, the automatic method employed image multiscale intensity texture analysis and segmentation to solve this problem. In this paper, firstly, SVM is applied to identify common pneumothorax. Features are extracted from lung images with the LBP (local binary pattern). Then, classification of pneumothorax is determined by SVM. Secondly, the proposed automatic pneumothorax detection method is based on multiscale intensity texture segmentation by removing the background and noises in chest images for segmenting abnormal lung regions. The segmentation of abnormal regions is used for texture transformed from computing multiple overlapping blocks. The rib boundaries are identified with Sobel edge detection. Finally, in obtaining a complete disease region, the rib boundary is filled up and located between the abnormal regions.

## 1. Introduction

Medical imaging refers to the technical and processing of noninvasive acquisition of internal tissue images of a part of the human or human body for medical or medical research. It is a kind of inverse reasoning calculation. As a science, medical images belong to biological images and include imaging diagnostics, radiology, endoscopy, medical thermography, medical photography, and microscopes. In addition, techniques such as brainwave mapping and magnetoencephalography, although focusing on measurement and recording, do not show significant images, but because of the locating characteristics (i.e., containing positional information) of the resulting data, it can be considered as another form of medical image.

In clinical application, also known as medical imaging, or imaging medicine, some hospitals will have to have a medical imaging center, medical imaging department, or medical imaging department; set up related equipment; and prepare for dedicated nurses, radiologists, and physicians, responsible for the operation of instrumentation, interpretation, and diagnosis of images, which is different from radiation therapy for radiology.

In medical science, medical engineering, medical physics, and biomedical information science, medical imaging usually refers to the science of researching and developing images and capturing and storing technologies and equipment. Research on how to interpret and diagnose medical images is a supplementary science belonging to the radiology department or other medical fields (e.g., neurology, cardiovascular diseases).

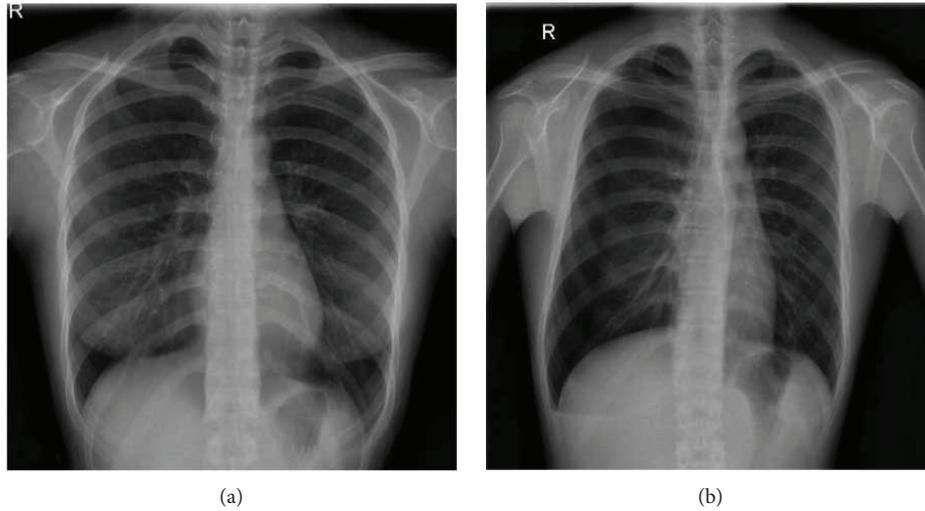


FIGURE 1: Chest radiographs of (a) normal and (b) pneumothorax.

Digital image processing has been widely applied in the medical domain. However, most of the methods still require manual processing. Automatic image segmentation and features analysis can assist doctors in the treatment and diagnosis of diseases with higher accuracy, accelerate diagnosis process, and improve efficiency. Automatic medical image segmentation is difficult due to the varied image quality caused by equipment and dosage. In this paper, the automatic method employed image multiscale intensity texture analysis and segmentation to solve this problem. The proposed methods automatically recognize and classify abnormal region without manual segmentation. Generally, automatic identification is based on the difference of textures and organ shapes, or any pathological changes of the lung area. Therefore, the important features could be retained to identify abnormal areas.

In this paper, the chest X-ray images are utilized for identifying lung-related health issues. The first proposed identifying common pneumothorax classification method is based on SVM. Features are extracted from the lung image by the local binary pattern. Then, classification of pneumothorax is determined by support vector machines. The second proposed automatic pneumothorax detection is based on multiscale intensity texture segmentation. The background and noises in the chest images are removed for segmenting the abnormal lung region. Since the area of rib boundaries can be affected easily, the rib boundaries are identified by Sobel edge detection. Finally, in order to cover the complete disease region, abnormal regions are located.

Medical imaging serves as an important source of information for anatomy and organ function along with the diagnosis of diseases. The integration of image processing techniques to medical imaging with machine learning methods has led to the development of computer-aided diagnostic and decision-making tools. The X-rays have the advantages of rapid reproducibility and low cost.

Among them, pneumothorax is a common and easy relapse disease in the diagnosis of chest X-rays. The pneumothorax is an abnormal accumulating of air in the pleural space

and leads lung discrete from the chest wall [1, 2]. In general, symptoms range from asymptomatic to life-threatening respiratory distress. If the pneumothorax is significant, it can cause a shift of the mediastinum and compromise hemodynamic stability. There are several types of pneumothorax: primary spontaneous pneumothorax, secondary spontaneous pneumothorax, and iatrogenic/traumatic pneumothorax. However, conditions such as smoking, a common cold, and a family history of pneumothorax may cause spontaneous pneumothorax. On the other hand, the blunt trauma, penetration trauma, and surgery would result in traumatic pneumothorax. According to the United States research, pneumothorax is prevalently occurring in humans, and its relapse ratio is 35% in men [3]. Furthermore, critical pneumothorax may lead to hypoxia, shock, and even death [4].

The chest radiograph has been used as a general adjunct for pneumothorax screening and diagnosis. According to ionizing radiation form of X-rays, the chest images have been generated with the absorption of different spectrums based on the tissue density [5]. Chest radiograph scans the whole chest and produces abundant information and radiologists have to review the data [6]. However, the pneumothorax may be missed or misclassified as other diseases easily in using chest radiographs, because pneumothorax has the characteristics of curved contour and smooth regions within dark region against the chest wall and ribs, and clavicles may overlap [7]. Previous work [7] that proposed the method is based on the image processing gradient-histogram analysis, and its use of many methods is quite complicated, and even many steps that require manual adjustment of parameters in the patent did not put forward the prediction of pneumothorax picture accuracy. This study mainly uses SVM, LBP, Sobel edge detection, and other methods to automate small sample prediction and achieve 85.8% accuracy on patch size  $5 \times 5$  blocks. If clinicians use the method proposed in this study, there is considerable confidence in automation and efficiency. Examples of abnormality for normal chest and pneumothorax are shown in Figure 1.

In the analysis of radiographs, the doctor diagnosed pneumothorax mainly through the main observation of the affected area will have a visceral pleural line, the outer periphery of the pleura line because of the gas, so in the image cannot be seen by the pulmonary blood vessels constitute lung markings and mediastinal displacement, defined in the British Thoracic Society guideline as chest wall larger than 2 cm for large pneumothorax and less than 2 cm for small pneumothorax.

The advantage of X-ray is the fast and inexpensive imaging, and experienced physicians can quickly identify the pneumothorax from the images. In this study, the methods used by these physicians were considered expert knowledge to serve as the machine learning direction for image processing.

The computer-aided diagnosis (CAD) has the potential for improving the diagnostic accuracy of radiologists to distinguish between pneumothorax and other conditions [8]. Several automated analysis of chest radiograph image algorithms had been developed [9]. Geva et al. [10] proposed a visual bag-of-words representation for classification of chest radiograph with different pathologies. Based on the local and global texture signature scheme, the chest radiograph was used to generate localized texture analysis for the detection of local abnormalities. The method applied training and detection of the abnormality by using Gentle AdaBoost to classify normal and abnormal images suggested by Geva et al. [10].

Previous researchers were mostly developed for the diagnosis of the chest radiograph, not for image segmentation and detection for abnormal location. These methods, which concentrate on diagnosis, are viable only on significantly different diseases. Moreover, the location of abnormal region detection is necessary for evaluating treatment guidelines of symptoms and acute disease indicators. For the purpose of detecting the location of abnormal regions, this paper proposed an image segmentation based on texture and intensity distribution for the chest radiograph. The method uses the texture analysis based on intensity and gradient for region segmentation to find out the pneumothorax regions.

At present, there are skin cancers and tuberculosis on the automatic recognition of machine learning or depth learning algorithms. This research focuses on pneumothorax machine learning classification. At present, there is no literature that has focused on pneumothorax medical image classification. The reasons for choosing Sobel edge detection is that it calculates very fast and efficiently on the edge of the medical images. This study uses SVM for its high classification efficiency and deals with high-dimensional spatial data effectively.

## 2. Identifying Pneumothorax Using Machine Learning Algorithms

The SVM-based identifying common pneumothorax method using local binary pattern is explained step by step as follows.

*2.1. The Proposed Scheme.* This paper presents identifying common pneumothorax system consisting of lung object

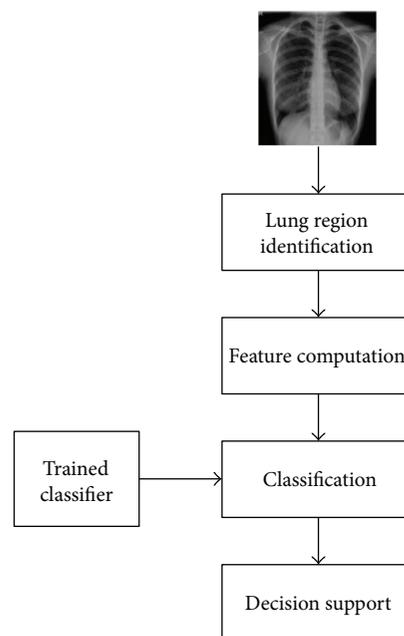


FIGURE 2: The flowchart of the proposed SVM-based lung classification.

identification, feature computation, and training and classification. Lung region identification processes the original image to determine the mean intensity of global lung regions. Feature computation processes the multiblock uniform local binary pattern (ULBP) to calculate each pixel of lung regions and count on the histogram. Finally, support vector machines train and classify normal lungs from abnormal lungs using ULBP features. The flowchart of the support vector machine- (SVM-) based common pneumothorax identification is shown in Figure 2.

*2.2. Identification of Lung Regions.* The lung region identification is used to segment the objective global lung regions from original X-ray chest image. After image loading, the mean intensity values of the original image have to be enhanced to the range of intensity from 0 to 1 based on gamma correction to provide useful image intensity details through adjustment of the gamma value [11]. In this process, we assign normalized gamma value to be 2, because the normalized gamma function can use the pixel intensity to provide multiequalization darkness preservation. Consequently, the higher gamma parameter generates a more significant adjustment to display the lung region. The enhancement formula is described as follows.

$$\text{enh}(h, w) = \left( \frac{I(h, w)}{\text{Max\_graylevel\_num}} \right)^\gamma, \quad (1)$$

where  $I(h, w)$  is the pixel value of the original image,  $\text{Max\_graylevel\_num}$  is the number of the original intensity range, and  $\gamma$  is the gamma value. Figure 3(a) shows an example of original image, and Figure 3(b) shows enhancement result. Because several lighter pixels surrounded by dark pixels lead

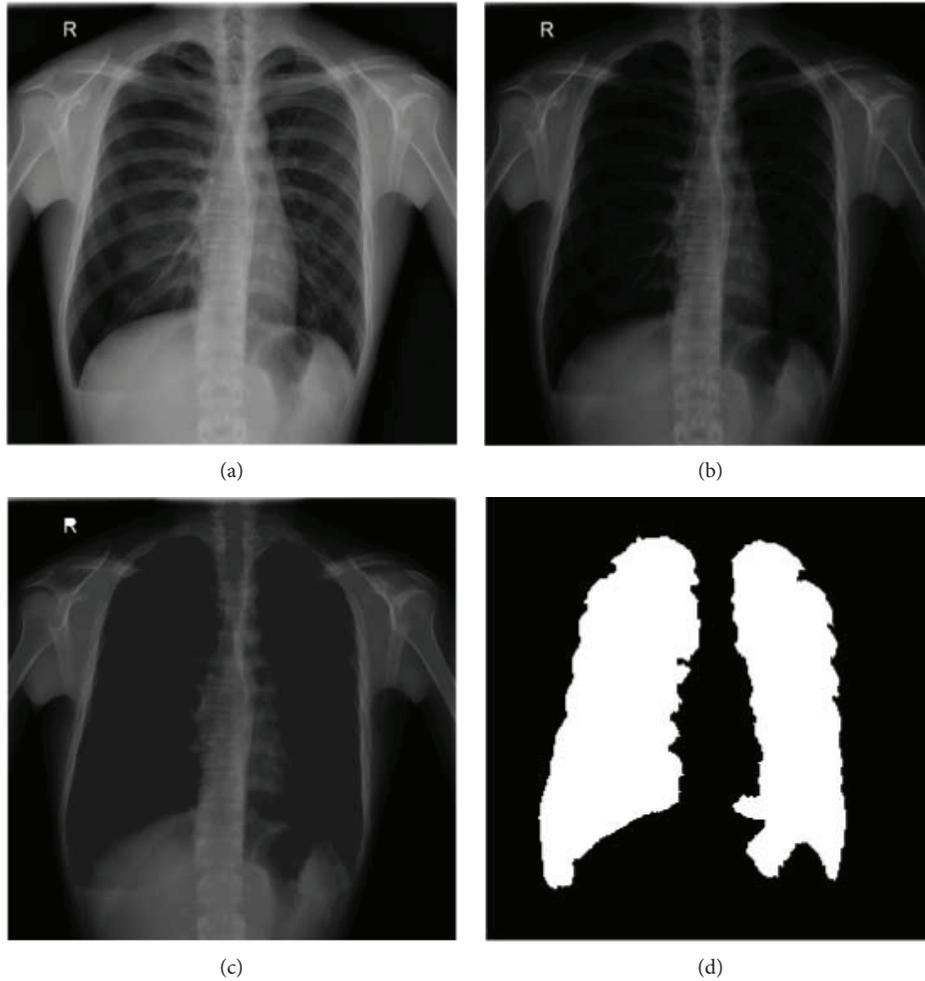


FIGURE 3: The results of an example (a) original image, (b) the enhanced image, (c) enhanced by hole-filling image, and (d) image lung region identification.

121	255	60
46	110	10
51	111	105

(a)

1	1	0
0	C	0
0	1	1

(b)

FIGURE 4: An example of LBP. This result values encoded as 11001100 by starting from the upper left and reading clockwise.

to generate holes of intensity pixels in the enhanced image, the connect-component method is applied to fill the holes in the lighter pixels surrounded by dark pixels [12].

Finally, the enhanced image defines the integrated lung regions via the Otsu algorithm [13] and transforms into a binary image, and then the lung regions are identified based on 8-connected neighborhood method to find two maximum objects that are able to fill the empty spaces within the regions; Figure 3 shows an example.

**2.3. Local Binary Pattern Feature Extraction.** By following the lung region identification, local binary pattern (LBP) is the primary technique in the section. LBP is a local texture operation technique that is considered to be a simple and efficient texture operator. LBP operator uses the center as thresholding for each pixel and compares with neighbor pixel, then the results transformed into the value of binary. Each result of binary value is counted on a histogram. For example, original values are shown in Figure 4(a) in the 3 × 3-sized

block. Figure 4(b) uses the center as the threshold in the 3×3-sized block. Finally, each 8-bit LBP is counted on 256-bin histogram.

In addition, LBP operator could adopt the uniform patterns to reduce the length of the feature vector and implement rotation invariant descriptor in a simple way. In this section, the features are calculated through a uniform local binary pattern (ULBP). Uniform pattern is contained at most two bitwise transitions from 0 to 1. For example, the patterns 00000000 (0 transitions), 01111100 (2 transitions), and 11001111 (2 transitions) are uniform. And the patterns 11001101 (4 transitions) and 01010110 (6 transitions) are not uniform. Computation of the ULBP histogram has a separate bin (e.g., the pattern 01000000 and the pattern 00100000 have the same bin).

There are 8 bins for each combination with no more than two conversions (e.g., 10000000 > 01000000 ... > 00000001, 11,000,000 > 01100000 > ... > 0000001111100000 > 0111000-0 > ... > 00000111), and each will have 8 patterns for every uniform pattern. And all nonuniform patterns are assigned to a single bin. ULBP image is constructed by  $d \times d$  size. ULBP operator is used for overlapping and scanning lung regions, then generates pattern histogram from the lung image for overlapping and scanning lung regions in lung identification. The neighboring correlation of pixel pattern histogram after encoding transformation in a moving window is defined as

$$X_n = \sum_{i=1}^h \sum_{j=1}^w H\{ULBP(i, j) == k(n)\}, \quad n = 1, 2, \dots, 58, \quad (2)$$

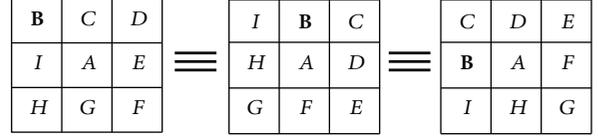
where  $X_n$  is the number of the  $n$ th bin in the histogram.  $h$  is the height of the lung image and  $w$  is the width of the lung image.  $k$  is a set, which includes 58 uniform binary patterns corresponding to the integers 0, 1, 2, 3, 4, 6, 7, 8, 12, 14, 15, 16, 24, 28, 30, 31, 32, 48, 56, 60, 62, 63, 64, 96, 112, 120, 124, 126, 127, 128, 129, 131, 135, 143, 159, 191, 192, 193, 195, 199, 207, 223, 224, 225, 227, 231, 239, 240, 241, 243, 247, 248, 249, 251, 252, 253, 254, and 255.  $i$  and  $j$  are the pixel coordinates of the lung image.  $H\{S\}$  is equal to 1 if  $S$  is true, and  $H\{S\}$  is equal to 0 if  $S$  is false.  $H$  is the part of the histogram and  $n$  is the value of bin that judging  $(i, j)$  as the center of the ULBP value is equal to which bin in the histogram corresponding position accumulation

$$ULBP(i, j) = \min \left( \begin{array}{c} \text{ROT} \\ 0 \leq b < 8 \end{array} [LBP(i, j)]_b \right). \quad (3)$$

ROT is a function fixed with the central pixel, while the neighbors are circularly rotated;  $b$  is the number of  $b$ th rotations. An example is shown in Figure 5.

$$LBP(i, j) = \sum_{k=0}^d 2^k \text{Win}(x_k - x_c), \quad (4)$$

$$\text{Win}(x_k - x_c) = \begin{cases} 1, & \text{if } x_k \geq x_c, \\ 0, & \text{otherwise,} \end{cases}$$



ROT: "ABCDEF GHI"  $\equiv$  "AIB CDEFGH" ...  $\equiv$  "ACDEFGHIB"

FIGURE 5: The ROT function, as the same type in the case of clockwise rotation.

where LBP is a function value where  $x_c$  is a center of the  $d \times d$ -sized block, and  $x_k$  is the neighboring pixel.  $\text{Win}$  is a binary thresholding function. Calculate the LBP straight:  $x_c$  as the center point with the adjacent point  $x_k$  comparison; if greater than or equal to the adjacent point, set to 1 and vice versa 0.

**2.4. Machine Learning Algorithm.** After feature computation, the classification of common pneumothorax is determined by support vector machines. Generally, classification includes input data and divides into training and testing sets. The training set has one label value (class labels) in each lung image in which normal lung image is set to 0, otherwise is set to 1. And each lung image has multiple attributes (features). The attributes are equal to each bin value on ULBP histogram from the feature computation. The features obtained 118 dimensions by using ULBP operator with the  $9 \times 9$ -sized block and  $11 \times 11$ -sized block. These constitute the training vectors through the use of the RBF kernel function. Finally, SVM constructs a model classification for normal lung or abnormal lung.

### 3. Rapid Pathology Detection for Chest X-Ray Images Using Texture Segmentation Method

**3.1. The Proposed Scheme.** The multiscale intensity texture segmentation algorithm is composed of lung object identification, target region texture analysis, multiscale region segmentation, and target region detection from chest X-ray images. The lung object identification processes the original image to discover the global lung regions. The target region texture analysis processes the global lung regions to calculate the texture distribution and defines the initial target regions based on the local binary pattern (LBP) [14]. The identification of the multiscale region definition; the smooth, complex regions; and the rib boundaries is performed by the different direction intensity distribution of global lung regions. Finally, the target region detection processes the initial target regions and results of multiscale region segmentation to estimate the relationships and to detect the final target regions. The flowchart of the proposed multiscale intensity texture segmentation is shown in Figure 6.

**3.2. Identification of Lung Regions.** The lung region identification is used to segment the objective global lung regions from original X-ray chest image. After image loading, the mean intensity values of the original image have to be enhanced to the range of intensity from 0 to 1 based on gamma correction to obtain the useful enhanced image

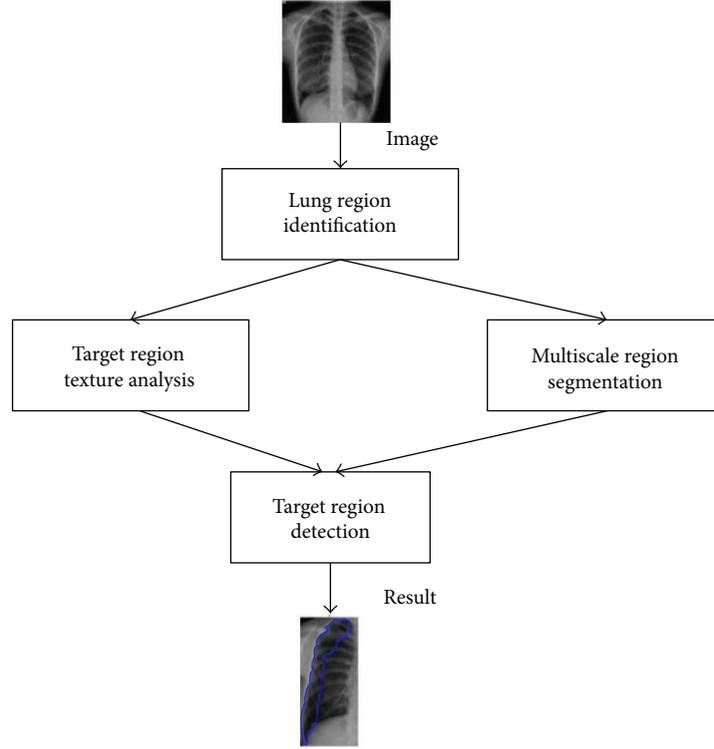


FIGURE 6: The flowchart of the proposed multiscale intensity texture segmentation.

through adjustment of the gamma value [11]. Finally, the enhanced image defines the integrated lung regions via the Otsu algorithm [13] to transform the binary image, and then the lung regions are identified based on 8-connected neighborhood method to find two maximum objects that are able to fill empty spaces within the regions.

**3.3. Texture Analysis of Abnormal Regions.** Radiograph construction according to X-ray radiation and interactions with human tissue are composed of 12 bits per pixel whose value corresponds to 0–4095 [15]. The high-density human tissue, such as the bones and lung, can display light intensity by using absorption spectrum. Reversely, the low-density human tissue, such as pneumothorax, displayed dark intensity using attenuation spectrum. The pixel intensity distribution of the original image between 0 and 4095 has sufficient information to describe human tissues, but that may lead to highlight noise. Based on the reason, the 12 bits of the original image are transformed to 8-bit image to reduce the noise for subsequent processing.

By following lung region identification, the abnormal regions are recognized by target region texture analysis within the low density of the abnormal region. The proposed target region texture analysis according to the local binary pattern (LBP) defines the smooth and complex regions. LBP image is constructed by  $d \times d$ -sized LBP operator for overlapping and scanning lung regions, then generates pattern histogram from neighboring correlation of patches with  $R \times R$  size for overlapping and scanning lung regions in lung identification. The neighboring correlation of pixel

pattern histogram  $\text{patchPattern}_{R,d}\{X_0, X_1, \dots, X_{255}\}$  after decimal transformation in a moving window is defined as

$$\text{Target}(\text{patch}(x_c, y_c)) = \begin{cases} 1, & \text{if } \text{patchPattern}_{R,d}(X_0) = R \times R, \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where  $\text{patch}(x_c, y_c)$  is a patch with center  $(x_c, y_c)$  and  $R \times R$ -sized block, and if the value of  $\text{patchPattern}(X_0)$  is equal to  $R \times R$ , then  $\text{Target}(\text{patch}(x_c, y_c))$  is set to 1; otherwise,  $\text{Target}(\text{patch}(x_c, y_c))$  is 0. The center point of 1 in the  $R \times R$  region will be marked as 1. It is judged that the matrix sum of the size  $R \times R$  centering on  $x_c$  and  $x_y$  is set to 1 if it is equal to  $R \times R$  and conversely 0.

$$X_k = \sum_{i=1}^R \sum_{j=1}^R H\{LBP(i, j) == k\}, \quad k = 0, 1, \dots, 255, \quad (6)$$

where  $X_k$  is the number of the  $k$ th pattern.  $i$  and  $j$  are pixel positions of a patch.  $H\{S\}$  is equal to 1 if  $S$  is true, and  $H\{S\}$  is equal to 0 if  $S$  is false.

$$\begin{aligned} LBP(i, j) &= \sum_{k=0}^d 2^k \text{Win}(x_k - x_c), \\ \text{Win}(x_k - x_c) &= \begin{cases} 1, & \text{if } |x_k - x_c| \geq 5, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (7)$$

where LBP is a function value where  $x_c$  is a center of the  $d \times d$ -sized block, and  $x_k$  is the neighboring point pixel.

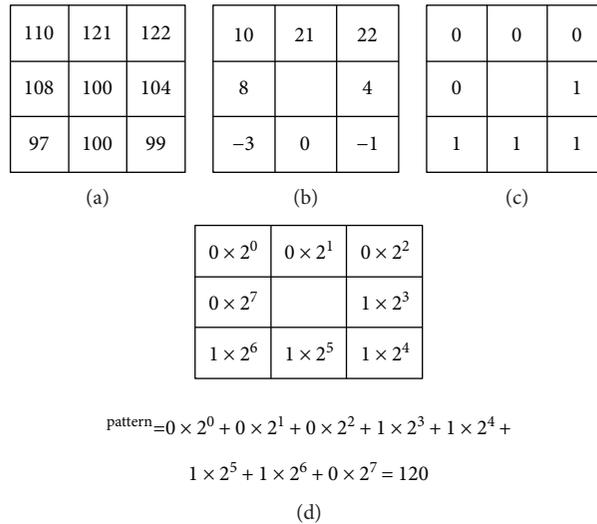


FIGURE 7: An example of the pattern generation: (a) the original pixel values, (b) difference value of the neighboring point and center point, (c) result of the Win function, and (d) the pattern value of the LBP function.

Win is a binary thresholding function. An example is shown in Figure 7.

And the formula is similar to (4) but after subtracting the absolute value to see whether greater than or equal to 5 is set to 1 instead of 0.

**3.4. Multiscale Region Segmentation.** After lung region identification, the noise of lung regions is eliminated by Gaussian filter [16] with  $5 \times 5$  mask and standard deviation set as 6 to smooth boundaries of the lung regions. The intensity distributions of horizontal and vertical patterns are constructed from calculating difference values of gradients of horizontal and vertical direction through the Gaussian-filtered image. To segment the regions of different pixel intensity distributions, the constructing patterns of horizontal and vertical direction pattern set are employed for region segmentation.

The noise elimination of image is to calculate the difference of intensity between adjacent pixels at horizontal and vertical directions, respectively, for the difference value is able to describe the distribution of pixel intensity. When concerning the horizontal direction, the distribution of pixel intensity is smooth if the difference of value is positive, while the distribution of pixel intensity is complicated if the difference of value is negative. When the distribution of pixel intensity is equal to zero, the distribution of pixel intensity is invariant. At the vertical direction, the distribution of pixel intensity is regarded as the upper boundary of the rib ( $V_1$ ) if the difference of value is positive while the distribution of pixel intensity is regarded as the lower boundary of the rib ( $V_2$ ) if the difference of value is negative. The results are shown in Figures 8(d) and 8(c). The different parts of the lung tissue have different pixel intensity distribution.

Next, smooth region, complex region, and rib boundary are determined using the intensity distribution of horizontal scanning and vertical scanning individually, based on the

patch of  $5 \times 5$ . According to the calculation of difference values, a positive value is set to 1, and a negative value is set to  $-1$ . In the horizontal direction, if the number of 1's is larger, the centroid point of the patch is set to 1. Reversely, if the number of  $-1$ 's is larger, the centroid point of the patch is set to  $-1$ . Otherwise, the centroid point of the patch is set to 0. The points set to 1 are defined as  $H_1$  region and the points set to  $-1$  are defined as  $H_2$  region. The results are shown in Figures 8(b) and 8(c). Finally, consider the intersection region of the target and  $H_1$  and the intersection region of the target and  $H_2$ . The results which are larger are regarded as smooth regions, and the rest are complex regions. The result of  $V_1$  union  $V_2$  is rib boundary.

**3.5. Detection of Abnormal Regions.** After lung regions are determined, the intersection of smooth regions and complex regions will be labeled. Then, more precise smooth regions, constructed by calculating difference value of the standard deviation of the target regions and overlapping region, can be spotted. Firstly, combine the  $V_1$  and  $V_2$  patterns to generate the rib boundaries and remove the rib boundaries for smooth regions and complex regions, as shown in Figures 9(a) and 9(b). Because the rib boundaries do not belong to smooth or complex regions, the process leads to identifying the overlapping regions of smooth and complex regions, as shown in Figure 9(d). For the definition of the new overlapping regions, consider the standard deviation of the overlapping region and the target region. This process calculates the standard deviation for all regions. Because the intensity distribution of the overlapping regions and intensity distribution of the target regions are similar, their standard deviation is naturally similar. Therefore, use the standard deviation of the target region to redefine and determine the overlapping region to be smooth or complex region. If the standard deviation of overlapping region is less than plus or minus 10 percent of the target region, the overlapping region is

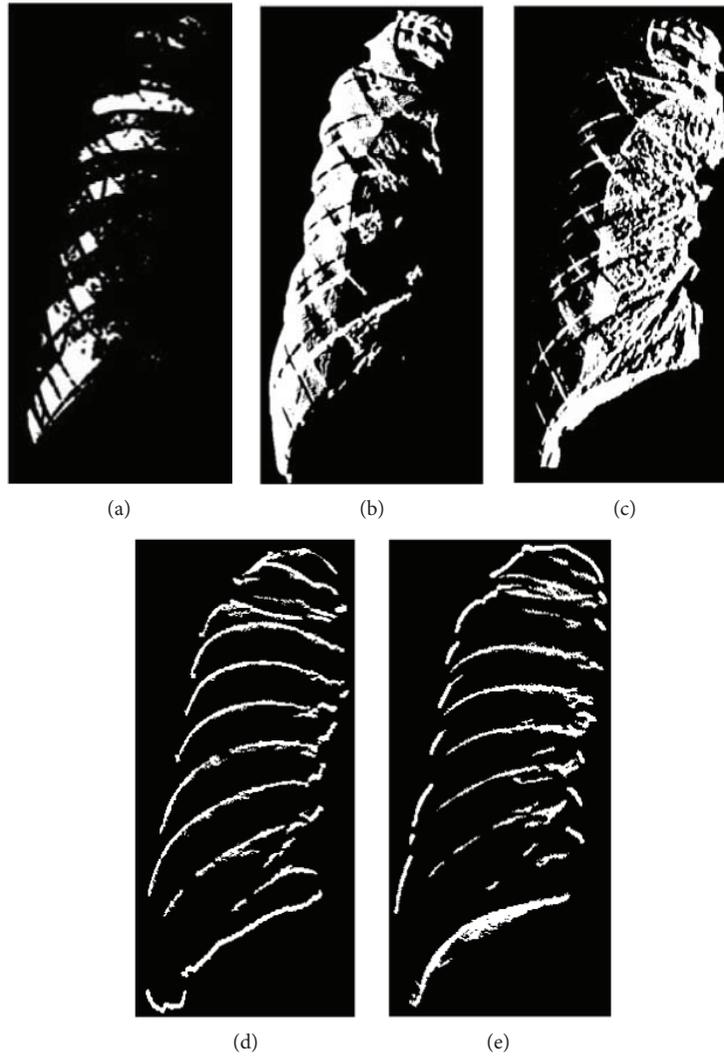


FIGURE 8: The segmentation results: (a) target region, (b)  $H_1$  intensity distribution pattern set, (c)  $H_2$  intensity distribution pattern set, (d)  $V_1$  intensity distribution pattern set, and (e)  $V_2$  intensity distribution pattern set.

recognized as the smooth region; otherwise, the overlapping region is removed from the smooth region. Finally, each smooth region is determined as a pneumothorax region. In the rib boundary that is located between the pneumothorax regions, as shown in Figure 9(e), the morphology operations are applied to remove the noises for obtaining final pneumothorax region, as shown in Figure 9(e). The results are then marked on the original image as shown in Figures 9(g) and 9(h).

## 4. Experimental Results

**4.1. Data Exploratory.** 32 chest radiographs, including traumatic pneumothorax and spontaneous pneumothorax from 32 patients (age range, 18–65 years) and 10 normal chest radiographs, used in this study are acquired from the Department of Medical Imaging, Chung Shan Medical University Hospital, in Taichung, Taiwan. The dataset was

approved by the Institutional Review Board of the Chung Shan Medical University Hospital, in Taichung, Taiwan.

**4.2. Pneumothorax Prediction Accuracy.** The ULBP histogram for  $11 \times 11$  block and  $9 \times 9$  block merged into the final histogram is calculated by using SVM for training and classification. With a sample of 42 images, each image was divided into two lung images (A total of 32 pneumothorax lung cases and 52 normal lung cases). 70% lung cases (36 normal cases and 22 pneumothorax cases) were used for the training phase, while 30% lung cases (16 normal cases and 10 pneumothorax cases) were used for the testing phase. The 5-fold cross-validation result with accuracy (Acc.) variation from 76.9%–88.4% is shown Table 1.

**4.3. Pneumothorax Segmentation Accuracy.** The multi-scale intensity texture is calculated by the intensity of each pixel in each block. Finally, the pneumothorax region is detected through standard deviation to evaluate

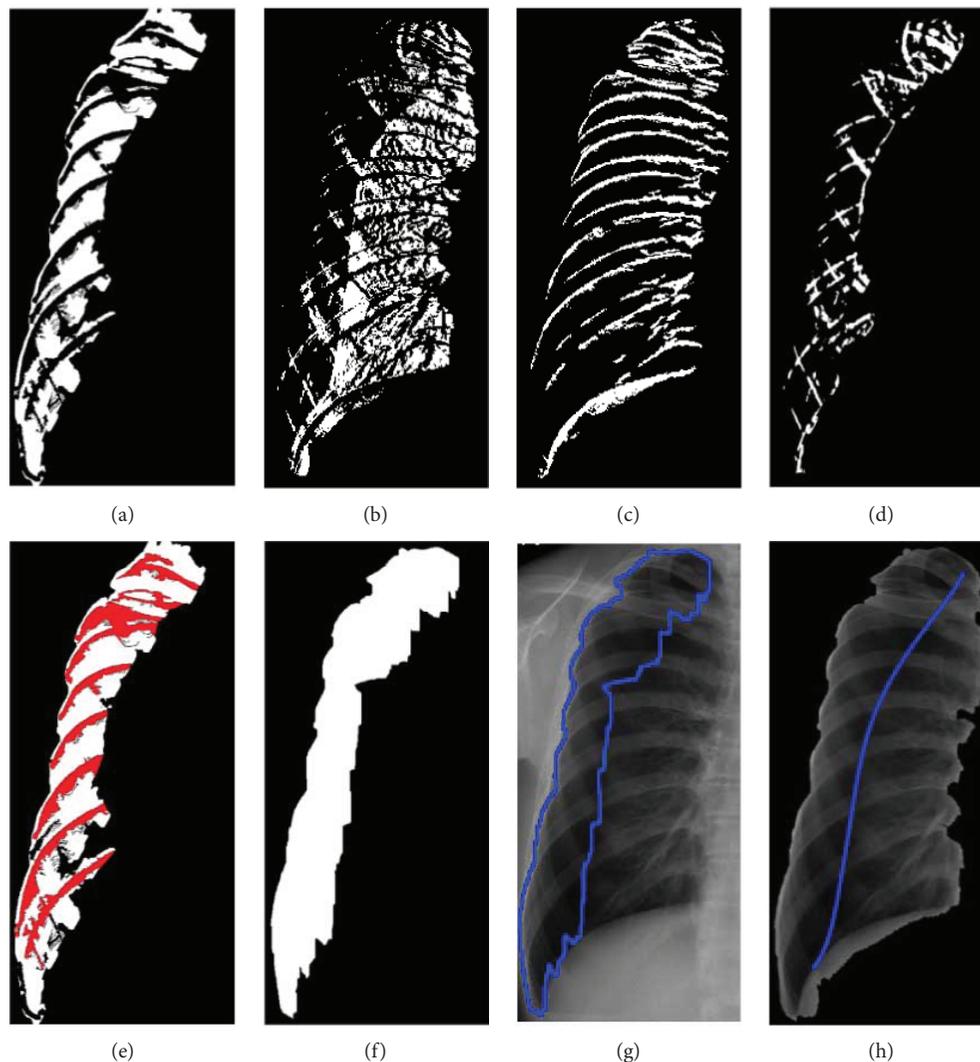


FIGURE 9: The results of an example of (a) smooth regions, (b) complex regions, (c) rib boundary regions, (d) intersection region between smooth and complex regions, (e)–(f) final segmentation region, and (g)–(h) original image.

TABLE 1: 5-fold cross-validation result.

Fold	1	2	3	4	5	Acc.
Acc.	76.9%	88.4%	88.4%	80.7%	76.9%	82.2%

the final recognition of the region of pneumothorax. Three objective evaluations of the accuracy, precision, and sensitivity ratio for comparing the segmentation effects are defined as

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\
 \text{Precision} &= \frac{TP}{TP + FP}, \\
 \text{Sensitivity} &= \frac{TP}{TP + FN},
 \end{aligned} \tag{8}$$

where TP is when both the prediction outcome and the actual value are positive, TN is when both the prediction outcome and the actual value are negative, FP is when the prediction

outcome is positive while the actual value is negative, and FN is when the prediction outcome is negative while the actual value is positive.

Figure 10 shows the comparison between the area that is manually depicted of radiologists and the proposed method. Figures 10(a) and 10(d) are original images, (b) and (e) are pneumothorax regions by manual of radiologists, and (c) and (f) are pneumothorax regions by the proposed method. The accuracy, precision, and sensitivity of all dataset images by using the proposed method are shown in Figure 11.

*4.4. Comparisons of the Proposed Method.* From the experimental results in Figure 11, the accuracy is affected to discern the region between smooth and complex areas in different block sizes. The block sizes of different judgments of the image intensity provide different information; the segmentation will be different in the result. In particular, the image intensity at the boundary and in the vicinity of the clavicle is relatively complicated. In Table 2, the average accuracy of the  $5 \times 5$  block compared to the  $11 \times 11$  block decreases from

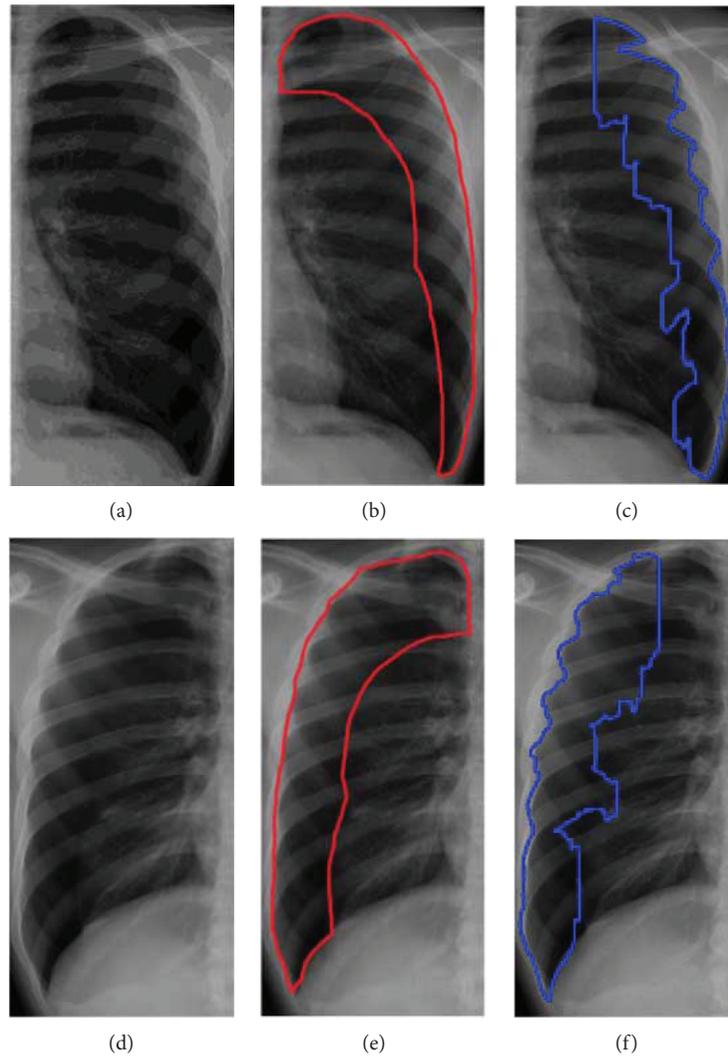


FIGURE 10: Comparison between the area that is manually depicted of radiologists and the proposed method. (a–c) Segmentation results of the left lung; (d–f) segmentation results of the right lung.

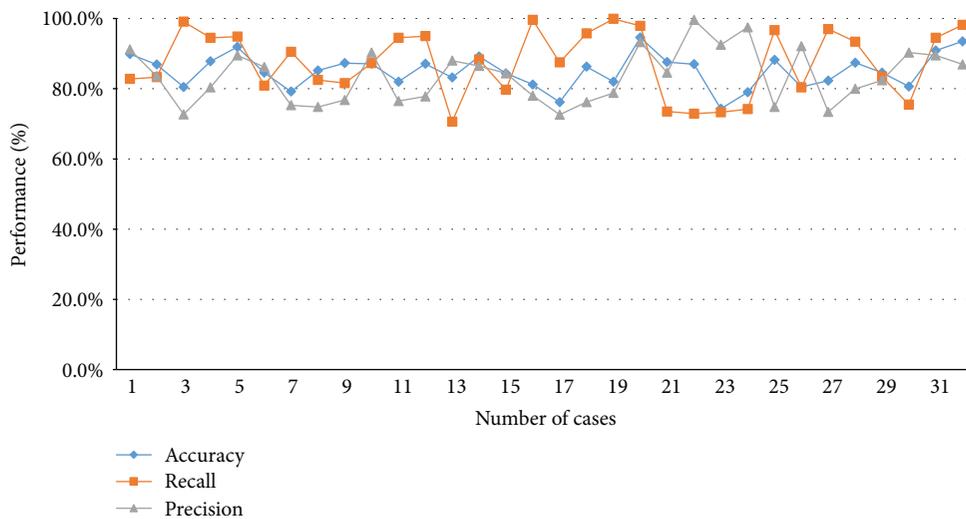


FIGURE 11: Accuracy, precision, and sensitivity of all dataset image using the proposed method.

TABLE 2: Comparisons of the accuracy, precision, and sensitivity for different patch sizes of the proposed method.

Patch size	Accuracy	Precision	Sensitivity
5 × 5	85.8%	83.6%	87.4%
7 × 7	84.1%	84.4%	85.5%
9 × 9	83.2%	85.1%	82.3%
11 × 11	81.1%	86.2%	81.6%

85.8% to 81.1%. However, the block size is too small to be used because the region will reveal too little information and then the pneumothorax region cannot be accurately judged. And the block size will increase when the precision increases, but the sensitivity reduces for the evaluation of the precision; it referenced FP and TP, and the sensitivity referenced FN and TP. In the results, when the size of the selected block is larger, the FP decreases but the FN increases. In conclusion, the final cutout pneumothorax region provides a better result for 5 × 5 size of the patch.

## 5. Conclusion

The primary method in the paper is to segment the lung in the abnormal region through multiple overlapping blocks. The abnormal region is found by texture transformed from computing multiple overlapping blocks. Finally, this method effectively analyses lung diseases of the area in the chest X-ray image and improves the possible diagnosis of the missing problem of the pneumothorax area. This increases the efficiency for physicians to assess the extent of the treatment of pneumothorax, so as to support the radiologist to reduce workload.

This study presents a novel framework for automatic pneumothorax detection in CXRs. The texture analysis is based on intensity and gradient for pneumothorax detection. The pneumothorax case was a difficult judgment when pneumothorax region is extremely stenotic and close to the chest boundaries. In addition, pixels located near the chest boundaries tend to have less discriminative texture on image indication, because the bones and pleura existed in obvious edges, which reduced their correspondence of textures. Consequently, the texture characteristic in chest boundaries area is not as prominent as in the inner lung region. Discrimination in different lung regions and adding the texture weight may be the future research focus. The segmentation can increase the accuracy rate for the segmentation of pneumothorax region.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## References

[1] M. Noppen and T. De Keukeleire, "Pneumothorax," *Respiration*, vol. 76, no. 2, pp. 121–127, 2008.

[2] A. MacDuff, A. Arnold, J. Harvey, and on behalf of the BTS Pleural Disease Guideline Group, "Management of spontaneous pneumothorax: British Thoracic Society pleural disease guideline 2010," *Thorax*, vol. 65, Supplement 2, pp. ii18–ii31, 2010.

[3] A. P. Wakai, "Spontaneous Pneumothorax," *BMJ Clinical Evidence*, 2011, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3275306/>.

[4] N. Bellaviti, F. Bini, L. Pennacchi et al., "Increased incidence of spontaneous pneumothorax in very young people: observations and treatment," *Chest*, vol. 150, no. 4, p. 560A, 2016.

[5] T. Weber, F. Bayer, W. Haas et al., *Investigation of the Signature of Lung Tissue in X-Ray Grating-Based Phase-Contrast Imaging*, 2012, <http://arxiv.org/abs/1212.5031>.

[6] L. Cardinale, G. Volpicelli, A. Lamorte, and J. Martino, "Revisiting signs, strengths and weaknesses of standard chest radiography in patients of acute dyspnea in the emergency department," *Journal of Thoracic Disease*, vol. 4, no. 4, pp. 398–407, 2012.

[7] K. Doi and S. Sanada, *Method and System for Automatic Detection of Ribs and Pneumothorax in Digital Chest Radiographs*—US Patent 5,668,888, 1997.

[8] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4–5, pp. 198–211, 2007.

[9] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 733–746, 2011.

[10] O. Geva, G. Zimmerman-Moreno, S. Lieberman, E. Konen, and H. Greenspan, "Pneumothorax detection in chest radiographs using local and global texture signatures," in *Medical Imaging 2015: Computer-Aided Diagnosis*, Renaissance Orlando, Florida, at SeaWorld, March 2015.

[11] S. C. Huang, F. C. Cheng, and Y. S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1032–1041, 2013.

[12] A. Bush, H. Gray, and D. M. Denison, "Diagnosis of pulmonary hypertension from radiographic estimates of pulmonary arterial size," *Thorax*, vol. 43, no. 2, pp. 127–131, 1988.

[13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[14] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[15] T. Kimpe and T. Tuytschaever, "Increasing the number of gray shades in medical display systems—how much is enough?," *Journal of Digital Imaging*, vol. 20, no. 4, pp. 422–432, 2007.

[16] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 296–310, 1993.

## Research Article

# Design of a Clinical Decision Support System for Fracture Prediction Using Imbalanced Dataset

Yung-Fu Chen <sup>1,2,3,4</sup> Chih-Sheng Lin,<sup>1</sup> Kuo-An Wang,<sup>5,6</sup> La Ode Abdul Rahman,<sup>2</sup> Dah-Jye Lee <sup>4</sup> Wei-Sheng Chung,<sup>3,7</sup> and Hsuan-Hung Lin <sup>6</sup>

<sup>1</sup>Department of Radiology, BenQ Medical Center, The Affiliated BenQ Hospital of Nanjing Medical University, Nanjing, Jiangsu Province 210017, China

<sup>2</sup>Department of Healthcare Administration and Department of Dental Technology and Materials Science, Central Taiwan University of Science and Technology, Taichung 40601, Taiwan

<sup>3</sup>Department of Health Services Administration, China Medical University, Taichung 40402, Taiwan

<sup>4</sup>Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602, USA

<sup>5</sup>Department of Industrial Education and Technology, National Changhua University of Education, Changhua County 50007, Taiwan

<sup>6</sup>Department of Management Information Systems, Central Taiwan University of Science and Technology, Taichung 40601, Taiwan

<sup>7</sup>Department of Internal Medicine, Taichung Hospital, Ministry of Health and Welfare, Taichung 40343, Taiwan

Correspondence should be addressed to Hsuan-Hung Lin; [shlin@ctust.edu.tw](mailto:shlin@ctust.edu.tw)

Received 26 August 2017; Revised 11 January 2018; Accepted 23 January 2018; Published 22 March 2018

Academic Editor: Emiliano Schena

Copyright © 2018 Yung-Fu Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

More than 1 billion people suffer from chronic respiratory diseases worldwide, accounting for more than 4 million deaths annually. Inhaled corticosteroid is a popular medication for treating chronic respiratory diseases. Its side effects include decreased bone mineral density and osteoporosis. The aims of this study are to investigate the association of inhaled corticosteroids and fracture and to design a clinical support system for fracture prediction. The data of patients aged 20 years and older, who had visited healthcare centers and been prescribed with inhaled corticosteroids within 2002–2010, were retrieved from the National Health Insurance Research Database (NHIRD). After excluding patients diagnosed with hip fracture or vertebrate fractures before using inhaled corticosteroid, a total of 11645 patients receiving inhaled corticosteroid therapy were included for this study. Among them, 1134 (9.7%) were diagnosed with hip fracture or vertebrate fracture. The statistical results showed that demographic information, chronic respiratory diseases and comorbidities, and corticosteroid-related variables (cumulative dose, mean exposed daily dose, follow-up duration, and exposed duration) were significantly different between fracture and nonfracture patients. The clinical decision support systems (CDSSs) were designed with integrated genetic algorithm (GA) and support vector machine (SVM) by training and validating the models with balanced training sets obtained by random and cluster-based undersampling methods and testing with the imbalanced NHIRD dataset. Two different objective functions were adopted for obtaining optimal models with best predictive performance. The predictive performance of the CDSSs exhibits a sensitivity of 69.84–77.00% and an AUC of 0.7495–0.7590. It was concluded that long-term use of inhaled corticosteroids may induce osteoporosis and exhibit higher incidence of hip or vertebrate fractures. The accumulated dose of ICS and OCS therapies should be continuously monitored, especially for patients with older age and women after menopause, to prevent from exceeding the maximum dosage.

## 1. Introduction

Chronic respiratory diseases, including chronic obstructive pulmonary disease (COPD), asthma, bronchiectasis, allergic

rinitis and sinusitis, obstructive sleep apnoea syndrome, pulmonary hypertension, and other occupational lung diseases, are caused by disorder of the airways and other structures of the lung [1]. More than 1 billion people

suffer from chronic lung diseases, accounting for more than 4 million deaths annually [2]. Among them, more than 200 million people were afflicted by COPD causing 3 million deaths globally; more than 235 million people were affected by asthma resulting in 0.18 million deaths [3, 4] with over 80% of deaths found in countries with low- and lower-middle income [5]; and the incidence rate of bronchiectasis ranged from 4.2 per 100000 persons aged 18–34 years to 272 per 10000 persons aged 75 years and above in the US [6] with an increasing rate of 8.7%/year from 2001 to 2007 [7].

COPD is an inflammatory disease of the lung characterized by progressive airflow obstruction, systematic chronic inflammation, and recurrent acute exacerbation [8, 9]. WHO predicted that COPD will become the 3rd leading cause of death worldwide in 2030 [10]. Bronchiectasis is characterized by abnormal bronchial dilation and bronchial wall thickness, chronic infection and inflammation, recurrent cough and sputum production, and bacterial colonization and airflow obstruction, resulting in a decline in respiratory function [11]. Compared to other chronic diseases, asthma has a relatively lower fatality rate. Asthma is characterized by chronic airway inflammation with the history of respiratory symptoms like wheeze, shortness of breath, chest tightness, and cough accompanied with expiratory airflow limitation [12]. Oral corticosteroids (OCS) and inhaled corticosteroids (ICS) are usually prescribed by physicians to improve symptom, lung function, and quality of life, as well as to reduce repeated exacerbations for patients with asthma, COPD, and bronchiectasis. However, it was reported that corticosteroid use may increase the risk of fracture [13–15].

Clinical decision support systems (CDSSs) provide useful information and expert knowledge to assist healthcare providers to improve diagnosis and treatment outcomes, disease managements, and healthcare quality for patients in both home and clinical settings and have been shown to be capable of improving healthcare outcomes in medical practice [16]. CDSSs have been widely applied in disease diagnosis [17, 18], disease treatment and management [19–21], medical alerting or event reminding [22–24], and drug dosing or medication prescribing [25, 26].

*1.1. Problems Encountered in CDSS Design with Imbalanced Datasets.* Taiwanese National Health Insurance program is a single-payer, compulsory insurance system that was established in 1995 by the Bureau of National Health Insurance (NHI), Ministry of Health and Welfare. The insurance program provides healthcare to 99% of the 23.74 million citizens of Taiwan and maintains contracts with 97% of the nation's healthcare institutions. The National Health Research Institute is authorized to establish the National Health Insurance Research Database (NHIRD), as well as to manage registration and claim data for the 23 million insured citizens. Most studies used a subset of the NHIRD that consisted of 1 million randomly sampled beneficiaries enrolled in the NHI program. According to a PubMed website search conducted in July 2016, around 4000 studies investigated using NHIRD have been published [27]. The high accuracy and validity of

ICD-9-CM diagnoses in the NHIRD have been described in previous studies [28]. Recently, we have conducted several studies to discover the association between an individual disease with the risk of acquiring another diseases based on the data retrieved from the NHIRD, for example, the investigation of inhaled corticosteroids with pulmonary tuberculosis [29], sleep disorders with erectile dysfunction [30], gout with erectile dysfunction in men [31], and bronchiectasis with ischemic stroke [32]. Because most of the datasets retrieved from the NHIRD for the above-mentioned studies are imbalanced with ratios of positive samples (minority) to negative samples (majority) ranging from 1:4 to 1:5, the decision hyperplane will bias toward the majority class when adopting the accuracy-driven algorithm to design the CDSS.

Minority cases refer to rare patterns or abnormal behaviors that are difficult to detect but are often important. In real application domains, such as oil spill detection in satellite radar images [33], text classification [34], financial fault detection [35, 36], customer identification [37], medical diagnosis [38, 39], and others, classification of imbalanced datasets incurs critical problems. One problem encountered in classifying unbalance datasets is that samples of one class (majority) outnumber the samples in another class (minority) that is of often more interest or importance, making the algorithms driven by accuracy bias toward the majority class [40]. Generally, although the classification accuracies are satisfactory in the classifiers designed based on the accuracy-driven (treating accuracy as the fitness function) algorithms, their sensitivities are quite low. Hence, alternative fitness functions, such as area under ROC curve (AUC) [41] or weighted sum of accuracy, sensitivity, and specificity, are proposed to solve these problems.

*1.2. State-of-the-Art Methods for CDSS Design.* The classifiers built based on logistic regression, decision tree, standard neural network, and support vector machine are generally suitable for balanced data only. When dealing with imbalanced data, these classifiers often bias toward the majority cases while distorting the minority cases. To improve the predictive performance, the methods for modeling imbalanced datasets include data preprocessing, cost-sensitive learning, and kernel-based methods [42].

Resampling data in the sample space and selecting features in the feature space are commonly used preprocessing strategies for dealing with imbalanced datasets. There are 3 categories of resampling methods, including undersampling, oversampling, and hybrid sampling methods, applied for rebalancing the imbalanced datasets. Oversampling methods are used to deal with the minority samples by generating new minority samples, while undersampling strategies are applied to randomly discard the majority samples to balance the datasets. The hybrid sampling is a combination of both methods.

Undersampling includes random undersampling and informed undersampling [42]. The major drawback of undersampling is that important information may be lost due to the removal of some data points. There is no specific mechanism under random undersampling, which only

functions randomly. Other undersampling methods like one-sided selection [43], BalancedCascade, and EasyEnsemble [44] are called informed undersampling. One-sided selection obtains the balanced training set from an imbalanced one by keeping the minority samples untouched while selectively removing the borderline majority samples by applying Tomek link concept [43]. It eliminates boundary samples and deals with only a subset of majority samples. Devi et al. [45] proposed a modified Tomek link-based undersampling scheme to eliminate, besides boundary samples, outlier and redundant samples to improve the one-sided selection method. In EasyEnsemble and BalancedCascade, several subsets of training data are sampled from the majority samples and each subset is combined with the minority samples for training a model. The generated models are then combined for making the final decision [44].

The simplest oversampling method is random oversampling, in which the minority samples are randomly duplicated. A critical problem of random oversampling is overfitting. Another major approach in oversampling is synthetic minority oversampling technique (SMOTE), which generates synthetic samples on the line segments connecting each minority sample to its  $k$ -nearest neighbors in minority class [46]. A major problem of SMOTE is that it blindly generates synthetic samples without considering the majority data points located close to the minority samples, resulting in overlapping between classes. Later, several methods, such as Borderline-SMOTE [47], Safe-Level-SMOTE [48], and Cluster-SMOTE [49], extending the conventional SMOTE were proposed to improve the performance.

Cost-sensitive learning is useful for handling imbalanced healthcare data since classifying a minority (positive) sample to the majority (negative) class often costs more than classifying a majority sample to the minority class. There are 3 major categories generally found in cost-sensitive learning: approaches that assign different weights to samples, ensemble schemes that integrate with cost-sensitive methods, and methods that incorporate the misclassified costs directly into the classifiers [42]. The first approach is motivated by the AdaBoost scheme which trains an initial model based on the original imbalanced dataset and identifies misclassified samples. More weight will be assigned to the misclassified samples in the following iterations until the classifier significantly improves. In the second scheme, boosting approach is generally used for improving the imbalanced data problem and multiple base learners are trained to solve the classification problem. Many types of ensemble learning methods like bagging, boosting, and stacking have been proposed to combine base learners according to different strategies. In the third method, by considering costs of misclassification differ among different classes, the classifier is designed by minimizing the total misclassification cost. In the current study, the strategy for designing the CDSSs is to weight more on sensitivity in the objective function or to maximize the AUC for achieving higher sensitivity in detecting more patients (minority cases) to increase their well-being.

Support vector machine (SVM) is mostly integrated in kernel-based methods for imbalanced data learning. In

Farquad and Bose [40], oversampling of the minority class was achieved by training an SVM model with lower accuracy but higher sensitivity by increasing the value of SVM model parameter  $C$  to convert the misclassified majority samples into minority samples. The reconstructed more balanced training dataset was then applied for training the classification models with better predictive performance using different artificial techniques like multilayer perceptron (MLP), logistic regression (LR), and random forest (RF). In contrast, Jian et al. [50] applied the biased SVM to identify the support vectors and nonsupport vectors of the imbalanced training samples and then used SMOTE and random undersampling methods to resample the support vectors in the minority class and nonsupport vectors in the majority class, respectively. Recently, Piri et al. [42] proposed a new SMOTE algorithm by oversampling the informative minority samples near the SVM decision boundary. Additionally, they focused on misclassified informative minority samples by oversampling them with a higher degree than the correctly classified minority samples. The algorithm generates much less synthetic samples and is more efficient than SMOTE [46], Borderline-SMOTE [47], Safe-Level-SMOTE [48], and Cluster-SMOTE [49].

According to the aforementioned description, the datasets retrieved from the NHIRD for public health studies are mostly imbalanced with minority to majority sample ratios ranging from 1:4 to 1:5, and the decision hyperplane tends to bias toward the majority class when adopting the accuracy-driven algorithm to design the CDSS. The dataset adopted in this study is even more imbalanced with the minority to majority ratio approximates to 1:9. The objectives include the following: (1) investigate the association between corticosteroid use and fracture using NHIRD and (2) design the CDSSs to predict fracture occurrence for patients with chronic respiratory diseases prescribed with corticosteroids by dealing with the imbalanced dataset. The preliminary results were reported in [51].

## 2. Materials and Methods

*2.1. Data Sources.* Data of the patients who were 20 years old or older visiting healthcare centers (outpatients or inpatients) and had been prescribed with inhale corticosteroids because of diagnosed asthma (ICD-9-CM 493), chronic obstructive pulmonary disease (COPD) (ICD-9-CM 491, 492, and 496), or bronchiectasis (ICD-9-CM 494) within 2002–2010 were retrieved from the NHIRD for this investigation. Patients diagnosed with hip or vertebrate fractures before using inhaled corticosteroid were excluded. The outcome measure was hip fracture (ICD-9-CM 820) or vertebral fracture (ICD-9-CM 805 and 806). The data of 11645 patients, including 1134 patients with the fracture (614 men and 520 women) and 10511 patients without fracture (6211 men and 4300 women), were retrieved from the NHIRD within 2001–2013 and were used for statistical analysis and CDSS design. The patients were divided into 4 groups according to age: 20–40, 41–50, 51–64, and  $\geq 65$  years. The respiratory diseases included were asthma (ICD-9-CM codes 493),

COPD (ICD-9-CM codes 491, 492, and 496), and bronchiectasis (ICD-9-CM codes 494). Other comorbidities included diabetes (ICD-9-CM codes 250), cancer (ICD-9-CM codes 140–208), liver cirrhosis (ICD-9-CM codes 571.2 and 571.5), end-stage renal disease (ICD-9-CM codes 585), and osteoporosis (ICD-9-CM codes 722.0).

SPSS 22.0 software (IBM.SPSS) was adopted for all statistical analyses. Difference in the proportional distribution of the demographic characteristics and comorbidities of the patients with fracture and those without fracture were compared and tested using the Chi-square test. The overall, age-specific, and comorbidity-specific incidences of fractures in both fracture and nonfracture groups were also compared. The cumulative dose, mean follow-up daily dose, and mean exposed daily dose of prescribed corticosteroid were also analyzed using unpaired Student's *t*-test. The statistical significance was defined as  $p < 0.05$ .

**2.2. Design of Clinical Decision Support Systems.** As shown in Figure 1, a wrapper method integrating genetic algorithm (GA) and support vector machine (SVM) was used for designing the CDSSs with the former adopted for selecting salient features and adjusting the SVM parameters (cost value and kernel parameter) whereas the latter for classifying different classes and calculating fitness values based on the objective functions [19]. For each iteration, the  $n$  chromosomes were updated by combining  $n/2$  new chromosomes generated from crossover with the other  $n/2$  chromosomes obtained from mutation. The parameters of the genetic algorithm were set as follows: number of initial chromosome population ( $n$ ) = 10, maximum number of iterations with unimproved fitness value (MaxNotImproved) = 500, and maximum number of total iterations (MaxIteration) = 100000. When the current iteration (CurrentIteration)  $\geq$  MaxIteration or the number of iterations with unimproved fitness (NumOfNotImproved)  $\geq$  MaxNotImproved, the program terminated.

**2.2.1. Data Preprocessing and Preparation.** The data retrieved from the NHIRD were divided into training and testing datasets, each contains 50% of the samples in majority class and minority class, respectively, that is, 567 samples with fracture and 5255 samples without fracture. Random undersampling, cluster-based undersampling [52], and one-sided selection [43] methods were used to prepare the training datasets before designing the CDSSs. After random undersampling, the balanced training set consisted of  $2m$  samples by including all the  $m$  minority samples and randomly selecting  $m$ -nonduplicated samples from the  $M$  ( $M > m$ ) samples in the majority class. In contrast, in the cluster-based undersampling, the samples in the majority class were clustered into  $m$  clusters that each consisted of  $\lfloor M/m \rfloor$  or  $\lfloor M/m \rfloor + 1$  samples. And then, the kNN algorithm was applied to select the sample which was nearest to the center of gravity in each cluster, resulting in a balanced training set containing  $2m$  samples for cross-validation. After the CDSSs had been trained, the imbalanced testing dataset was used for testing the predictive performance of the models.

**2.2.2. Model Training and Testing.** Independent training and testing (ITT) was adopted for training and testing effectiveness of the CDSS [53] that 50% of the data were used for training and validating and the rest 50% for testing its predictive performance by calculating the accuracy, sensitivity, specificity, G-mean, and AUC. In the training phase, tenfold cross-validation was adopted for training and validating the models for obtaining a model with the best performance. For cross-validation, all the sample data in the training set were randomly divided into 10 clusters (folds), in which any combined nine folds were used for training while the remaining one for validating the CDSSs. The procedure was repeated for 10 times.

In the testing phase, the imbalanced testing dataset was applied for testing the designed CDSSs obtained in the testing phase. When designing the CDSSs, selection of the objective function is crucial in obtaining optimal CDSS models. In this study, the cost-sensitive objective functions, including AUC and combined accuracy, sensitivity, and specificity, were used to obtain the optimal CDSSs with imbalanced datasets. Equation (1) shows the objective function which combines the accuracy, sensitivity, and specificity. Equation (2) shows the AUC treated as the objective function.

$$\text{OB1} = \text{AC} - |\text{SE} - \text{SP}|, \quad (1)$$

$$\text{OB2} = \text{AUC}. \quad (2)$$

In (1), we intended to maximize the value of the objective function by increasing the accuracy (AC) and decreasing the difference between the sensitivity (SE) and specificity (SP) at the same time to avoid the decision hyperplane to bias toward the majority class, whereas in (2), only AUC is maximized.

In order to verify the feasibility of the proposed method, a pilot study which used the CoIL challenge dataset [40] and WDBC dataset [54] was conducted and compared with the results obtained in previous studies [40] and one-sided selection [43], respectively. Table 1 shows the predictive performance of the CDSSs trained by using training datasets obtained by random undersampling. As shown in the table, the predictive performances of the 10 training subsets are very similar with an average predictive accuracy, sensitivity, specificity, G-mean, and AUC of 66.17%, 62.85%, 66.38%, 0.6439, and 0.7069, respectively.

Table 2 compares the predictive performance among CDSS models designed using different methods based on the CoIL challenge dataset. As shown in the table, unlike biases exhibited in models designed with accuracy-driven algorithms, the method proposed in this study presents higher AUC with more unbiased sensitivity and specificity than the methods proposed in [40], that is, SVM+100% oversampling, MLP+SMOTE, hybrid SVM-MLP+100% oversampling, LR+SMOTE, and hybrid SVM-LR+100% oversampling, as well as decision tree (J48)+cluster-based kNN undersampling and one-sided selection. Notice that the one-sided selection methods achieve 0% sensitivity and 100% specificity in the testing phase for its inability of obtaining a balanced training set after undersampling

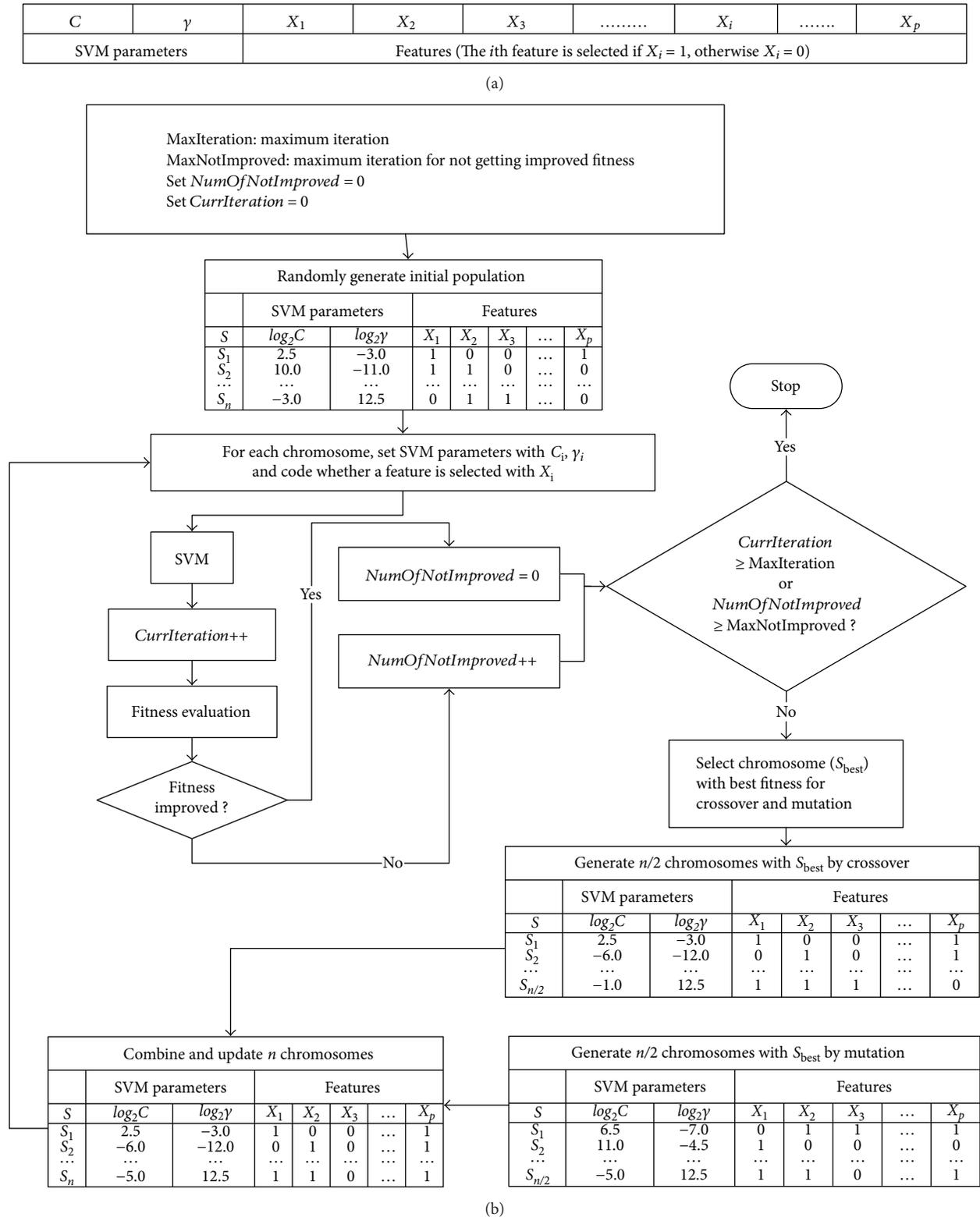


FIGURE 1: The wrapper method combining genetic algorithm and SVM: (a) a chromosome example and (b) the flowchart of the adopted wrapper method with genetic algorithm used for selecting features and adjusting SVM parameters as well as SVM for classifying different classes and calculating fitness values based on the objective functions.

(5474 majority samples and 348 minority samples before and 5236 majority samples and 348 minority samples after one-sided selection).

Table 3 compares the predictive performance among the CDSS models designed based on our proposed methods and the one-sided selection methods using the WDBC dataset. As

TABLE 1: Predictive performance by randomly undersampling the imbalanced training dataset.

Training subset	AC (%)	SE (%)	SP (%)	G-mean	AUC
1	66.32	63.02	66.53	0.6484	0.704
2	65.67	63.86	65.78	0.6481	0.698
3	67.75	60.92	68.18	0.6445	0.704
4	65.07	64.28	65.12	0.6470	0.700
5	65.42	62.18	65.62	0.6388	0.705
6	65.78	62.18	66.10	0.6411	0.713
7	67.75	60.92	68.18	0.6445	0.710
8	65.57	62.18	65.78	0.6396	0.710
9	67.29	65.54	67.29	0.6641	0.716
10	65.12	63.44	65.23	0.6433	0.709
Average	66.17	62.85	66.38	0.6439	0.707

noted in the table, the predictive performance of the models designed with one-sided selection methods demonstrates similar performance with models built using our proposed methods. Compared with the one-sided selection methods, the results indicate that our proposed methods are more robust with the predictive performance less influenced by different datasets adopted for designing CDSSs.

### 3. Results and Discussions

**3.1. Statistical Analysis.** Comparisons of demographic characteristics, comorbid respiratory diseases, and other comorbidities between the patients with and without fracture are shown in Table 4. As shown in the table, age and gender distributions of the patients in fracture cohort and comparison cohort were significantly different ( $p < 0.01$ ). The patients with fracture were older than those without fracture ( $p < 0.001$ ). The mean ages of the patients with fracture were  $70.5 \pm 12.5$  years, and those without fracture were  $58.5 \pm 18.1$  years. The prevalence of respiratory diseases and other comorbidities, including COPD (64.7% versus 51.6%), DM (14.9% versus 12.4%), end-stage renal disease (2.8% versus 1.4%), and osteoporosis (7.1% versus 3.6%), was significantly higher in the fracture cohort than in the comparison cohort ( $p < 0.05$ ), while the prevalence of asthma was significantly lower in the fracture cohort than in the comparison cohort ( $p < 0.001$ ). On the other hand, bronchiectasis (5.8% versus 4.7%), liver cirrhosis (1.1% versus 0.8%), and cancer (4.7% versus 4.4%) were not significantly different between the 2 groups (Table 4).

Table 5 exhibits the association between prescribed oral or intravenous corticosteroid and fracture. The cumulative dose ( $2355.86 \pm 4049.43$  versus  $1345.77 \pm 2831.30$ ) and mean exposed daily dose ( $94.04 \pm 314.98$  versus  $67.65 \pm 235.62$ ) of patients with fracture were significantly higher than the non-fracture cohort ( $p < 0.01$ ), while the mean follow-up daily dose ( $59.56 \pm 285.58$  versus  $59.74 \pm 218.39$ ) did not reach significance ( $p > 0.05$ ). Regarding the follow-up ( $999.65 \pm 997.09$  versus  $706.77 \pm 907.35$  days) and exposed ( $154.04 \pm 320.91$  versus  $90.11 \pm 194.80$  days) durations, the patients

with fracture were significantly longer ( $p < 0.001$ ) than those without fracture.

Table 6 shows the association between prescribed inhaled corticosteroid and fracture. As indicated in the table, the cumulative dose ( $230.14 \pm 442.76$  versus  $171.97 \pm 318.13$ ) of patients with fracture was significantly higher than the patients without fracture ( $p < 0.001$ ), while the mean follow-up daily dose ( $0.44 \pm 2.31$  versus  $1.00 \pm 3.91$ ) was significantly lower ( $p < 0.001$ ). On the other hand, the mean exposed daily dose ( $0.51 \pm 0.81$  versus  $0.55 \pm 1.35$ ) was not significantly different between patients with and without fracture ( $p > 0.05$ ). With regard to the follow-up ( $1777.17 \pm 1039.82$  versus  $1370.91 \pm 1097.76$  days) and exposed ( $453.79 \pm 584.65$  versus  $342.35 \pm 503.32$  days) durations of inhaled corticosteroid, the patients with fracture were significantly longer ( $p < 0.001$ ) than those without fracture.

**3.2. Clinical Decision Support System.** Table 7 compares the predictive performance in both training and testing phases of the CDSSs designed using balanced training sets obtained from different undersampling methods. Table 8 shows the optimal SVM parameters and selected features for CDSS design.

**3.3. Discussions.** Although allergic rhinitis and sinusitis are also a type of chronic respiratory diseases, they are not included for analysis in the current study for they are usually treated with intranasal corticosteroid [55] having lower concentration and smaller effect than oral or inhaled corticosteroids [56]. Neither pulmonary hypertension nor occupational lung diseases are not treated with corticosteroid. It was observed that intranasal corticosteroid suppresses children's growth in short-term study [57], but not affecting adult height for children in long-term treatment [58]. Among the elderly people, administration of low-dose inhaled corticosteroids exhibited small but significant excess risk of cataracts but not observed in use of nasal corticosteroids [56]. Patients with asthma who received the ICS were reported to have higher risk of contracting pneumonia or lower respiratory infection [59] and have effect on adult height in children [60]. Similar to asthma, patients with COPD receiving ICS therapy also exhibited increased risk of pneumonia [61] and TB [62].

As shown in Table 4, the majority of patients with fractures in this study were men (54.1%) with mean age of  $70.5 \pm 12.5$ , which is similar to the results obtained from some previous studies, for example, 94% (mean age:  $62.7 \pm 12.4$ ) reported in [63] and 60% (mean age:  $61.2 \pm 9.0$ ) in [64], while contradicting to another study conducted by Pujades-Rodríguez et al. [15] with only 40% (mean age =  $69.3 \pm 10.03$ ) were male in patients with COPD.

**3.3.1. Effects of OCS and ICS on Fracture.** OCS was reported to be associated with an increased dose-response risk of fracture or osteoporosis in patients with chronic respiratory disease [65] and patients with asthma [66]. Long-term use of OCS may also induce other adverse effects; for example, it was reported that morbidity rates for patients with severe asthma comorbid with other diseases, such as type-2

TABLE 2: Comparisons of predictive performance between models designed based on the methods proposed in this study and other studies [40] with the CoIL challenge dataset [54].

CDSS model	AC (%)	SE (%)	SP (%)	G-mean	AUC
SVM + 100% oversampling	50.08	66.39	49.04	0.5706	0.5772
MLP + SMOTE	82.48	34.87	85.49	0.5460	0.6018
Hybrid SVM-MLP + 100% oversampling	52.1	63.87	51.36	0.5727	0.5762
LR + SMOTE	72.4	56.3	73.42	0.6429	0.6486
Hybrid SVM-LR + 100% oversampling	50.18	66.39	49.15	0.5712	0.5777
Decision tree (J48) + cluster-based kNN undersampling	55.68	68.50	54.86	0.6130	0.6020
One-sided selection + OB1	94.05	0	100	0	0.4997
One-sided selection + OB2	94.05	0	100	0	0.5000
GA-SVM + Rand undersampling with OB1	63.22	66.80	62.99	0.6487	0.7071
GA-SVM + Rand undersampling with OB2	62.92	67.64	62.62	0.6508	0.6885
GA-SVM + cluster-based kNN undersampling with OB1	65.72	65.12	65.76	0.6544	0.6997
GA-SVM + cluster-based kNN undersampling with OB2	62.67	65.96	62.46	0.6419	0.6599

TABLE 3: Comparisons of predictive performance between models designed based on the methods proposed in this study and the one-sided selection methods [43] with the WDBC dataset [54].

Group	Accuracy	Sensitivity	Specificity	G-Mean	AUC
GA-SVM + Rand undersampling with OB1	97.64	97.64	97.64	0.9764	0.9915
GA-SVM + Rand undersampling with OB2	97.40	97.64	97.16	0.9740	0.9970
GA-SVM + cluster-based kNN undersampling with OB1	98.11	98.11	98.11	0.9811	0.9945
GA-SVM + cluster-based kNN undersampling with OB2	97.40	96.22	98.58	0.9739	0.9965
One-sided selection with OB1	97.84	97.97	97.64	0.9780	0.9956
One-sided selection with OB2	97.49	96.22	98.26	0.9723	0.9984

TABLE 4: Comparisons of demographic characteristics, comorbid respiratory diseases, and other comorbidities between patients with and without fracture.

	Fracture		<i>p</i> value
	No ( <i>N</i> = 10511)	Yes ( <i>N</i> = 1134)	
<i>Sex</i>			0.001
Men	6211 (59.1%)	614 (54.1%)	
Women	4300 (40.9%)	520 (45.9%)	
<i>Age (mean ± SD, year)</i>	58.5 ± 18.1	70.5 ± 12.5	<0.001
<i>Age (year)</i>			
20–40	1985 (18.9%)	41 (3.6%)	
41–50	1364 (13.0%)	35 (3.1%)	
51–64	2456 (23.4%)	182 (16.0%)	
≥65	4706 (44.8%)	876 (77.2%)	
<i>Chronic respiratory diseases</i>			
Asthma	7172 (68.2%)	687 (60.6%)	<0.001
COPD	5424 (51.6%)	734 (64.7%)	<0.001
Bronchiectasis	489 (4.7%)	66 (5.8%)	0.079
<i>Other comorbidities</i>			
DM	1302 (12.4%)	169 (14.9%)	0.015
Cancer	464 (4.4%)	53 (4.7%)	0.687
Liver cirrhosis	88 (0.8%)	13 (1.1%)	0.286
ESRD	151 (1.4%)	32 (2.8%)	<0.001
Osteoporosis	377 (3.6%)	81 (7.1%)	<0.001

diabetes, dyspeptic disorder, and cataract, prescribed with higher OCS dose were significantly higher than those with mild/moderate asthma. Glucocorticoid highly decreases bone formation by inhibiting cell differentiation, and increasing apoptosis might be the mechanism causing such an effect [67]. Consistent to previous investigations, our study showed that the cumulative doses of OCS and ICS were significantly higher for patients with fracture than those without fracture, indicating the dose-response effects (Tables 5 and 6). Whether ICS use is associated with osteoporosis or fracture is still controversial. Some previous studies supported that ICS use increased the risk of fracture in patients with COPD [13–15] and patients with asthma [68], while others were against the above conclusion in COPD [65, 69] and asthma [70]. Differences in study design, duration of ICS use and cumulative dose, and frequency of systemic OCS prescriptions might be the reason causing such controversy [71].

The effect of corticosteroid use remains controversial. It was reported that corticosteroid use may increase the risk of fracture for patients with COPD [13–15]. However, a recent study reported that COPD was an independent risk factor for osteoporosis and fracture regardless of ICS use [72]. Moreover, cumulative ICS was reported not to increase the risk of fracture hospitalization [73]. Surprisingly, one study even showed that female COPD patients with ICS therapy exhibited dose-response protective effect on osteoporosis [69]. Similar controversy was also found in asthma. Monadi et al. [68] noted that the BMD in both the spinal cord and

TABLE 5: Association between prescribed oral or intravenous corticosteroid and fracture.

	Fracture		<i>p</i> value
	No ( <i>N</i> = 10511)	Yes ( <i>N</i> = 1134)	
<i>Medication (oral or intravenous corticosteroids)</i>	10511 (90.3%)	1134 (9.7%)	
<i>Cumulative dose (mg)</i>			
Mean ± SD	1345.77 ± 2831.30	2355.86 ± 4049.43	<0.001
Median	390	810	
<i>Mean follow-up daily dose (mg/d) (cumulative dose/follow-up duration)</i>			
Mean ± SD	59.74 ± 218.39	59.56 ± 285.58	0.979
Median	4.54	3.275	
<i>Mean exposed daily dose (mg/d) (Cumulative dose/cumulative exposed duration)</i>			
Mean ± SD	67.65 ± 235.62	94.04 ± 314.98	0.001
Median	13.45	15	
<i>Duration (day), median (range)</i>			
Follow-up duration	706.77 ± 907.35	999.65 ± 997.09	<0.001
Median	240	731.0	
Exposed duration	90.11 ± 194.80	154.04 ± 320.91	<0.001
Median	28	38.5	

TABLE 6: Association between prescribed inhaled corticosteroid and fracture.

	Fracture		<i>p</i> value
	No ( <i>N</i> = 10511)	Yes ( <i>N</i> = 1134)	
<i>Medication (inhaled corticosteroids)</i>	10511 (90.3%)	1134 (9.7%)	
<i>Cumulative dose (mg)</i>			
Mean ± SD	171.97 ± 318.13	230.14 ± 442.76	<0.001
Median	45	54	
<i>Mean follow-up daily dose (mg/d) (cumulative dose/follow-up duration)</i>			
Mean ± SD	1.00 ± 3.91	0.44 ± 2.31	<0.001
Median	0.07	0.05	
<i>Mean exposed daily dose (mg/d) (Cumulative dose/cumulative exposed duration)</i>			
Mean ± SD	0.55 ± 1.35	0.51 ± 0.81	0.261
Median	0.42	0.39	
<i>Duration (day), median (range)</i>			
Follow-up duration	1370.91 ± 1097.76	1777.17 ± 1039.82	<0.001
Median	1239.0	1862.0	
Exposed duration	342.35 ± 503.32	453.79 ± 584.65	<0.001
Median	121.0	199.0	

TABLE 7: Comparisons of predictive performance among different sampling methods.

Group	Training phase			Testing phase				
	AC (%)	SE (%)	SP (%)	AC (%)	SE (%)	SP (%)	G-mean	AUC
Random undersampling with OB1	68.57	74.82	62.32	63.16	77.00	62.36	0.6929	0.7590
Random undersampling with OB2	68.92	68.92	68.92	68.30	70.03	68.20	0.6909	0.7495
Clustering-based kNN undersampling with OB1	71.25	75.66	66.84	63.40	76.19	62.20	0.6884	0.7526
Clustering-based kNN undersampling with OB2	71.25	71.25	71.25	67.28	69.84	67.00	0.6840	0.7515
One-sided selection + OB1	77.80	82.59	35.62	94.54	0	100	0	0.7007
One-sided selection + OB2	89.75	1.41	99.79	71.43	41.98	73.13	0.5541	0.6626

TABLE 8: Optimal SVM parameters and selected features for CDSS design.

Undersampling + objective function	Random + OB1	Random + OB2	Clustered + OB1	Clustered + OB2	One sided + OB1	One sided + OB2
$\alpha$	3.3	11.9	22.0	3.6	25.0	16.8
$\gamma$	-5.6	-6.8	-22.0	-2.8	-25.0	-13.5
Sex	x	x	x	x	x	x
Age	x	x	x	x		x
Asthma						
COPD						
Bronchiectasis						
DM						
Cancer						
Liver cirrhosis						
ESRD						
Osteoporosis						
OCS_followup_days	x		x		x	x
OCS_exposed_days	x	x	x	x		x
ICS_followup_days	x	x	x	x		x
ICS_exposed_days	x	x	x	x		x
OCS_dose	x	x	x			
ICS_dose	x		x	x		
OCS_follow_daily_dose	x			x		x
OCS_exposed_daily_dose		x	x			
ICS_follow_daily_dose	x	x	x	x		x
ICS_exposed_daily_dose						

hip of asthma patients less than 50 years old under 6.5 years of treatment decreased significantly compared to the controls. On the other hand, Loke et al. [70] found that long-term use ( $\geq 12$  months) of ICS in patients with asthma was not significantly associated with fracture and bone mineral density. Such controversies might be caused by differences in study design, duration of ICS use and cumulative dose, and frequency of systemic OCS prescriptions [71]. Frequent OCS administration may significantly increase the risk of osteoporosis and fracture. Investigations of ICS administration and osteoporosis or fracture for patients with non-cystic fibrosis bronchiectasis are scant. It was reported that osteoporosis and osteopenia were prevalent in patients with bronchiectasis [74–76], which might be attributed to the intakes of proton pump inhibitor or inhaled corticosteroids although it was also associated with increased age and female gender [76].

**3.3.2. Prevalence of Fracture in Patients with Chronic Respiratory Disease Receiving OCS and ICS Therapy.** Table 4 compares the prevalence of asthma (60.6% versus 68.2%,  $p < 0.001$ ), COPD (64.7% versus 51.6%,  $p < 0.001$ ), and bronchiectasis (5.8% versus 4.7%,  $p = 0.079$ ) for patients with and without fracture receiving OCS and ICS. It was noted that the prevalence of asthma in patients with fracture was significantly lower than those without fracture, while COPD was more prevalent in patients with fracture than those without fracture. With regard to bronchiectasis,

although it was more prevalent in patients with fracture, but not reaching significance, the inconsistency might be caused by a difficulty in differentiating skeletal effects of corticosteroid medications and COPD itself [67] and symptoms of COPD overlapped with asthma [77, 78] or bronchiectasis [79, 80] for some patients. The mortality rate of bronchiectasis was reported to be as high as 20.4% and increased to 55% if comorbid with COPD [81].

**3.3.3. Comorbidity of Patients with Chronic Respiratory Disease Receiving OCS and ICS Therapy.** Diabetes (DM), end-stage renal disease (ESRD), and osteoporosis were shown to be comorbidities for patients with fracture who had been receiving OCS and ICS treatments. As presented in Table 4, the prevalence of DM (14.9% versus 12.4%,  $p < 0.05$ ), ESRD (2.8% versus 1.4%,  $p < 0.001$ ), and osteoporosis (7.1% versus 3.6%,  $p < 0.001$ ) for patients with fracture was significantly higher than those without fracture. Bountiful literatures have shown the associations between diabetes and bone fracture. Forsen et al. [82] reported that women aged 50–74 years with type 2 diabetes for more than 5 years exhibited higher risk of hip fracture. It was also shown that patients with diabetes presenting diabetic retinopathy, advanced cortical cataract, longer acquired duration, and insulin treatment had higher risk of fracture [83]. Patients with ESRD were noted to have 4.4-fold risk of hip fracture than the general population [84]. Stehman-Breen et al. [85] reported that age, female, BMI, and peripheral vascular disease were

associated with independent factors associated with hip fracture for patients with ESRD. The reason might be that renal osteodystrophy, amyloid, aluminum intoxication, and acidosis commonly found in ESRD patients receiving dialysis increased the risk of bone loss, resulting in hip fracture [84].

**3.3.4. Effectiveness of Clinical Decision Support Systems.** Recently, we have conducted a prospective study to verify the effectiveness of a designed CDSS in ventilator weaning [19] and designed a predictive model for predicting erectile dysfunction using the Taiwan National Health Insurance Database [86]. For predicting successful ventilator weaning, a predictive sensitivity of 87.7% has been achieved by using CDSS, which is significantly higher ( $p < 0.01$ ) than the weaning determined by physicians (sensitivity: 61.4%). Furthermore, the days using mechanical ventilator for the study group ( $38.41 \pm 3.35$ ) are significantly ( $p < 0.001$ ) shorter than the control group ( $43.69 \pm 14.89$ ), with a decrease of 5.2 days in average, resulting in a saving of healthcare cost of NT\$45,000 (US\$1500) per patient in the current Taiwanese National Health Insurance setting [19]. The CDSS was demonstrated to be effective in identifying the earliest time of ventilator weaning for patients to resume and sustain spontaneous breathing, thereby avoiding unnecessary prolonged ventilator use and decreasing healthcare cost.

The CDSS for predicting ventilator weaning was designed based on the datasets collected in respiratory care centers which were more balanced; in contrast, the dataset adopted in this study was retrieved from NHIRD and was highly imbalanced. As shown in Table 2, although the CDSSs designed with our proposed method outperform the other methods [40], the performance is only fair with the accuracy, sensitivity, and specificity all lower than 70% and AUC smaller or a little greater than 0.7. As indicated in Table 7, CDSSs designed with integrated GA and SVM using different undersampling methods for obtaining balanced training sets and applying objective functions for tuning SVM parameters and selecting salient feature exhibit similar and satisfactory predictive performance with AUCs approximate to 0.75.

Although reaching statistically significant difference ( $p < 0.05$ ) between fracture and nonfracture groups (Table 4), the variables regarding respiratory diseases (asthma and COPD) and other comorbidities (DM, ESRD, and osteoporosis) were not selected for designing the CDSSs. Moreover, as shown in Table 8, OCS and ICS variables, such as ICS\_followup\_days and ICS\_exposed\_days, which are statistically dependent with each other were selected. As argued in [53], filter methods like statistical analysis,  $F$  score, and entropy were not good at selecting salient features for CDSS design. Additionally, variables which are highly correlated can be used together to enforce the predictive performance of the CDSS [53].

Patients who had been prescribed with ICS might also be prescribed with oral corticosteroid (OCS) because of intermittent exacerbations. OCS administrated for treating acute exacerbation may have stronger effect than ICS on osteoporosis and fracture. Moreover, previous studies mainly focused on the effects of ICS on osteoporosis or fracture for

patients with asthma and COPD, respectively. The effect of ICS on osteoporosis or fracture for patients with asthma-COPD and bronchiectasis-COPD overlap syndromes should also be considered. Future works will account for the effect of accumulated OCS and ICS dose, not just OCS or ICS dose, on osteoporosis or fracture. Patients with asthma-COPD overlap syndrome and bronchiectasis-COPD overlap syndromes will be separately considered for investigating the effect of OCS and ICS on fracture.

## 4. Conclusions

Based on the analytical results, it was concluded that long-term use of inhaled corticosteroids may induce osteoporosis and exhibit higher incidence of hip or vertebrate fractures. The designed CDSSs exhibited satisfactory performance in the prediction of fractures for patients who were prescribed with corticosteroids. We suggested that the accumulated dose of ICS and OCS therapies should be continuously monitored to prevent from exceeding the maximum dosage, especially for patients with older age and women after menopause.

## Conflicts of Interest

The authors declare that there is no competing interest.

## Authors' Contributions

Yung-Fu Chen and Chih-Sheng Lin contributed equally to this work.

## Acknowledgments

This work was supported in part by the Fulbright Foundation for scholarly exchange (Program no. G-1-00005) and Ministry of Science and Technology, Taiwan (Grant nos. MOST 105-2410-H-166-006, MOST 106-2410-H-166-004, and MOST 107-2918-I-166-001) for Yung-Fu Chen.

## References

- [1] World Health Organization, "Chronic Respiratory Diseases," August 2017, <http://www.who.int/respiratory/en/>.
- [2] World Health Organization, "Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach," World Health Organization, Geneva, Switzerland, 2013.
- [3] T. Ferkol and D. Schraufnagel, "The global burden of respiratory disease," *Annals of the American Thoracic Society*, vol. 11, no. 3, pp. 404–406, 2014.
- [4] R. Lozano, M. Naghavi, K. Foreman et al., "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010," *The Lancet*, vol. 380, no. 9859, pp. 2095–2128, 2012.
- [5] World Health Organization, "Chronic respiratory diseases, asthma," August 2017, <http://www.who.int/respiratory/asthma/en/>.

- [6] D. Weycker, J. Edelsberg, G. Oster, and G. Tino, "Prevalence and economic burden of bronchiectasis," *Clinical Pulmonary Medicine*, vol. 12, no. 4, pp. 205–209, 2005.
- [7] A. E. Seitz, K. N. Olivier, J. Adjemian, S. M. Holland, and D. R. Prevots, "Trends in bronchiectasis among medicare beneficiaries in the United States, 2000 to 2007," *Chest*, vol. 142, no. 2, pp. 432–439, 2012.
- [8] S. P. Bhatt and M. T. Dransfield, "Chronic obstructive pulmonary disease and cardiovascular disease," *Translational Research*, vol. 162, no. 4, pp. 237–251, 2013.
- [9] N. A. Negewo, P. G. Gibson, and V. M. McDonald, "COPD and its comorbidities: impact, measurement and mechanisms," *Respirology*, vol. 20, no. 8, pp. 1160–1171, 2015.
- [10] World Health Organization, "Chronic respiratory diseases, chronic obstructive pulmonary disease (COPD)," August 2017, <http://www.who.int/respiratory/copd/en/>.
- [11] R. J. Boyton, "Bronchiectasis," *Medicine*, vol. 40, no. 5, pp. 267–272, 2012.
- [12] Global Initiative for Asthma, "Global strategy for asthma management and prevention," 2016, <http://ginasthma.org/wp-content/uploads/2016/04/wms-GINA-2016-main-report-final.pdf>.
- [13] S.-W. Huang, W.-T. Wang, L.-C. Chou, H.-C. Chen, T.-H. Liou, and H.-W. Lin, "Chronic obstructive pulmonary disease increases the risk of hip fracture: a nationwide population-based cohort study," *Scientific Reports*, vol. 6, no. 1, article 23360, 2016.
- [14] Y. K. Loke, R. Cavallazzi, and S. Singh, "Risk of fractures with inhaled corticosteroids in COPD: systematic review and meta-analysis of randomised controlled trials and observational studies," *Thorax*, vol. 66, no. 8, pp. 699–708, 2011.
- [15] M. Pujades-Rodriguez, C. J. P. Smith, and R. B. Hubbard, "Inhaled corticosteroids and the risk of fracture in chronic obstructive pulmonary disease," *QJM*, vol. 100, no. 8, pp. 509–517, 2007.
- [16] A. X. Garg, N. K. J. Adhikari, H. McDonald et al., "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review," *JAMA*, vol. 293, no. 10, pp. 1223–1238, 2005.
- [17] T. Porat, O. Kostopoulou, A. Woolley, and B. C. Delaney, "Eliciting user decision requirements for designing computerized diagnostic support for family physicians," *Journal of Cognitive Engineering and Decision Making*, vol. 10, no. 1, pp. 57–73, 2016.
- [18] S. Horng, D. A. Sontag, Y. Halpern, Y. Jernite, N. I. Shapiro, and L. A. Nathanson, "Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning," *PLoS One*, vol. 12, no. 4, article e0174708, 2017.
- [19] J.-C. Hsu, Y.-F. Chen, W.-S. Chung, T.-H. Tan, T. Chen, and J. Y. Chiang, "Clinical verification of a clinical decision support system for ventilator weaning," *BioMedical Engineering Online*, vol. 12, article S4, Supplement 1, 2013.
- [20] G. Luo, F. L. Nkoy, P. H. Gesteland, T. S. Glasgow, and B. L. Stone, "A systematic review of predictive modeling for bronchiolitis," *International Journal of Medical Informatics*, vol. 83, no. 10, pp. 691–714, 2014.
- [21] K. Dunn Lopez, S. M. Gephart, R. Raszewski, V. Sousa, L. E. Shehorn, and J. Abraham, "Integrative review of clinical decision support for registered nurses in acute care settings," *Journal of the American Medical Informatics Association*, vol. 24, no. 2, pp. 441–450, 2017.
- [22] A.-M. J. Scheepers-Hoeks, R. J. Grouls, C. Neef, E. W. Ackerman, and E. H. Korsten, "Physicians' responses to clinical decision support on an intensive care unit—comparison of four different alerting methods," *Artificial Intelligence in Medicine*, vol. 59, no. 1, pp. 33–38, 2013.
- [23] T. Oluoch, A. Katana, D. Kwaro et al., "Effect of a clinical decision support system on early action on immunological treatment failure in patients with HIV in Kenya: a cluster randomised controlled trial," *The Lancet HIV*, vol. 3, no. 2, pp. e76–e84, 2016.
- [24] A. K. Otto, A. A. Dyer, C. M. Warren, M. Walkner, B. M. Smith, and R. S. Gupta, "The development of a clinical decision support system for the management of pediatric food allergy," *Clinical Pediatrics*, vol. 56, no. 6, pp. 571–578, 2017.
- [25] E. Ammenwerth, P. Schnell-Inderst, C. Machan, and U. Siebert, "The effect of electronic prescribing on medication errors and adverse drug events: a systematic review," *Journal of the American Medical Informatics Association*, vol. 15, no. 5, pp. 585–600, 2008.
- [26] F. Baypinar, H. J. Kingma, R. T. M. van der Hoeven, and M. L. Becker, "Physicians' compliance with a clinical decision support system alerting during the prescribing process," *Journal of Medical Systems*, vol. 41, no. 6, p. 96, 2017.
- [27] C. C. Chang, C. C. Liao, and T. L. Chen, "Perioperative medicine and Taiwan National Health Insurance Research Database," *Acta Anaesthesiologica Taiwanica*, vol. 54, no. 3, pp. 93–96, 2016.
- [28] C.-L. Cheng, Y.-H. Y. Kao, S.-J. Lin, C.-H. Lee, and M. L. Lai, "Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan," *Pharmacoepidemiology & Drug Safety*, vol. 20, no. 3, pp. 236–242, 2011.
- [29] W.-S. Chung, Y.-F. Chen, J.-C. Hsu, W.-T. Yang, S.-C. Chen, and J. Y. Chiang, "Inhaled corticosteroids and the increased risk of pulmonary tuberculosis: a population-based case-control study," *International Journal of Clinical Practice*, vol. 68, no. 10, pp. 1193–1199, 2014.
- [30] H.-H. Lin, F.-M. Ho, Y.-F. Chen, C.-M. Tseng, C.-C. Ho, and W.-S. Chung, "Increased risk of erectile dysfunction among patients with sleep disorders: a nationwide population-based cohort study," *International Journal of Clinical Practice*, vol. 69, no. 8, pp. 846–852, 2015.
- [31] Y. F. Chen, H. H. Lin, C. C. Lu et al., "Gout and a subsequent increased risk of erectile dysfunction in men aged 64 and under: a nationwide cohort study in Taiwan," *The Journal of Rheumatology*, vol. 42, no. 10, pp. 1898–1905, 2015.
- [32] Y.-F. Chen, H.-H. Lin, C.-S. Lin, B. Turbat, K.-A. Wang, and W.-S. Chung, "Bronchiectasis and increased risk of ischemic stroke: a nationwide population-based cohort study," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. Volume 12, pp. 1375–1383, 2017.
- [33] M. Kubat, R. C. Holte, and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," *Machine Learning*, vol. 30, no. 2-3, pp. 195–215, 1998.
- [34] F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2002.
- [35] J. Perols, "Financial statement fraud detection: an analysis of statistical and machine learning algorithms," *AUDITING: A Journal of Practice & Theory*, vol. 30, no. 2, pp. 19–50, 2011.

- [36] Y. J. Kim, B. Baik, and S. Cho, "Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning," *Expert Systems with Applications*, vol. 62, pp. 32–43, 2016.
- [37] B. Zhu, Y. Niu, J. Xiao, and B. Baesens, "A new transferred feature selection algorithm for customer identification," *Neural Computing and Applications*, vol. 28, no. 9, pp. 2593–2603, 2017.
- [38] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89–109, 2001.
- [39] P. H. Abreu, M. S. Santos, M. H. Abreu, B. Andrade, and D. C. Silva, "Predicting breast cancer recurrence using machine learning techniques: a systematic review," *ACM Computing Surveys*, vol. 49, no. 3, pp. 1–40, 2016.
- [40] M. A. H. Farquod and I. Bose, "Preprocessing unbalanced data using support vector machine," *Decision Support Systems*, vol. 53, no. 1, pp. 226–233, 2012.
- [41] J. Hu, H. Yang, M. R. Lyu, I. King, and A. M. C. So, "Online nonlinear AUC maximization for imbalanced data sets," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–14, 2017.
- [42] S. Piri, D. Delen, and T. Liu, "A synthetic informative minority over-sampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets," *Decision Support Systems*, vol. 106, pp. 15–29, 2018.
- [43] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: one-sided selection," *In ICML*, vol. 97, pp. 179–186, 1997.
- [44] X. Y. Liu, J. Wu, and Z. H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, 2009.
- [45] D. Devi, S. Biswas, and B. Purkayastha, "Redundancy-driven modified Tomek-link based undersampling: a solution to class imbalance," *Pattern Recognition Letters*, vol. 93, pp. 3–12, 2017.
- [46] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [47] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in *Advances in Intelligent Computing*, pp. 878–887, Springer, Berlin, Heidelberg, 2005.
- [48] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-level-SMOTE: safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," in *Advances in Knowledge Discovery and Data Mining*, T. Theeramunkong, B. Kijssirikul, N. Cercone, and T.-B. Ho, Eds., pp. 475–482, Springer, Berlin, Heidelberg, 2009.
- [49] D. A. Cieslak, N. V. Chawla, and A. Striegel, "Combating imbalance in network intrusion datasets," in *2006 IEEE International Conference on Granular Computing*, pp. 732–737, Atlanta, GA, USA, 2006.
- [50] C. Jian, J. Gao, and Y. Ao, "A new sampling method for classifying imbalanced data based on support vector machine ensemble," *Neurocomputing*, vol. 193, pp. 115–122, 2016.
- [51] K. A. Wang, Y. F. Chen, L. O. A. Rahman, H. H. Lin, and W. S. Chung, "Using Taiwan NHIRD for studying the effect of inhaled corticosteroids on fracture," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2667–2671, Banff, AB, Canada, October 2017.
- [52] W.-C. Lin, C.-F. Tsai, Y.-H. Hu, and J.-S. Jhang, "Clustering-based undersampling in class-imbalanced data," *Information Sciences*, vol. 409–410, pp. 17–26, 2017.
- [53] Y.-F. Chen, P.-C. Huang, K.-C. Lin et al., "Semi-automatic segmentation and classification of pap smear cells," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 1, pp. 94–108, 2014.
- [54] "Breast Cancer Wisconsin (Diagnostic) Data Set," <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>.
- [55] C. Cingi, P. Gevaert, R. Mösges et al., "Multi-morbidities of allergic rhinitis in adults: European Academy of Allergy and Clinical Immunology task force report," *Clinical and Translational Allergy*, vol. 7, no. 1, p. 17, 2017.
- [56] P. Ernst, M. Baltzan, J. Deschênes, and S. Suissa, "Low-dose inhaled and nasal corticosteroid use and the risk of cataracts," *European Respiratory Journal*, vol. 27, no. 6, pp. 1168–1174, 2006.
- [57] D. P. Skoner, G. S. Rachelefsky, E. O. Meltzer et al., "Detection of growth suppression in children during treatment with intranasal beclomethasone dipropionate," *Pediatrics*, vol. 105, no. 2, article e23, 2000.
- [58] L. Agertoft and S. Pedersen, "Effect of long-term treatment with inhaled budesonide on adult height in children with asthma," *The New England Journal of Medicine*, vol. 343, no. 15, pp. 1064–1069, 2000.
- [59] T. McKeever, T. W. Harrison, R. Hubbard, and D. Shaw, "Inhaled corticosteroids and the risk of pneumonia in people with asthma: a case-control study," *Chest*, vol. 144, no. 6, pp. 1788–1794, 2013.
- [60] H. W. Kelly, A. L. Sternberg, R. Lescher et al., "Effect of inhaled glucocorticoids in childhood on adult height," *The New England Journal of Medicine*, vol. 367, no. 10, pp. 904–912, 2012.
- [61] K. M. Kew and A. Seniukovich, "Inhaled steroids and risk of pneumonia for chronic obstructive pulmonary disease," *Cochrane Database of Systematic Reviews*, no. 3, article CD010115, 2014.
- [62] Y.-H. Dong, C.-H. Chang, F.-L. L. Wu et al., "Use of inhaled corticosteroids in patients with COPD and the risk of TB and influenza: a systematic review and meta-analysis of randomized controlled trials," *Chest*, vol. 145, no. 6, pp. 1286–1297, 2014.
- [63] T. A. Lee and K. B. Weiss, "Fracture risk associated with inhaled corticosteroid use in chronic obstructive pulmonary disease," *American Journal of Respiratory and Critical Care Medicine*, vol. 169, no. 7, pp. 855–859, 2004.
- [64] J. D. Jaramillo, C. Wilson, D. S. Stinson et al., "Reduced bone density and vertebral fractures in smokers. Men and COPD patients at increased risk," *Annals of the American Thoracic Society*, vol. 12, no. 5, pp. 648–656, 2015.
- [65] P. Vestergaard, L. Rejnmark, and L. Mosekilde, "Fracture risk in patients with chronic lung diseases treated with bronchodilator drugs and inhaled and oral corticosteroids," *Chest*, vol. 132, no. 5, pp. 1599–1607, 2007.
- [66] P. Li, L. Ghazala, E. Wright, J. Beach, D. Morrish, and D. Vethanayagam, "Prevalence of osteopenia and osteoporosis in patients with moderate to severe asthma in western Canada," *Clinical and Investigative Medicine*, vol. 38, no. 2, pp. E23–E30, 2015.

- [67] S. A. Sutter and E. M. Stein, "The skeletal effects of inhaled glucocorticoids," *Current Osteoporosis Reports*, vol. 14, no. 3, pp. 106–113, 2016.
- [68] M. Monadi, Y. Javadian, M. Cheraghi, B. Heidari, and M. Amiri, "Impact of treatment with inhaled corticosteroids on bone mineral density of patients with asthma: related with age," *Osteoporosis International*, vol. 26, no. 7, pp. 2013–2018, 2015.
- [69] S.-F. Liu, H.-C. Kuo, G.-H. Liu et al., "Inhaled corticosteroids can reduce osteoporosis in female patients with COPD," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 11, pp. 1607–1614, 2016.
- [70] Y. K. Loke, D. Gilbert, M. Thavarajah, P. Blanco, and A. M. Wilson, "Bone mineral density and fracture risk with long-term use of inhaled corticosteroids in patients with asthma: systematic review and meta-analysis," *BMJ Open*, vol. 5, no. 11, article e008554, 2015.
- [71] S. A. Aljubran, G. J. Whelan, M. C. Glaum, and R. F. Lockey, "Osteoporosis in the at-risk asthmatic," *Allergy*, vol. 69, no. 11, pp. 1429–1439, 2014.
- [72] S. J. Chen, W. C. Liao, K. H. Huang et al., "Chronic obstructive pulmonary disease and allied conditions is a strong independent risk factor for osteoporosis and pathologic fractures: a population-based cohort study," *QJM*, vol. 108, no. 8, pp. 633–640, 2015.
- [73] R. W. Flynn, T. M. MacDonald, A. Hapca, I. S. MacKenzie, and S. Schembri, "Quantifying the real life risk profile of inhaled corticosteroids in COPD by record linkage analysis," *Respiratory Research*, vol. 15, no. 1, p. 141, 2014.
- [74] R. K. Dave, A. C. Arroliga, and M. Prince-Fiocco, "Osteoporosis and osteopenia in patients with non-cystic fibrosis (non-CF) bronchiectasis," *Chest*, vol. 136, no. 4, p. 9S, 2009.
- [75] S. S. Sunny, A. Burns, H. Jary, T. Aspray, and A. De Soyza, "P90 osteoporosis in non-cystic fibrosis bronchiectasis (NCFBr) adults," *Thorax*, vol. 67, Supplement 2, pp. A102.3–A1A103, 2012.
- [76] N. Diehl and M. M. Johnson, "Prevalence of osteopenia and osteoporosis in patients with noncystic fibrosis bronchiectasis," *Southern Medical Journal*, vol. 109, no. 12, pp. 779–783, 2016.
- [77] M. J. Abramson, J. L. Perret, S. C. Dharmage, V. M. McDonald, and C. F. McDonald, "Distinguishing adult-onset asthma from COPD: a review and a new approach," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 9, pp. 945–962, 2014.
- [78] D. S. Postma and K. F. Rabe, "The asthma–COPD overlap syndrome," *The New England Journal of Medicine*, vol. 373, no. 13, pp. 1241–1249, 2015.
- [79] J. R. Hurst, J. S. Elborn, A. de Soyza, and BRONCH-UK Consortium, "COPD–bronchiectasis overlap syndrome," *European Respiratory Journal*, vol. 45, no. 2, pp. 310–313, 2015.
- [80] M. Á. Martínez-García, M. J. Selma-Ferrer, C. Navarro-Soriano, and A. Muñoz-Reina, "Bronchiectasis phenotype in COPD patients," *Clinical Pulmonary Medicine*, vol. 22, no. 3, pp. 123–127, 2015.
- [81] P. C. Goeminne, T. S. Nawrot, D. Ruttens, S. Seys, and L. J. Dupont, "Mortality in non-cystic fibrosis bronchiectasis: a prospective cohort analysis," *Respiratory Medicine*, vol. 108, no. 2, pp. 287–296, 2014.
- [82] L. Forsen, H. E. Meyer, K. Midthjell, and T. H. Edna, "Diabetes mellitus and the incidence of hip fracture: results from the Nord-Trøndelag Health Survey," *Diabetologia*, vol. 42, no. 8, pp. 920–925, 1999.
- [83] R. Q. Ivers, R. G. Cumming, P. Mitchell, A. J. Peduto, and Blue Mountains Eye Study, "Diabetes and risk of fracture: the Blue Mountains Eye Study," *Diabetes Care*, vol. 24, no. 7, pp. 1198–1203, 2001.
- [84] A. M. Alem, D. J. Sherrard, D. L. Gillen et al., "Increased risk of hip fracture among patients with end-stage renal disease," *Kidney International*, vol. 58, no. 1, pp. 396–399, 2000.
- [85] C. O. Stehman-Breen, D. J. Sherrard, A. M. Alem et al., "Risk factors for hip fracture among patients with end-stage renal disease," *Kidney International*, vol. 58, no. 5, pp. 2200–2205, 2000.
- [86] H. H. Lin, P. C. Chan, M. H. Lee et al., "Decision support systems for predicting erectile dysfunction," in *Proceedings of the International Conference on Computer Networks and Communication Technology (CNCT 2016)*, pp. 606–612, Xiamen, China, 2016.

## Research Article

# Automatic Semantic Segmentation of Brain Gliomas from MRI Images Using a Deep Cascaded Neural Network

Shaoguo Cui <sup>1,2</sup>, Lei Mao,<sup>1</sup> Jingfeng Jiang <sup>2,3</sup>, Chang Liu,<sup>1</sup> and Shuyu Xiong<sup>1</sup>

<sup>1</sup>College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China

<sup>2</sup>Medical Physics Department, University of Wisconsin, Madison, WI 53705, USA

<sup>3</sup>Biomedical Engineering Department, Michigan Technological University, Houghton, MI 49931, USA

Correspondence should be addressed to Shaoguo Cui; [cuishaoguo2002@163.com](mailto:cuishaoguo2002@163.com)

Received 30 November 2017; Accepted 11 February 2018; Published 19 March 2018

Academic Editor: Weide Chang

Copyright © 2018 Shaoguo Cui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Brain tumors can appear anywhere in the brain and have vastly different sizes and morphology. Additionally, these tumors are often diffused and poorly contrasted. Consequently, the segmentation of brain tumor and intratumor subregions using magnetic resonance imaging (MRI) data with minimal human interventions remains a challenging task. In this paper, we present a novel fully automatic segmentation method from MRI data containing *in vivo* brain gliomas. This approach can not only localize the entire tumor region but can also accurately segment the intratumor structure. The proposed work was based on a cascaded deep learning convolutional neural network consisting of two subnetworks: (1) a tumor localization network (TLN) and (2) an intratumor classification network (ITCN). The TLN, a fully convolutional network (FCN) in conjunction with the transfer learning technology, was used to first process MRI data. The goal of the first subnetwork was to define the tumor region from an MRI slice. Then, the ITCN was used to label the defined tumor region into multiple subregions. Particularly, ITCN exploited a convolutional neural network (CNN) with deeper architecture and smaller kernel. The proposed approach was validated on multimodal brain tumor segmentation (BRATS 2015) datasets, which contain 220 high-grade glioma (HGG) and 54 low-grade glioma (LGG) cases. Dice similarity coefficient (DSC), positive predictive value (PPV), and sensitivity were used as evaluation metrics. Our experimental results indicated that our method could obtain the promising segmentation results and had a faster segmentation speed. More specifically, the proposed method obtained comparable and overall better DSC values (0.89, 0.77, and 0.80) on the combined (HGG + LGG) testing set, as compared to other methods reported in the literature. Additionally, the proposed approach was able to complete a segmentation task at a rate of 1.54 seconds per slice.

## 1. Introduction

Although brain cancers are less prevalent, they are very lethal. Among them, gliomas are the most common brain tumors. They can be graded into low-grade gliomas (LGG) and high-grade gliomas (HGG), with the latter being more aggressive and infiltrative than the former [1]. A glioma is highly invasive because it tends to aggressively grow and could quickly invade the central nervous system (CNS). According to US National Cancer Institute, approximately 18,000 Americans are diagnosed with a glioma every year; many of them die within 14 months [2]. In clinical practice, medical imaging, mainly computed tomography (CT) and magnetic resonance imaging (MRI), has been used to

determine (1) the presence of a tumor, (2) the inclusion of peritumoral edema, and (3) the spread into other locations such as the CNS [3].

Compared to CT, MRI or contrast-enhanced MRI becomes the imaging modality of choice for diagnosis and treatment planning in the brain because of its sensitivity and superior image contrast in soft tissues. However, the multiplicity and complexity of the brain tumors under MRI often make tumor recognition and segmentation difficult for radiologists and other clinicians [4]. Consequently, automatic segmentation of heterogeneous tumors can greatly impact the clinical medicine by freeing physicians from the burden of the manual depiction of tumors. Furthermore, if computer algorithms can provide robust and quantitative

measurements of tumor depiction, these automated measurements will greatly aid in the clinical management of brain tumors.

In the past few decades, significant research efforts in the computer vision and image processing community have been devoted to developing computer-aided systems that can be used for automated tumor characterization/classification [5–21]. Although some systems were tested and showed good performance, the fully automatic detection and subsequent diagnosis of brain tumors have not been massively used in the clinical settings, thereby indicating that some major developments are still needed [21].

Based on MRI data, our primary goal of this paper was to propose a new fast and accurate computer system that could first localize complete tumor region and then segment the more detailed intratumor structure. Our computer system contained two major steps. First, by leveraging an FCN [22], a tumor location map was first obtained. In the second step, a deep learning ensemble of the CNN was used to classify the tumor region into four subregions: (1) necrosis, (2) edema, (3) nonenhancing tumor, and (4) enhancing tumor. In this study, the performance of the proposed algorithm was assessed in a public database containing 274 cases of *in vivo* gliomas.

The paper is structured as follows: Section 2 presents the related works in the automated brain cancer segmentation. Particularly, attention was given to computer systems based on machine learning. The proposed two-step (cascaded) neural network is described in Section 3. The emphases are on the design methodology and training methods for the performance assessment. In Section 4, results of our numerical experiments are summarized followed by some closing remarks in Section 5.

## 2. Relevant Work and Our Contributions

In recent years, many methods have been proposed to automatically segment brain tumors based on MRI data. These methods can be largely divided into two categories: (1) hand-crafted feature and classifier methods based on traditional machine learning such as support vector machine (SVM) and random forests (RF) [5–13] and (2) fully automatic methods based on deep learning using the CNN [14–21].

Methods in the first category use manually extracted features, and these features are input to classifiers. In other words, once these hand-crafted features are solely determined by human operators, classifiers “weigh” them during the training but cannot modify these features in any way. One significant concern of hand-crafted features stems from the fact that these features could have significant inter- and intrauser variability. A brief summary of these methods can be found in Table 1.

In contrast, methods in the second category can self-learn the feature representations adapted to a specific task from training data. Recently, deep learning neural networks, especially CNNs, are rapidly gaining their popularity in the computer vision community. This trend has certainly been accelerated after the recent record-shattering performance

of the CNN in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [23]. Recent deep learning methods for automatic brain tumor segmentation are summarized below in Table 2.

However, the above-mentioned CNN methods were all based on the patch-wise method in which (medical) images were often divided into patches during the training and testing. The advantage of this method was that it could take advantage of the existing classification model of the natural image and solve the problem of the class label imbalance in MRI images. Despite its popularity, operating on image patches was computationally time-consuming. Recalling, given a typical image size (e.g.,  $256 \times 256$ ), a large number of patches (65535) were required as inputs for prediction. Furthermore, this method was not end-to-end and performed the segmentation task by independently classifying the central pixel of a patch, which will result in some errors and need postprocessing. Thus, the expensive computation and postprocessing become the bottleneck of its real-time clinic application.

Recently, Shelhamer et al. [22] presented a novel FCN for semantic segmentation of natural scene images. This model can be trained in an end-to-end manner (also known as pixel-wise). Their results showed that the FCN outperformed the previous methods for semantic segmentation of a natural scene image in performance and speed. Inspired by the work in [22], we proposed a hybrid approach by constructing a deep cascaded neural network.

Our main contribution of this work is to propose a hybrid cascaded neural network for the purpose of segmentation of brain tumors including segmentation of intratumor subregions, from MRI data. This model consists of one FCN and one CNN. This combination enables us to perform pixel semantic predictions by taking advantage of both a pixel-wise method and a patch-wise method. Formally, in this cascaded neural network, an FCN was first used to localize the tumor region from an MRI slice and then a CNN with deeper architecture and smaller kernels was used to classify brain tumor into multiple subregions. This approach can not only obtain the better segmentation accuracy but can also speed the prediction efficiency.

## 3. Methods

**3.1. Construction of the Deep Cascaded Neural Network.** The starting point of the proposed system is *in vivo* MRI data consisting of four different sequences (FLAIR, T1, T1c, and T2), and the endpoint becomes a characterized tumor (see Figure 1). In the output image, a brain tumor is classified into four different zones: necrosis, edema, nonenhancing tumor, and enhancing tumor.

More specifically, the architecture of the proposed system includes an FCN followed by a CNN which accompanies small convolution kernels (see Figure 1). So the segmentation task based on this cascaded network can be divided into two major steps. In the first step, the pixel-wise FCN was used to quickly localize the tumor by marking the tumor region. Then, the patch-wise CNN was used to further categorize the above-identified tumor region into different subregions

TABLE 1: A summary of brain tumor segmentation methods based on traditional machine learning. Only methods using MRI data were included in this table.

Number	Publication	Database	Summary of method	Performance
1	Corso et al. [5]	20 cases of <i>in vivo</i> brain tumors; T1, T1-C, T2, FLAIR	A hybrid method combining an affinity-based segmentation method with a generative model	0.62–0.69 (Jaccard)
2	Hamamci et al. [6]	Synthetic data from Utah + <i>in vivo</i> data from Harvard	A cellular automata method combining a probability framework	0.72 (DICE complete tumor)
3	Mehmood et al. [7]	BrainWeb data + <i>in vivo</i> brain tumors; T1, T1-weighted, T2, T2-weighted	A novel saliency model for lesion localization and an N-cut graph segmentation model for classification	83%–95% (classification accuracy)
4	Havaei et al. [8]	MICCAI-BRATS 2013 dataset	Hand-crafted features + a support vector machine	0.86 (DICE complete tumor)
5	Usman and Rajpoot [9]	MICCAI-BRATS 2013 dataset	Automated wavelet-based features + a random forest classifier	0.88 (DICE complete tumor)
6	Tustison et al. [10]	MICCAI-BRATS 2013 dataset	Combine a random forest model with a framework of regularized probabilistic segmentation	0.88 (DICE complete tumor)
7	Zikic et al. [11]	40 multichannel MR images, including DTI	Decision forests using context-aware spatial features for automatic segmentation of high-grade gliomas	GT: 0.89 AC: 0.84 NE: 0.70 E: 0.72 (10/30 tests)
8	Pinto et al. [12]	MICCAI-BRATS 2013 dataset	Using appearance- and context-based features to feed an extremely randomized forest	0.83 (DICE complete tumor)
9	Bauer et al. [13]	10 multispectral patient datasets	Combines support vector machine classification with conditional random fields	GT: 0.84 AC: 0.84 NE: 0.70 E: 0.72 (Inpatient regularized)

TABLE 2: A summary of brain tumor segmentation methods based on deep-learning neural networks. Only methods using MRI data were included in this table.

Number	Publication	Database	Summary of method	Performance (DICE)		
				Complete	Core	Enh
1	Urban et al. [14]	MICCAI-BRATS 2013 dataset	3D CNN with 3D convolutional kernels	0.87	0.77	0.73
2	Zikic et al. [15]	MICCAI-BRATS 2013 dataset	Apply a CNN in a sliding-window fashion in the 3D space	0.84	0.74	0.69
3	Davy et al. [16]	MICCAI-BRATS 2013 dataset	A CNN with two pathways of both local and global information	0.85	0.74	0.68
4	Dvorak and Menze [17]	MICCAI-BRATS 2013 dataset	Structured prediction was used together with a CNN	0.83	0.75	0.77
5	Pereira et al. [18]	MICCAI-BRATS 2013 dataset	A CNN with small $3 \times 3$ kernels	0.88	0.83	0.77
6	Havaei et al. [19]	MICCAI-BRATS 2013 dataset	A cascade neural network architecture in which “the output of a basic CNN is treated as an additional source of information for a subsequent CNN”	0.88	0.79	0.73
7	Lyksborg et al. [20]	MICCAI-BRATS 2014 dataset	An ensemble of 2D convolutional neural networks +doing a volumetric segmentation by three steps	0.80	0.64	0.59
8	Kamnitsas et al. [21]	MICCAI-BRATS 2015 dataset	Using 3D CNN, two-scale extracted feature, 3D dense CRF as postprocessing	0.85	0.67	0.63

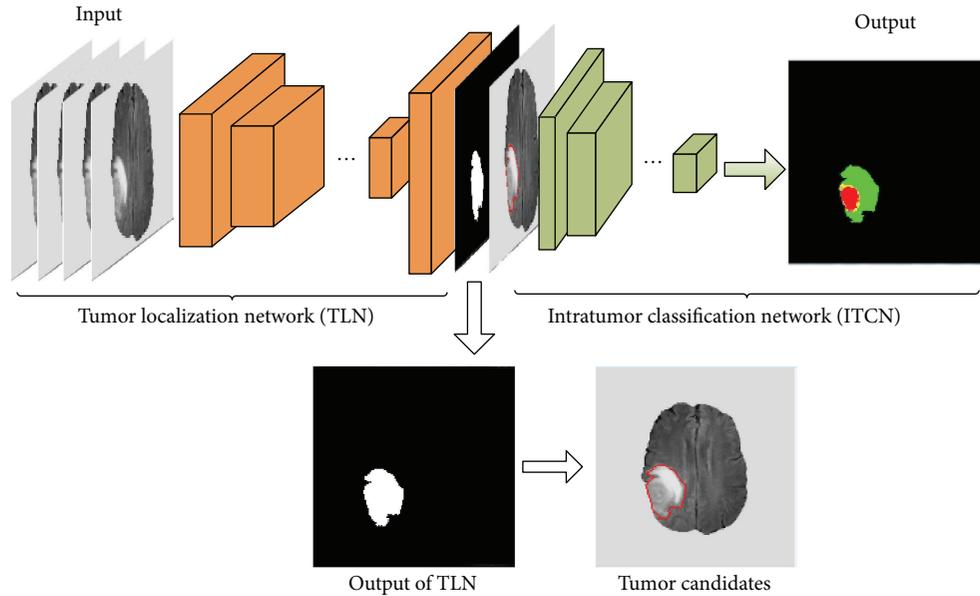


FIGURE 1: An illustrative overview of the proposed deep cascaded convolutional neural network for a fast and accurate tumor segmentation.

representing different pathologies. This system design was motivated and justified as follows. First, the FCN can take a whole image as the input and localization of a complete tumor only requires one-pass of the forward propagation. Thus, it can remarkably improve the segmentation efficiency. Second, this combination of FCN and CNN can alleviate the pixel sample class imbalance problem which is serious in MRI images. Thus, it can capture better segmentation details. Third, the intratumor characterization in the second step will only need to be applied to the tumor regions localized in the first step instead of the entire image, thereby significantly reducing forward computing time. Hereafter, the FCN and the CNN are referred as to tumor localization network (TLN) and intratumor classification network (ITCN), respectively.

**3.1.1. A Description of TLN.** We modified the FCN-8s architecture [22] to model our TLN. The input channels (RGB) in the original FCN-8s were changed to 4 channels in order to account for 4 different MRI modalities. And the 21 output channels in the original FCN-8s were changed to 2, corresponding to either the tumor region or the nontumor region. As shown in Figure 2, after the operations of the convolution and pooling, the feature map became smaller in size (see Table 3). To obtain a higher resolution of the final features, the input images (size  $240 \times 240$ ) were padded to  $438 \times 438$  using zero padding [22]. Additionally, the deconvolution was applied so that the size of output image matched with that of the input image. It is worth noting that multiple convolutional kernels were used in each convolutional layer for a better feature extraction (e.g., edges, curves, and corner).

We observed that a significant amount of low-level feature details such as location and edge could be lost after convolution striding and pooling. However, these lost features were valuable for semantic segmentation. Thus, two skip connections [22] were introduced for two purposes: (1)

mitigating the loss of local image features and (2) combining local information obtained from intermediate layers (i.e., max pooling 4 and max pooling 3, resp.) with the global information in these deep layers (i.e., after 7 convolution layers). All relevant parameters used in the subnet TLN are shown in Table 3 below.

**3.1.2. A Description of ITCN.** The proposed ITCN includes two convolutional layer groups (3 layers each), two max pooling layers, and three fully connected layers. Recall that the TLN yields a binary tumor map for a given MRI image and the ITCN (see Figure 3) further classifies the identified tumor into 4 different subregions. Formally, for each location  $(i, j)$  within the identified tumor map, 4 patches (size of  $33 \times 33$ ) centered on the  $(i, j)$  location were extracted from the original 4 input channels (FLAIR, T1, T1c, and T2) and subsequently used as the input to the ITCN. More details of this ITCN subnet are listed in Table 4.

In the ITCN, as inspired by the work of Simonyan and Zisserman [24], multiple convolutional layers with small kernels ( $3 \times 3$  pixels) were used. An alternative approach would be an architecture with fewer layers and larger kernels. Theoretically, two cascaded convolutional layers with two  $3 \times 3$  kernels have similar effects on the receptive fields, as compared to one convolutional layer with a  $5 \times 5$  kernel. But two cascaded layers with two  $3 \times 3$  kernels result in more complex nonlinearities and fewer weights. Fewer weights lead to a less computing cost and can also alleviate the possibility of overfitting. It is generally understood that, with the increase of the CNN's depth, a CNN can gain higher representation capacity. As shown in Figure 3, in each of the two pooling layers, a  $3 \times 3$  overlapping subwindow with a stride of 2 was applied to the feature maps for reducing feature dimension and integrating higher-level features. The detailed hyperparameters of the ITCN can be found in Table 4 below.

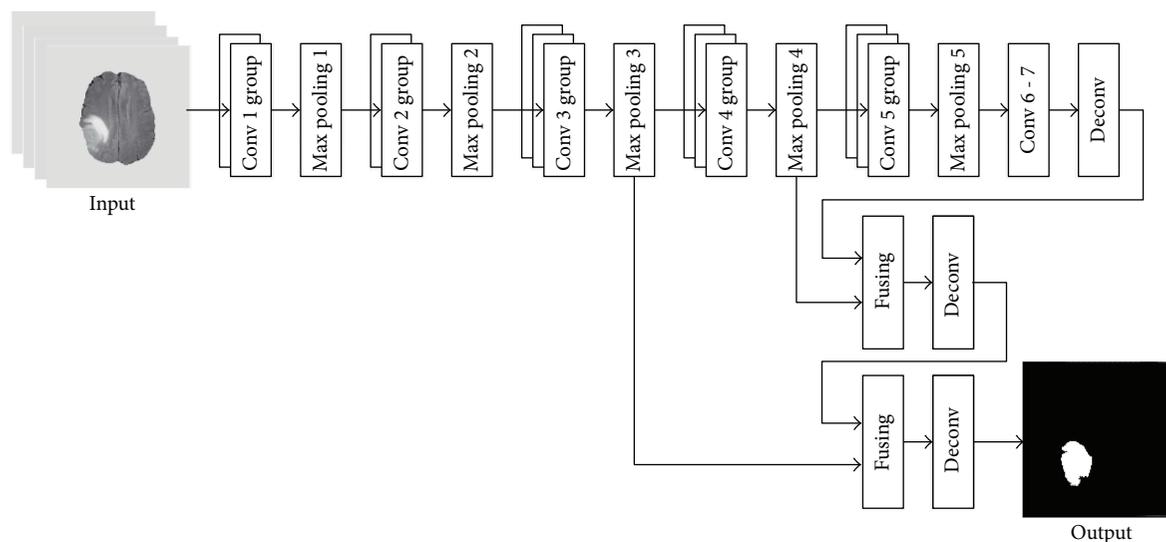


FIGURE 2: An illustration of the architecture of the TLN subnet for pixel-wise prediction.

TABLE 3: Parameters used in the subnet TLN. In each convolutional layer, the feature maps had been padded by 1 prior to the convolution so that all intermediate feature maps do not change their sizes before and after the convolution.

Number	Layer name	Filter size	Stride	Number of Filters	Output
1	Conv 1_1 + ReLU	3*3	1	64	438*438*64
2	Conv 1_2 + ReLU	3*3	1	64	438*438*64
3	Max pooling 1	2*2	2	—	219*219*64
4	Conv 2_1 + ReLU	3*3	1	128	219*219*128
5	Conv 2_2 + ReLU	3*3	1	128	219*219*128
6	Max pooling 2	2*2	2	—	110*110*128
7	Conv 3_1 + ReLU	3*3	1	256	110*110*256
8	Conv 3_2 + ReLU	3*3	1	256	110*110*256
9	Conv 3_3 + ReLU	3*3	1	256	110*110*256
10	Max pooling 3	2*2	2	—	55*55*256
11	Conv 4_1 + ReLU	3*3	1	512	55*55*512
12	Conv 4_2 + ReLU	3*3	1	512	55*55*512
13	Conv 4_3 + ReLU	3*3	1	512	55*55*512
14	Max pooling 4	2*2	2	—	28*28*512
15	Conv 5_1 + ReLU	3*3	1	512	28*28*512
16	Conv 5_2 + ReLU	3*3	1	512	28*28*512
17	Conv 5_3 + ReLU	3*3	1	512	28*28*512
18	Max pooling 5	2*2	2	—	14*14*512
19	Conv 6 + ReLU	7*7	1	4096	8*8*4096
20	Conv 7 + ReLU	1*1	1	4096	8*8*4096

**3.2. Implementation.** All numerical experiments were conducted using a Dell workstation equipped with dual Intel E5-2603 CPUs and a middle-end GPU graphic card (GeForce GTX 1080, NVIDIA, CA, USA). The operation system of the workstation is Ubuntu (version 14.04). The proposed cascaded neural network has been implemented using Python (version 2.7) under the framework of Caffe, an open-source deep learning platform (<http://caffe.berkeleyvision.org/>). Some essential details are discussed below.

**3.2.1. Preprocessing.** As recommended by the literature [25], MRI data were preprocessed before the proposed cascaded neural network was applied. Basically, the N4ITK method was first used to correct the distortion of MRI data, followed by data normalization.

Given an image  $X$ ,  $x(i, j)$  is the intensity corresponding to the  $j$ th column at the  $i$ th row of  $X(i, j = 1, 2, \dots, 240)$ . The data intensity normalization procedure is briefly described below:

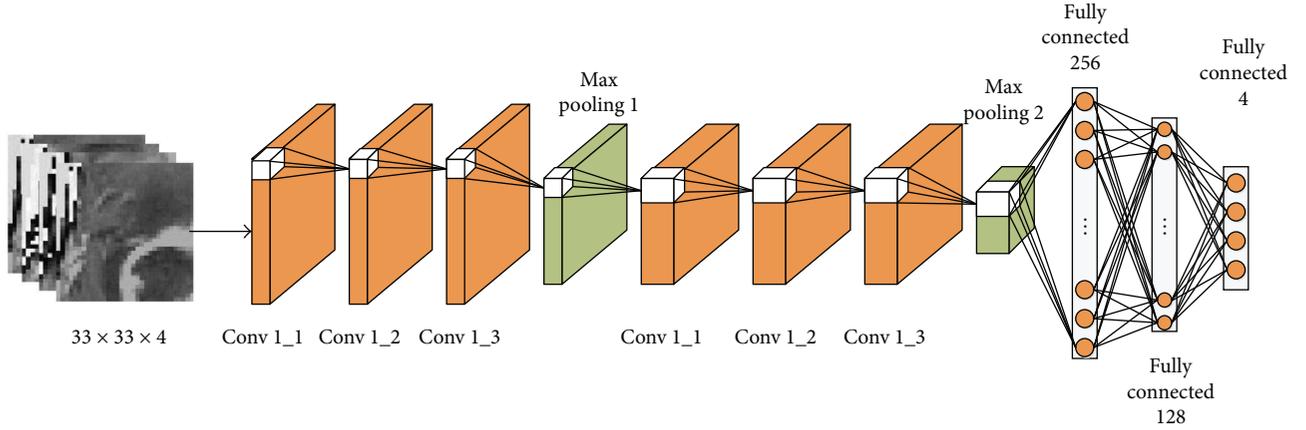


FIGURE 3: An illustration of the second subnet ITCN for the intratumoral classification. The classification was done in a patch-to-patch fashion.

TABLE 4: A list of parameters used in the proposed subnet ITCN. In each convolutional layer, the feature maps had been padded by 1 prior to the convolution so that the convolution do not change the size of the resultant feature map.

Number	Layer name	Filter size	Stride	Number of filters	FC units	Output
1	Conv 1_1 + LReLU	3*3	1	64	—	33*33*64
2	Conv 1_2 + LReLU	3*3	1	64	—	33*33*64
3	Conv 1_3 + LReLU	3*3	1	64	—	33*33*64
4	Max pooling 1	3*3	2	—	—	16*16*64
5	Conv 2_1 + LReLU	3*3	1	128	—	16*16*128
6	Conv 2_2 + LReLU	3*3	1	128	—	16*16*128
7	Conv 2_3 + LReLU	3*3	1	128	—	16*16*128
8	Max pooling 2	3*3	2	—	—	8*8*128
9	FC1 + dropout	—	—	—	8192	256
10	FC2 + dropout	—	—	—	256	128
11	FC3 + softmax	—	—	—	128	4

- (1) Removed the top 1% and bottom 1% from each slice of the MRI data.
- (2) For each slice of MRI data  $X$ , a normalized image  $X'$  was obtained. In the scaled image  $X'$ , each intensity value  $x'(i, j)$  can be obtained as follows:

$$x'(i, j) = \frac{(x(i, j) - \bar{X})}{X_s}, \quad (1)$$

where  $x(i, j)$  is the gray value of pixel  $(i, j)$  prior to the normalization and  $\bar{X}$  and  $X_s$  are the mean and standard deviation of the unscaled image  $X$ , respectively.

The above-mentioned preprocessing method was used to process each modality MRI data including FLAIR, T1, T1c, and T2. Particularly, the FLAIR images were generated using fluid-attenuated inversion recovery protocol and useful in terms of differentiating the brain tumor from its normal background. Figure 4 presents some FLAIR slices before and after using the proposed image intensity normalization.

We randomly selected 3 different cases from the FLAIR dataset. As shown in Figure 4 below, it is easy to find that the above-mentioned data normalization can improve the comparability of different slices.

**3.2.2. Convolution Operation.** Each feature map  $Z$  shown in Figures 1, 2, and 3 was associated with one convolution kernel.  $Z$  was computed as follows:

$$Z = b + \sum_{r=1}^k W_r * X_r, \quad (2)$$

where  $k$  is the number of input channels,  $b$  is a bias term,  $X_r$  is an image from the  $r$ th input channel, and  $W_r$  is the weight associated with the  $r$ th channel. In (2),  $*$  denotes a convolution operator.

**3.2.3. Nonlinear Activation Function.** In our study, the TLN used rectified linear unit (ReLU) function [23] to perform nonlinear transformations. This selection was because ReLU could achieve better results as compared to the classical sigmoid and hyperbolic tangent functions. The use of ReLU

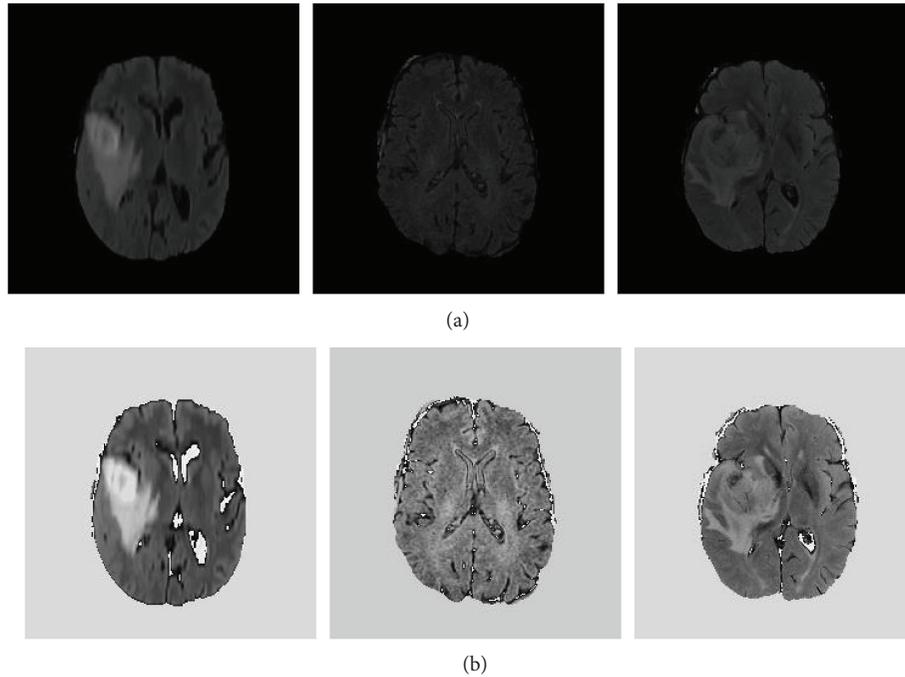


FIGURE 4: Randomly selected examples of FLAIR slices before (a) and after (b) the above-mentioned intensity normalization.

was also able to accelerate the training [26]. Mathematically, the ReLU function is defined below:

$$f(z) = \max(0, z). \quad (3)$$

In the ITCN, the leaky rectifier linear unit (LReLU) [27] was used. This was because imposing zeros (see (3)) could negatively affect the calculation of gradients. During the training of this neural network, zero gradients will significantly slow down the adjustments of weights. The LReLU function reads

$$f(z) = \max(0, z) + \alpha \min(0, z), \quad (4)$$

where  $\alpha$  is the leakiness parameter [18].

To address the multiclassification problem, a well-known softmax function was used to transform the neural network outputs to probability distributions. Softmax is defined as follows:

$$Y_i = \text{soft max}(Z_i) = \frac{e^{Z_i}}{\|e^Z\|}, \quad (5)$$

where  $Z_i$  is the output from the  $i$ th neuron and  $Y_i$  is the probability of input pixel corresponding to the  $i$ th class. In the TLN,  $i = 1$  or  $2$  because the TLN was to perform a binary classification in the first step. In the ITCN,  $i = 1, 2, 3, 4$  since the ITCN was to classify the MRI data into four classes.

**3.2.4. Loss Function.** Given a set of weights of the proposed neural network  $\theta$ , a categorical cross-entropy loss function was used to compute the loss of ground truth and predicted probability distribution. Mathematically, under an

arbitrary prediction for the  $i$ th pixel, the prediction loss can be defined as

$$L(\theta) = - \sum_{j=1}^C \mathbf{Y}'_{ij} \log(Y_{ij}), \quad (6)$$

where  $\mathbf{Y}'$ ,  $Y$ , and  $C$  are a one-hot vector, the predicted probability distribution, and the number of classes, respectively.

In the TLN, predictions were made for each pixel of the input image so that the loss function can be written as follows:

$$L(\theta') = - \frac{1}{S} \sum_{i=1}^S \sum_{j=1}^C \mathbf{Y}'_{ij} \log(Y_{ij}), \quad (7)$$

where  $C = 2$  and  $S$  is the pixel number of the input image. In every training, only one input image was used (the size of minibatch was 1).

Now referring to the ITCN, the loss function was calculated in conjunction with the concept of mini-batch. Thus, the loss function has the following form,

$$L(\theta'') = - \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^C \mathbf{Y}'_{ij} \log(Y_{ij}), \quad (8)$$

where  $C = 4$  and  $M$  is the size of minibatch. Of note, in this study,  $M = 256$ .

To achieve better generation ability and avoid overfitting, L2 regularization terms were also added to (7) and (8). Thus, the final forms of the loss functions are

$$L(\theta') = -\frac{1}{S} \sum_{i=1}^S \sum_{j=1}^C \mathbf{Y}'_{ij} \log(Y_{ij}) + \frac{\lambda}{S} \sum_{k=1}^Q |\theta'_k|^2, \quad (9)$$

$$L(\theta'') = -\frac{1}{M} \sum_{i=1}^M \sum_{j=1}^C \mathbf{Y}'_{ij} \log(Y_{ij}) + \frac{\lambda}{M} \sum_{k=1}^Q |\theta''_k|^2, \quad (10)$$

where  $\lambda$  is a regularization constant and  $Q$  is the number of model parameter.

**3.2.5. Optimization Method.** Equations (9) and (10) were minimized using the minibatch stochastic gradient descent (SGD) algorithm. To avoid numerical oscillations and accelerate convergence, the momentum method [23] was used. This process can be described as iterations from (11) to (13).

$$\mathbf{g}_t = \nabla_{\theta_{t-1}} L(\theta_{t-1}), \quad (11)$$

$$\mathbf{m}_t = \mu * \mathbf{m}_{t-1} - \eta_t \mathbf{g}_t, \quad (12)$$

$$\theta_t = \theta_{t-1} + \mathbf{m}_t. \quad (13)$$

In (11), (12), and (13), the subscript  $t$  is the iteration number and  $\theta$  corresponds to  $\theta'$  in (9) or  $\theta''$  in (10).  $L(\theta_{t-1})$  is the loss function when a parameter set  $\theta_{t-1}$  is used.  $\mathbf{g}_t$ ,  $\mathbf{m}_t$ , and  $\mu$  are the gradient, momentum, and momentum coefficient, respectively. We set  $\mu = 0.99$  and  $\mu = 0.9$  in the TLN and ITCN, respectively. Here,  $\eta_t$  is the learning rate.

To suppress the SGD noise and guarantee convergence, the learning rate  $\eta_t$  attenuates linearly from the initial learning rate  $\eta_0$  to the final learning rate  $\eta_\tau$  as the iteration progresses:

$$\eta_t = (1 - \gamma)\eta_0 + \gamma\eta_\tau, \quad (14)$$

$$\gamma = \frac{t}{\tau}, \quad (15)$$

where  $\tau$  is the total iteration number. In this study, we set  $\eta_\tau = \eta_0/100$ .

**3.2.6. Training Details.** The initial and final learning rates of the TLN model were set to  $1e-8$  and  $1e-10$ , respectively. The total iteration  $\tau = 2e6$ , and the momentum coefficient was 0.99. In the ITCN subnet, the initial and final learning rates were set to  $1e-3$  and  $1e-5$ , respectively. In the ITCN subnet, the total iteration  $\tau = 2e6$  and the momentum coefficient  $\mu = 0.9$ .

During the training of the TLN subnet, we used the transfer learning technique [28, 29]. The initial weights were obtained from a pretrained model that was trained using ImageNet in [24]. But initial weights of the 4th input channel were initialized using the average of the original 3 input channel (RGB) weights. And the final two output channels were initialized with the Xavier method [30]. Then, fine-tuning of the TLN was performed by the optimization process described above ((11), (12), and (13)) using the MRI training data. However, the training of the ITCN subnet was started from scratch and the weights were initialized with the Xavier method [30]. To avoid overfitting, we used

the dropout regularization [31] and the dropout ratio was set to 0.5 in all fully connected layers. Weight decay was set as 0.005.

**3.3. Datasets and Evaluation Metrics.** In order to train and evaluate the proposed system, numerical experiments were carried out using *in vivo* human patient data provided by the BRATS 2015 database [32]. The BRATS 2015 database contains 220 HGG and 54 LGG. Experimental data have been labeled, and five labels were used: normal brain tissues (noncancerous zone), necrosis, edema, nonenhancing tumor, and enhancing tumor. These pixel-wise delineations were considered the ground truth in this study. Each case contains four sequences of MRI data, namely, T1, T1c, T2, and FLAIR. The dimension of each MRI modality is  $155 \times 240 \times 240$  (slice number  $\times$  length  $\times$  width). All MRI data were spatially registered and stored as signed 16-bit integers. But only positive values were used.

The tenfold crossvalidation method [33] was used to evaluate the proposed system. More specifically, the 274 cases were divided into a training set (240 cases) and a testing set (34 cases). The 240 training cases were equally divided into 10 subsets in which 9 subsets were used as the training and 1 subset was used as the validation. In the training phase of the TLN subnet, all subregions within a tumor were merged into one tumor region. Thus, in the binary ground truth, zero represents the noncancerous tissues while one represents cancerous regions. In the training phase of the ITCN subnet, we randomly selected 4,700,000 image patches ( $33 \times 33$ ) from the training set, which correspond to 1,175,000 patches for each label (4 different classes).

The quantitative evaluations were conducted for 3 different tumor regions: complete tumor region (including all four tumor subregions), core tumor region (including all tumor structures except edema), and enhancing tumor region (only including the enhanced tumor structure). For each type of regions, we compute DSC [34], PPV, and sensitivity [35] as quantitative evaluation metrics.

DSC measures the overlap between the ground truth and the automatic segmentation. It is defined as

$$\text{DSC} = \frac{P_1 \cap T_1}{(P_1 + T_1)/2}, \quad (16)$$

where  $P_1$  and  $T_1$  represent the positive values of the model prediction and the ground truth, respectively.

PPV is the proportion of the true positive in all segmentation tumor points. It is defined as

$$\text{PPV} = \frac{P_1 \cap T_1}{P_1}. \quad (17)$$

Sensitivity is the proportion of the detected tumor points in all ground truth tumor points. It is defined as

$$\text{Sensitivity} = \frac{P_1 \cap T_1}{T_1}. \quad (18)$$

The proposed system was compared with some other published methods. Those methods all have been validated

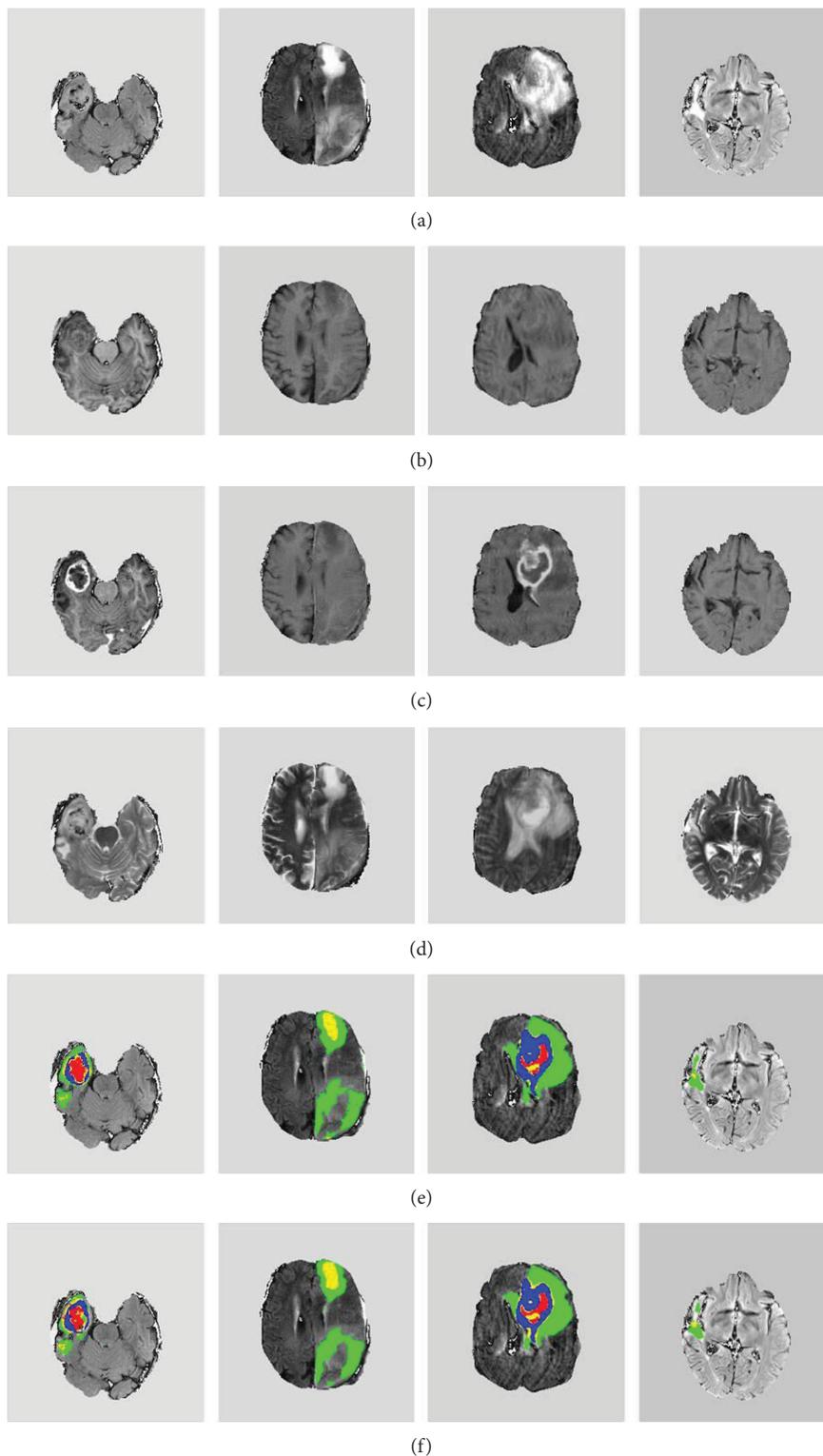


FIGURE 5: Representative examples of computer segmentation results of four brain tumors. (a–d) The original FLAIR, T1, T1c, and T2 slices, respectively. (e) The ground truth overlaid with the FLAIR image. (f) Segmentation results overlaid with the FLAIR image. (e, f) Red, green, yellow, and blue colors denote necrosis, edema, nonenhancing tumor, and enhancing tumor, respectively.

on the BRATS 2015 dataset. A one-step segmentation method based on the FCN-8s was also implemented for the purpose of comparison. The FCN-8s can segment the input MRI images into 5 classes in a single step.

## 4. Results

*4.1. Qualitative Observations.* Overall, we found that the proposed system can accurately delineate gliomas. Visual

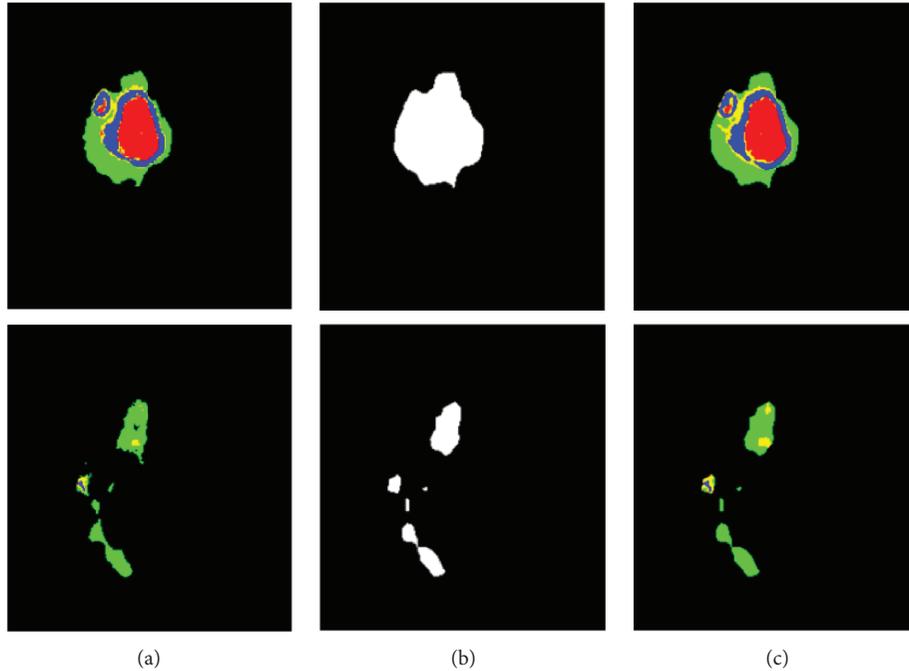


FIGURE 6: Two slices of computer segmentation result in a testing case: (a–c) the ground truth, results of tumor localization using the TLN subnet, and the intratumor segmentation results using the ITCN subnet, respectively. (a, c) Red, green, yellow, and blue colors denote necrosis, edema, nonenhancing tumor, and enhancing tumor, respectively.

inspections were conducted for testing data to validate the segmentation results of our proposed method. Figure 5 shows four selected examples. It can be observed that our method can effectively localize and segment brain tumors with vastly different shapes and sizes. Visually, the computer segmentation is comparable to the ground truth.

Also, the proposed system led to good details around boundaries. Figure 6 presents two representative examples of this observation. Since these brain tumors are complex, Figure 6 shows some good showcase examples. During the process, we found that the TLN subnet was able to effectively identify nearly all the tumor pixels. Subsequently, the ITCN subnet efficiently classified the tumor region into four subregions. Our method could largely detect the complete tumor and classify it to different tumor subregions from multimodality MRI images though there were a few misclassifications. This is not surprising because, pathologically, the brain glioma tumors invade their surrounding tissues rather than displacing them. Hence, the appearance of cancerous tissues and their surrounding (normal) tissues could be fairly similar under MRI.

We also found that, as compared to the FCN-8s with one-step segmentation, the proposed system could segment heterogeneous gliomas with a better boundary detail. The results of the proposed method and FCN-8s are compared in Figure 7. Five different typical slices representing significantly different tumor shapes and sizes are shown in this figure. It is easy to see that the results obtained from the proposed method (the third column) are more similar to the ground truth (the first column), as compared to the

classification results by the FCN-8s (the second column). Furthermore, boundaries of various subregions obtained by the FCN-8s were overly smoothed and, perhaps, inaccurate. But our method using the ITCN had better boundaries of the enhancing and nonenhancing regions.

**4.2. Evaluation and Comparison.** The quantitative comparisons with other methods in terms of DSC are summarized in Tables 5 and 6. All experiments were conducted on the BRATS 2015 dataset. The results of Table 5 were obtained by using the combined testing set of HGG and LGG, whereas results shown in Table 6 only used HGG data.

Obviously, the proposed cascaded neural network obtains the comparable and better DSC value on all tumor regions. Based on the combined testing dataset (see Table 5), our method obtained better comprehensive performance values (0.89, 0.77, and 0.80) as compared to other methods. Although the method proposed by Kamnitsas et al. [21] yields a slightly higher DSC value in the complete tumor, they obtained lower DSC values in core tumor and enhancing tumor. Actually, in their work, a 3D CNN and the structure prediction technology were adopted (i.e., conditional random field). Thus, it is computationally time-consuming and needs extra postprocessing. Furthermore, the method proposed by Dong et al. [36] yielded a slightly higher DSC value in core tumor and Yi et al. [37] yielded the same DSC value in enhancing tumor.

As can be seen in Table 6, based on the HGG testing dataset, our method obtained the highest DSC values in the complete tumor and enhancing tumor categories. Although the

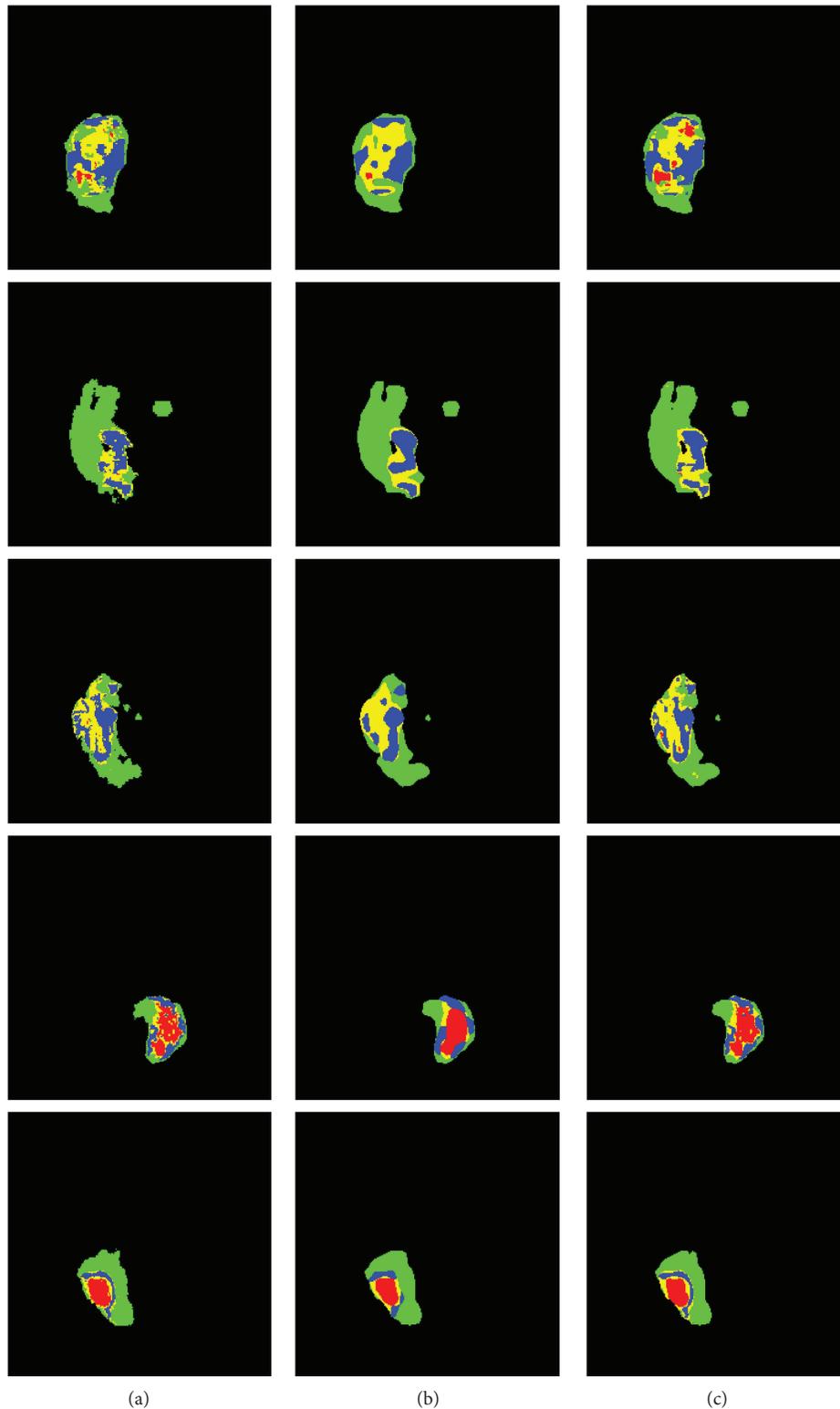


FIGURE 7: Examples of segmentation results from five typical slices comparing the FCN-8s (b) and the proposed method (c). (a) The ground truth. In this figure, red, green, yellow, and blue colors denote necrosis, edema, nonenhancing tumor, and enhancing tumor, respectively.

method proposed by Dong et al. [36] yielded a higher DSC value in the core tumor cases, it obtained a lower DSC value in the complete tumor category.

Recently, we found that Pereira et al. [39] also proposed a hierarchical brain tumor segmentation approach from MRI HGG images. The difference between their method and our

TABLE 5: A summary of DSC quantitative comparison on BRATS 2015 combined dataset (HGG and LGG).

Method	Dataset	Grade	DSC		
			Complete	Core	Enh
Pereira et al. [38]	BRATS 2015 Challenge	Combined	0.78	0.65	0.75
	BRATS 2015 Training	Combined	0.87	0.73	0.68
Havaei et al. [19]	BRATS 2015 Challenge	Combined	0.79	0.58	0.69
Kamnitsas et al. [21]	BRATS 2015 Challenge	Combined	0.85	0.67	0.63
	BRATS 2015 Training	Combined	0.90	0.76	0.73
Dong et al. [36]	BRATS 2015 Training	Combined	0.86	0.86	0.65
Yi et al. [37]	BRATS 2015 Training	Combined	0.89	0.76	0.80
FCN-8s	BRATS 2015 Training	Combined	0.84	0.71	0.63
Proposed	BRATS 2015 Training	Combined	0.89	0.77	0.80

TABLE 6: A summary of DSC quantitative comparison on BRATS 2015 HGG dataset.

Method	Dataset	Grade	DSC		
			Complete	Core	Enh
Pereira et al. [38]	BRATS 2015 Training	HGG	0.87	0.75	0.75
Havaei et al. [19]	BRATS 2015 Challenge	HGG	—	—	—
Kamnitsas et al. [21]	BRATS 2015 Training	HGG	—	—	—
Dong et al. [36]	BRATS 2015 Training	HGG	0.88	0.87	0.81
Yi et al. [37]	BRATS 2015 Training	HGG	0.89	0.79	0.80
FCN-8s	BRATS 2015 Training	HGG	0.88	0.76	0.71
Proposed	BRATS 2015 Training	HGG	0.90	0.81	0.81

TABLE 7: A comparison of our proposed method with hierarchical brain tumor segmentation [39] on DSC, PPV, and sensitivity metrics.

Method	DSC			PPV			Sensitivity		
	Complete	Core	Enh	Complete	Core	Enh	Complete	Core	Enh
Pereira et al. [39]	0.85	0.76	0.74	0.80	0.78	0.74	0.92	0.79	0.78
Proposed	0.90	0.81	0.81	0.91	0.77	0.87	0.87	0.84	0.76

TABLE 8: Comparisons of segmentation time among six different methods. The estimation of time for the proposed method was based on the acceleration of GPU.

Method	Time
Pereira et al. [18]	8 s–24 min
Havaei et al. [19]	8 min
Kamnitsas et al. [21]	30 s
Dong et al. [36]	2–3 s
FCN-8s	0.98 s
Proposed	1.54 s

method is that they adopted the FCN in both first and second steps. We compared the results of our method with their method (see Table 7). Our proposed approach obtained the better DSC values (0.90, 0.81, and 0.81) in all tumor regions. Furthermore, the proposed method also yielded higher PPV values in the complete and enhancing tumor categories and a higher sensitivity in the core tumor category. Of note, Pereira et al. [39] trained and tested on the BRATS 2013 dataset but we on the BRATS 2015 dataset.

Additionally, the segmentation speed for testing data was also documented (see Table 8). Computational performance of the first four methods was obtained through respective publications [18, 19, 21, 36]. The proposed method is efficient as compared to other methods. It only takes averagely 1.54 seconds in order to segment a slice and only runs slightly slower than the FCN-8s (0.98 seconds). This is understandable because the proposed method needs two-stage segmentation while the FCN-8s only needs a forward computation. However, the FCN-8s yields less accurate and overly smooth boundary maps. Of note, adopting the FCN for image semantic segmentation is faster than the traditional method based on patch-wise [22, 36]; despite computational efficiency, tests reported in the literature were done using slightly different computing platforms.

## 5. Discussions and Conclusions

In this work, a cascaded neural network was designed, implemented, and tested. The proposed system consists of two steps. In the first step, the TLN subnet was used to localize the brain tumor. Then, the ITCN subnet was applied to the

identified tumor regions to further classify the tumor into four subregions. We also adopted the advanced technologies to train and optimize the proposed cascaded neural network. Numerical experiments were conducted on 274 patient *in vivo* data sets from the BRATS 2015. DSC, PPV, and sensitivity were used as metrics for segmentation accuracy.

Based on quantitative and qualitative evaluations, we found that the proposed approach was able to accurately localize and segment complex brain tumors. We stipulate that there are two reasons. First, the ITCN subnet only represents and subsequently classifies the intratumoral region whereas other methods need to represent and classify all heterogeneous brain tissues. Second, intratumor subregions are usually very small proportions of the entire image. Other neural networks (e.g., FCN-8s) may suffer from the imbalance of different pixel labels. In the TLN subnet, our proposed method merged different tumor subregions into a whole tumor. Thus, the imbalance can be somewhat mitigated. In the ITCN subnet, we adopted the same quantity image patches of each class to train and optimize the model. In the future, deep learning neural networks could be expanded to include histological data and other data to further improve clinical management of brain cancers [40].

Furthermore, the proposed cascaded neural network can, on average, complete a segmentation task within 1.54 seconds. The proposed TLN subnet only requires a forward computation for localizing the whole tumor region in the first step. Then, the ITCN subnet only needs to classify tumor candidate pixels into different class subregions within a much-reduced region located by the TLN, thereby improving the computing efficiency.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research is funded by Chongqing Science and Technology Commission (Grant no. cstc2016jcyjA0383) and Humanity and Social Science Key Project of Chongqing Municipal Education Commission (Grant no. 16SKGH133). This research is also in part supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant no. KJ1709210) and Graduate Innovation Fund of Chongqing University of Technology (Grant no. YCX2016230).

## References

- [1] D. N. Louis, H. Ohgaki, O. D. Wiestler et al., "The 2007 WHO classification of tumours of the central nervous system," *Acta Neuropathologica*, vol. 114, no. 2, pp. 97–109, 2007.
- [2] S. Bauer, R. Wiest, L. P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Physics in Medicine & Biology*, vol. 58, no. 13, pp. R97–R129, 2013.
- [3] E. G. Van Meir, C. G. Hadjipanayis, A. D. Norden, H. K. Shu, P. Y. Wen, and J. J. Olson, "Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma," *CA: A Cancer Journal for Clinicians*, vol. 60, no. 3, pp. 166–193, 2010.
- [4] G. Tabatabai, R. Stupp, M. J. van den Bent et al., "Molecular diagnostics of gliomas: the clinical perspective," *Acta Neuropathologica*, vol. 120, no. 5, pp. 585–592, 2010.
- [5] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille, "Efficient multilevel brain tumor segmentation with integrated Bayesian model classification," *IEEE Transactions on Medical Imaging*, vol. 27, no. 5, pp. 629–640, 2008.
- [6] A. Hamamci, N. Kucuk, K. Karaman, K. Engin, and G. Unal, "Tumor-cut: segmentation of brain tumors on contrast enhanced MR images for radiosurgery applications," *IEEE Transactions on Medical Imaging*, vol. 31, no. 3, pp. 790–804, 2012.
- [7] I. Mehmood, N. Ejaz, M. Sajjad, and S. W. Baik, "Prioritization of brain MRI volumes using medical image perception model and tumor region segmentation," *Computers in Biology and Medicine*, vol. 43, no. 10, pp. 1471–1483, 2013.
- [8] M. Havaei, H. Larochelle, P. Poulin, and P.-M. Jodoin, "Within-brain classification for brain tumor segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 5, pp. 777–788, 2016.
- [9] K. Usman and K. Rajpoot, "Brain tumor classification from multi-modality MRI using wavelets and machine learning," *Pattern Analysis and Applications*, vol. 20, no. 3, pp. 871–881, 2017.
- [10] N. J. Tustison, K. L. Shrinidhi, M. Wintermark et al., "Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR," *Neuroinformatics*, vol. 13, no. 2, pp. 209–225, 2015.
- [11] D. Zikic, B. Glocker, E. Konukoglu et al., "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012. MICCAI 2012, vol 7512, Lecture Notes in Computer Science*, N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds., Springer, Berlin, Heidelberg, 2012.
- [12] A. Pinto, S. Pereira, H. Correia, J. Oliveira, D. M. Rasteiro, and C. A. Silva, "Brain tumour segmentation based on extremely randomized forest with high-level features," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, Italy, August 2015.
- [13] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2011. MICCAI 2011, vol 6893, Lecture Notes in Computer Science*, G. Fichtinger, A. Martel, and T. Peters, Eds., Springer, Berlin, Heidelberg, 2011.
- [14] G. Urban, M. Bendszus, F. Hamprecht, and J. Kleesiek, "Multimodal brain tumor segmentation using deep convolutional neural networks," in *MICCAI BraTS (Brain Tumor Segmentation) Challenge. Proceedings, Winning Contribution*, pp. 31–35, Boston, MA, USA, 2014.
- [15] D. Zikic, Y. Ioannou, M. Brown, and A. Criminisi, "Segmentation of brain tumor tissues with convolutional neural networks," in *Proceedings MICCAI-BRATS*, pp. 36–39, Boston, MA, USA, 2014.

- [16] A. Davy, M. Havaei, D. Warde-Farley et al., "Brain tumor segmentation with deep neural networks," in *Proceedings MICCAI-BRATS*, Boston, MA, USA, 2014.
- [17] P. Dvorak and B. Menze, "Structured prediction with convolutional neural networks for multimodal brain tumor segmentation," in *Proceedings MICCAI-BRATS*, pp. 13–24, Munich, Germany, 2015.
- [18] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [19] M. Havaei, A. Davy, D. Warde-Farley et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.
- [20] M. Lyksborg, O. Puonti, M. Agn, and R. Larsen, "An ensemble of 2D convolutional neural networks for tumor segmentation," in *Scandinavian Conference on Image Analysis. SCIA 2015, vol 9127*, Lecture Notes in Computer Science, Springer, Cham, 2015.
- [21] K. Kamnitsas, C. Ledig, V. F. J. Newcombe et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.
- [22] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.
- [23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, Carson, NV, USA, 2012.
- [24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <http://arxiv.org/abs/1409.1556>.
- [25] B. B. Avants, N. Tustison, and G. Song, "Advanced normalization tools (ANTS)," *Insight Journal*, vol. 2, pp. 1–35, 2009.
- [26] K. Jarrett, K. Kavukcuoglu, and Y. LeCun, "What is the best multi-stage architecture for object recognition?," in *2009 IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, September–October 2009.
- [27] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, GA, USA, 2013.
- [28] H.-C. Shin, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [29] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [30] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia, Italy, 2010.
- [31] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [32] B. H. Menze, A. Jakab, S. Bauer et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
- [33] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*, vol 7700pp. 437–478, Springer, Berlin, Heidelberg.
- [34] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [35] K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, "Joint sequence learning and cross-modality convolution for 3D biomedical segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.
- [36] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," in *Medical Image Understanding and Analysis. MIUA 2017. Communications in Computer and Information Science*, vol 723, M. Valdés Hernández and V. González-Castro, Eds., Springer, Cham, 2017.
- [37] D. Yi, M. Zhou, Z. Chen, and O. Gevaert, "3-D convolutional neural networks for glioblastoma segmentation," 2016, <http://arxiv.org/abs/1611.04534>.
- [38] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes*, A. Crimi, B. Menze, O. Maier, M. Reyes, and H. Handels, Eds., vol. 9556 of Lecture Notes in Computer Science, Springer, Cham, 2015.
- [39] S. Pereira, A. Oliveira, V. Alves, and C. A. Silva, "On hierarchical brain tumor segmentation in MRI using fully convolutional neural networks: a preliminary study," in *2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)*, Coimbra, Portugal, February 2017.
- [40] R. F. Barajas Jr, J. G. Hodgson, J. S. Chang et al., "Glioblastoma multiforme regional genetic and cellular expression patterns: influence on anatomic and physiologic MR imaging," *Radiology*, vol. 254, no. 2, pp. 564–576, 2010.

## Research Article

# A Deep Belief Network and Dempster-Shafer-Based Multiclassifier for the Pathology Stage of Prostate Cancer

Jae Kwon Kim <sup>1</sup>, Mun Joo Choi,<sup>2</sup> Jong Sik Lee,<sup>1</sup> Jun Hyuk Hong,<sup>3</sup> Choung-Soo Kim <sup>3</sup>,  
Seong Il Seo,<sup>4</sup> Chang Wook Jeong,<sup>5</sup> Seok-Soo Byun,<sup>6</sup> Kyo Chul Koo,<sup>7</sup> Byung Ha Chung,<sup>7</sup>  
Yong Hyun Park,<sup>8</sup> Ji Youl Lee,<sup>8</sup> and In Young Choi <sup>2</sup>

<sup>1</sup>Department of Computer Science and Information Engineering, Inha University, InhaRo 100, Nam-gu, Incheon, Republic of Korea

<sup>2</sup>Department of Medical Informatics, College of Medicine, The Catholic University of Seoul, 222 Banpo-daero, Seocho-gu, Seoul 06591, Republic of Korea

<sup>3</sup>Department of Urology, University of Ulsan College of Medicine, Seoul, Republic of Korea

<sup>4</sup>Department of Urology, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

<sup>5</sup>Department of Urology, Seoul National University College of Medicine, Seoul, Republic of Korea

<sup>6</sup>Department of Urology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

<sup>7</sup>Department of Urology, Yonsei University College of Medicine, Seoul, Republic of Korea

<sup>8</sup>Department of Urology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

Correspondence should be addressed to In Young Choi; [iychoi@catholic.ac.kr](mailto:iychoi@catholic.ac.kr)

Received 25 August 2017; Accepted 9 January 2018; Published 19 March 2018

Academic Editor: Ming-Yuan Hsieh

Copyright © 2018 Jae Kwon Kim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Object.** Pathologic prediction of prostate cancer can be made by predicting the patient's prostate metastasis prior to surgery based on biopsy information. Because biopsy variables associated with pathology have uncertainty regarding individual patient differences, a method for classification according to these variables is needed. **Method.** We propose a deep belief network and Dempster-Shafer- (DBN-DS-) based multiclassifier for the pathologic prediction of prostate cancer. The DBN-DS learns prostate-specific antigen (PSA), Gleason score, and clinical T stage variable information using three DBNs. Uncertainty regarding the predicted output was removed from the DBN and combined with information from DS to make a correct decision. **Result.** The new method was validated on pathology data from 6342 patients with prostate cancer. The pathology stages consisted of organ-confined disease (OCD; 3892 patients) and non-organ-confined disease (NOCD; 2453 patients). The results showed that the accuracy of the proposed DBN-DS was 81.27%, which is higher than the 64.14% of the Partin table. **Conclusion.** The proposed DBN-DS is more effective than other methods in predicting pathology stage. The performance is high because of the linear combination using the results of pathology-related features. The proposed method may be effective in decision support for prostate cancer treatment.

## 1. Introduction

Prostate cancer is the most common cancer in men, with around 1.1 million cases diagnosed and approximately 309,000 deaths in men worldwide in 2012 [1]. It is estimated that 40–50% of men may also have potentially extraprostatic disease [2].

Carcinectomy and radiotherapy are the typical treatments for prostate cancer [3]. The choice of treatment for

prostate cancer requires extensive experience and analysis of treatment cases. Pathological staging is the process of predicting the likelihood of prostate cancer disease spreading in a patient prior to treatment. The clinical stage evaluation is based on data gathered from clinical tests that are available prior to treatment or the surgical removal of the tumor. Cancer staging evaluation occurs both before and after the tumor is removed: the clinical and pathological stages, respectively [4]. Pathological staging is determined after the removal of

the tumor tissue and after surgery. This is more likely to be more accurate than clinical staging because it evaluates the direct nature of the disease. Therefore, the prediction of pathological stages using clinical data analysis is an important factor in the treatment of prostate cancer [5].

Pathologic staging prediction is very important because it provides physicians with optimal treatment and management strategies. For example, radical prostatectomy (RP), the surgical removal of the prostate gland, provides the best opportunity for cure when prostate cancer is localized and accurate prediction of the pathology stage can provide the most beneficial treatment approach [6–8]. Currently, Partin tables are used to predict the prognostic clinical outcome for prostate cancer, which are based on statistical methods such as logistic regression [9, 10]. The Partin tables use clinical test data including prostate-specific antigen (PSA) level, Gleason score, and clinical T stage to predict the pathology stage. While the Partin tables have been verified from 2001 to 2011, there are questions about their applicability to current patients following environmental changes [11]. Thus, a new classification method using machine learning is needed to provide an accurate prediction of the pathology stage [12].

Deep belief networks (DBN) are a deep learning technique and is an effective method for classification prediction [13, 14]. As DBN supports both unsupervised and supervised learning, it is possible to effectively learn about uncertain data relationships [15, 16]. Because PSA level, Gleason score, and clinical T stage for stage prediction have uncertainties in each patient, a combination of evidence for each variable is needed. The Dempster-Shafer theory (DS) is a technique used to fuse information based on trust values [17, 18]. The DS allows the combination of evidence from different sources to arrive at a degree of belief (represented by a mathematical object called a “belief function”) that considered all available evidence [19, 20]. This technique is a method for fusing information using a stochastic calculation method for belief values [21]. This allows fusion of the classification results of each variable to the pathology stage.

In this paper, we propose a DBN-DS-based multiclassifier for pathologic stage prediction of prostate cancer. The proposed DBN-DS uses patient PSA level, Gleason score, and clinical T stage and three DBNs to predict the pathology stage by combining the predicted information from the classifier. The classifiers are created by learning data according to features. When output values are generated using each learned DBN classifier, the final predicted result is provided by stochastically calculating the predicted output from each DBN classifier using DS. This paper is organized as follows: Section 2 presents the proposed technique and its process. Section 3 explains the experiments and presents their outcomes. Finally, Section 4 presents the conclusions.

## 2. Materials and Methods

**2.1. Data Set.** The study data comprised 6345 male patients extracted from the Korean Prostate Cancer Registry (KPCR) which is extended from Smart Prostate Cancer Data Base (SPCDB) at six tertiary medical centers in Korea [22]. The three input variables consist of initial PSA, Gleason score,

TRUS volume, and clinical T stage. Two output variables consisting of pathologic T stage (pT2a, pT2b, pT3a, pT3b, and pT3c) and N stage (pN1) were used. The output variables are transformed using the guidelines of the American Joint Committee on Cancer (AJCC), which were used to identify the pathologic stage as organ-confined disease (OCD; pT2+) or non-organ-confined disease (NOCD; pT3+ or N+) [23]. For the experiments, the data from the KPCR were divided into a training set 70% (4039 patients) and a validation set 30% (2306 patients).

**2.2. Deep Belief Network.** A deep belief network (DBN) is a generative graphical model or a type of deep neural network composed of multiple layers of latent variables, with connections between the layers but not between the units within each layer. The DBN is composed of restricted Boltzmann machine (RBM) layers. The learning method in the DBN is done by configuring the visible layer and hidden layer 1 into a single RBM. The DBN is composed of multiple layers of RBMs [24]. The RBMs consist of visible and hidden unit layers. Once learning is complete, hidden layers 1 and 2 are trained via the RBM by giving a new input as a value of the hidden layer 1. As such, learning is performed up to the last layer sequentially [25]. One classification technique using the DBN is back propagation, which is configured in the uppermost layer in the DBN [26]. This technique shows better results than an artificial neural network (ANN), which uses a connection intensity that is arbitrarily selected.

In this study, we constructed a classifier for three input and two output variables to construct a multiclassifier, as shown in Figure 1. We created one classifier for each variable. Our idea was to use multiclassifiers for each variable [27]. The purpose of this study was to make a linear combination of the predictions of the classifiers using DS [28]. Therefore, one variable must be converted into several input values. As PSA levels are continuous data, they were converted into binary numbers and configured as an input node. Because Gleason score and the clinical T stage are categorical data, they constitute an input node by constructing data in flag form.

**2.3. Dempster-Shafer-Based Information Fusion.** Dempster-Shafer (DS) is a mathematical theory that deals with the uncertainty and inaccuracy problems presented by Arthur Dempster and Glenn Shafer [29]. The DS provides an effective method for establishing evidence intervals using belief and likelihood values for the data set. The DS can support the combination of information. As a result, it is possible to use a combination rule to set various information as an evidence value and to calculate the result of all the evidence [30].

The DS expresses the degree of certainty as a section and sets mutually exclusive hypotheses such as probability. The set of objects is called the environment and is denoted by  $\theta$ . The  $\theta$  can have several elements such as  $\theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_k\}$ , and the number of subsets is  $2^k$ . When  $\theta$  has only one element, it is called an identification frame. A set of  $2^k$  subsets is called a power set and is denoted by  $\theta$ . The degree to which  $\theta$  is supported by any evidence is called the basic probability assignment function  $m(1)$ . The  $m$  is mapped to a probability

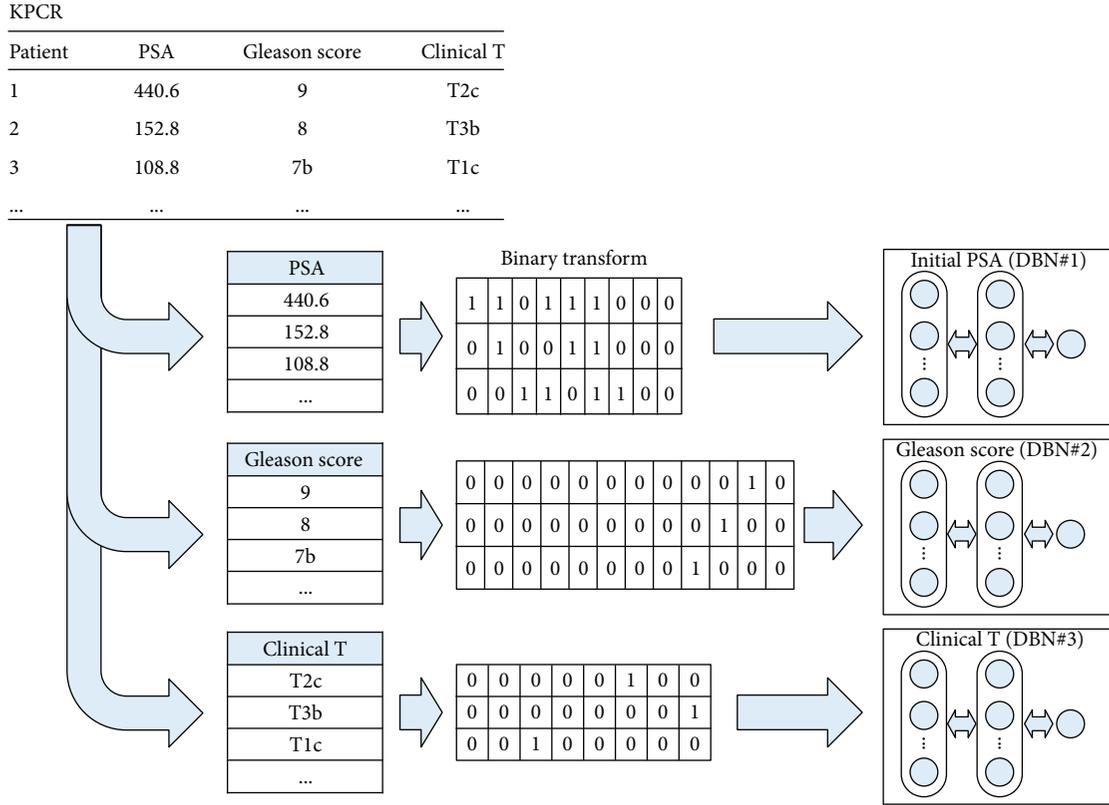


FIGURE 1: Multi DBN classifiers.

value of 0 for an empty set, and the sum of  $m$  is 1 for all subsets of  $\theta$  (2).

$$m : 2^k \rightarrow [0, 1], \quad (1)$$

$$m(\emptyset) = 0,$$

$$\sum_{U \in 2^k} m(U) = 1. \quad (2)$$

Belief ( $H$ ), which is the belief value for any hypothesis  $H$  (hypothesis; belief in a hypothesis is constituted by the sum of the masses of all sets enclosed by subjective probabilities) by given evidence, as shown in

$$\text{Bel}(H) = \sum_{U \in H} m(U). \quad (3)$$

The degree of trust depends on the reliability of the given evidence and on the overall environmental impact; the ratio of the degree is expressed by  $e$ .

$$m^e(A) = \begin{cases} (1-r)m(A), & A \subset \theta, \\ r + (1-r)m(\theta), & A = \theta, \end{cases} \quad (4)$$

where  $r$  is a value between 0 and 1 and is true if  $r=0$  and false if  $r=1$ . The DS calculates the value of a new belief through the process of fusion between different evidence. Thus, the convergence between the evidence can be expressed as (5); if  $X \cap Y = \emptyset$ , then the convergence value of the two evidence is zero.

$$m_1 \oplus m_2(A) = (1-p)^{-1} \sum_{X \cap Y = A} m_1(X)m_2(Y), \quad (5)$$

$$p = \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y). \quad (6)$$

The DS expresses the confidence measure for  $H$  as  $[\text{Bel}(H), \text{Pls}(H)]$  and the term as the interval. This interval is called the “evidential interval.” Plausibility (Pls) means the extent to which the hypothesis is not negated based on evidence (empty period except for true and false intervals), which means the maximum likelihood of being trusted. Bel has a range from 0 to 1 (true and false), Pls can be defined as in (7) and has a value of  $[0,1]$ . Likewise, the likelihood values can express the process of fusion from multiple evidence as well as the fusion of belief values.

$$\text{Pls}(H) = 1 - \text{Bel}(\neg H), \quad (7)$$

$$\text{Pls}(U) = (((\text{Pls}_1 \oplus \text{Pls}_2) \oplus \text{Pls}_3) \oplus \dots) \oplus \text{Pls}_n. \quad (8)$$

In this study, three output data predicted from a multiclassifier were fused and calculated. The calculation process using DS shown in the figure as  $DBN\#1$  (initial PSA) was set to  $m_1$ ,  $DBN\#2$  (Gleason score) was set to  $m_2$ , and  $DBN\#3$  (clinical T stage) was set to  $m_3$ . For the output data, the empty set of each of  $m_1$ ,  $m_2$ , and  $m_3$  is given by

$$m_1(\emptyset) = 1 - (m_1(\text{OCD}) + m_1(\text{NOCD})), \quad (9)$$

TABLE 1: Summary of initial PSA by pathology stage (organ-confined or non-organ-confined disease) in 6345 patients with clinically localized prostate carcinoma.

	Training set ( $n = 4039$ )		Validation set ( $n = 2306$ )	
	OCD ( $n = 2478$ )	NOCD ( $n = 1561$ )	OCD ( $n = 1414$ )	NOCD ( $n = 892$ )
Initial PSA				
Minimum	4	4	4	4
Maximum	160	440.60	81.13	164
Average	9.535 (0.173)	18.606 (0.622)	9.377 (0.197)	17.889 (0.653)

$$m_2(\emptyset) = 1 - (m_2(\text{OCD}) + m_2(\text{NOCD})), \quad (10)$$

$$m_3(\emptyset) = 1 - (m_3(\text{OCD}) + m_3(\text{NOCD})). \quad (11)$$

As described above,  $m_1$ ,  $m_2$ , and  $m_3$  were obtained, and then  $m_4$  is combined. The combination of  $m_4$  is shown in

$$\begin{aligned} m_4(\text{OCD}) &= m_1(\text{OCD}) \oplus m_2(\text{OCD}) \oplus m_3(\text{OCD}) \\ &= \frac{\sum_{\text{OCD} \cap \text{NOCD} = \text{OCD}} m_1(\text{OCD}) m_2(\text{OCD}) m_3(\text{OCD})}{\sum_{\text{OCD} \cap \text{NOCD} = \emptyset} m_1(\text{OCD}) m_2(\text{OCD}) m_3(\text{NOCD})}, \\ m_4(\text{NOCD}) &= m_1(\text{NOCD}) \oplus m_2(\text{NOCD}) \oplus m_3(\text{NOCD}) \\ &= \frac{\sum_{\text{OCD} \cap \text{NOCD} = \text{NOCD}} m_1(\text{NOCD}) m_2(\text{NOCD}) m_3(\text{NOCD})}{\sum_{\text{OCD} \cap \text{NOCD} = \emptyset} m_1(\text{NOCD}) m_2(\text{NOCD}) m_3(\text{NOCD})}. \end{aligned} \quad (12)$$

Next, the interval of the pass and fail of the evidential interval are summarized as

$$\begin{aligned} \text{Bel}(\text{OCD}) &= m_3(\text{OCD}), \\ \text{Pls}(\text{OCD}) &= 1 - \neg \text{Bel}(\text{OCD}), \\ \text{Bel}(\text{NOCD}) &= m_3(\text{NOCD}), \\ \text{Pls}(\text{NOCD}) &= 1 - \neg \text{Bel}(\text{NOCD}). \end{aligned} \quad (13)$$

As described above, the evidential interval section is constructed for OCD and NOCD, and the higher probability value of OCD and NOCD was set as the final output value.

Uncertainty data processing is a critical issue in the data fusion process. The DS and the Bayesian methods were compared to deal with this uncertainty. Unlike Bayesian inference, DS can contribute different levels of information to each source. In addition, a popular approach to data fusion has been established; unlike the Bayesian method, reliability can be assigned to all subsets of a hypothetical group, making it possible to form distributions for all subsets [31].

### 3. Result

**3.1. Dataset Description.** The characteristics of the initial PSA variable in the OCD and NOCD groups are shown in Table 1. Among the 6345 men, the average PSA levels in the OCD and NOCD groups in the training set were 9.535 and 18.606 ng/mL, respectively. In general, the level in the

TABLE 2: Distribution of Gleason scores by pathology stage (organ-confined or non-organ-confined disease) in 6345 patients with clinically localized prostate carcinoma.

Gleason score	Training set ( $n = 4039$ )		Validation set ( $n = 2306$ )	
	OCD (%) ( $n = 2478$ )	NOCD (%) ( $n = 1561$ )	OCD (%) ( $n = 1414$ )	NOCD (%) ( $n = 892$ )
3	3 (0.12)	0 (0.00)	0 (0.00)	1 (0.11)
4	5 (0.20)	5 (0.33)	6 (0.42)	1 (0.11)
5	6 (0.24)	11 (0.73)	8 (0.57)	1 (0.11)
6	1342 (54.16)	378 (24.93)	785 (55.52)	235 (26.35)
7 (3 + 4)	565 (22.80)	386 (25.46)	306 (21.64)	218 (24.44)
7 (4 + 3)	266 (10.73)	277 (18.27)	160 (11.32)	159 (17.83)
8	238 (9.60)	326 (21.50)	119 (6.42)	174 (19.51)
9	46 (1.88)	147 (9.70)	28 (1.98)	95 (10.65)
10	7 (0.28)	31 (2.04)	2 (0.14)	8 (0.90)

TABLE 3: Distribution of clinical T stage by pathology stage (organ-confined disease and non-organ-confined disease) in 6345 patients with clinically localized prostate carcinoma.

Clinical T stage	Training set ( $n = 4039$ )		Validation set ( $n = 2306$ )	
	OCD (%) ( $n = 2478$ )	NOCD (%) ( $n = 1561$ )	OCD (%) ( $n = 1414$ )	NOCD (%) ( $n = 892$ )
T1a	9 (0.36)	0 (0.00)	3 (0.21)	0 (0.00)
T1b	107 (4.32)	49 (3.23)	74 (5.23)	18 (2.02)
T1c	988 (39.87)	410 (27.04)	556 (39.32)	225 (25.22)
T2a	691 (27.89)	380 (25.07)	417 (29.49)	241 (27.02)
T2b	278 (11.22)	161 (10.62)	151 (10.68)	97 (10.87)
T2c	234 (9.44)	224 (14.78)	126 (8.91)	127 (14.24)
T3a	150 (6.05)	233 (15.37)	66 (4.67)	135 (15.13)
T3b	21 (0.85)	104 (6.86)	21 (1.49)	49 (5.49)

OCD group was higher, and the validation set also shows a difference of 9.377 and 17.899 ng/mL in the OCD and NOCD groups, respectively. The difference in values between the training and validation sets was not large. Although a high number of patients were observed at maximum, this is not a problem for analysis because they were only a fraction of the outlier compared to the mean.

The Gleason scores in the OCD and NOCD groups are shown in Table 2. Patients with OCD had a high Gleason score of 6. The NOCD group had scores of 6 or more. The difference between the OCD and NOCD groups was significant. In the scores below 5, OCD is more distributed than NOCD, and even more than 9 patients showed more NOCD patients.

The clinical T stages in the OCD and NOCD groups are shown in Table 3. Most patients were T2+. T1a occurred only in patients with OCD. In addition, many patients that are distributed in OCD until T1+ and patients with T3+ belong to NOCD. Although all variables are bounded by OCD and NOCD, there are many patients who belong to the same distributions.

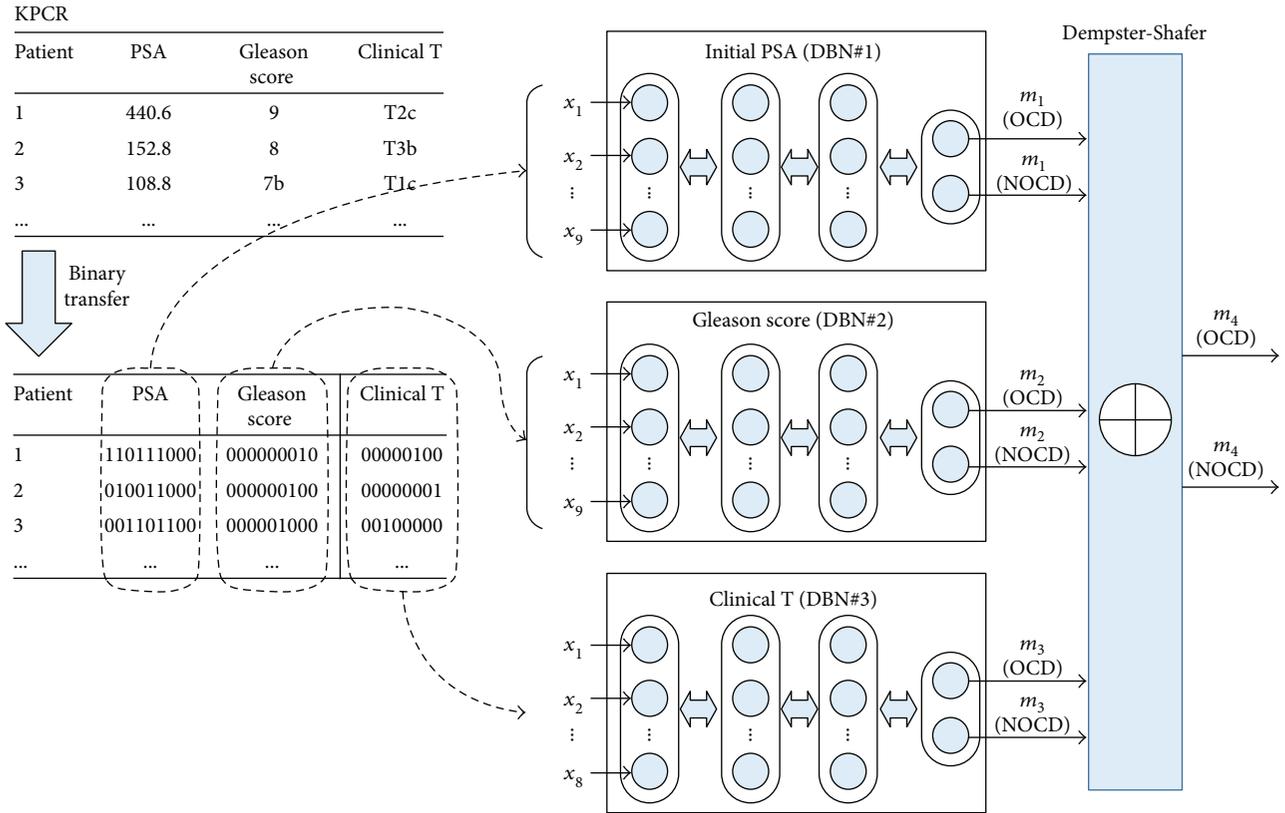


FIGURE 2: DBN-DS-based multiclassifier.

3.2. *DBN-DS Based Multiclassifier.* The proposed DBN and DS-based multiclassifier is shown in Figure 2. The training set was first changed to binary form. The initial PSA values were expressed as nine binary numbers based on the highest value (440 ng/mL). The Gleason score was composed of nine flags ranging from 3 to 10. The clinical T stage consisted of eight flags from T1a to T3b. The binary data of each of these variables was learned by the DBN classifier; that is, the first DBN consisted of nine input nodes because it was the input data of the initial PSA binary data. The output nodes of all classifiers were composed of two so that OCD and NOCD could be calculated with probability. The DBN consisted of three RBM layers, with the number of nodes of each RBM the same as the number of input nodes. Unsupervised learning was performed 100 times in total, while supervised learning using back propagation was performed 1000 times. Finally, we calculated the probability of the output variables as DS and determined the final number of  $m_4(\text{OCD})$  and  $m_4(\text{NOCD})$  as the final outputs.

3.3. *Experiments.* To evaluate the DBN-DS-based multiclassifier, the entire data set was divided into a 70% training set and a 30% testing set. The control groups included Decision Tree C4.5, naive Bayesian (NB), logistic regression (LR), back propagation (BP), support vector machine (SVM), random forest (RF), deep belief network, and Partin tables. The experiments compared the sensitivity, specificity, accuracy, and area under the curve (AUC) using confusion matrix [31]

and receiver operating characteristics (ROC) curve analysis [32]. The experimental results of confusion matrix are shown in Table 4.

In general, the results from a training set are better than those of a validation set because of differences in dataset volumes. Sensitivity was defined as the probability of correctly matching NOCD. Because NOCD has less data than OCD, it is difficult to match. The proposed method has a 61.77% improved performance compared to those of the other models. In other words, the probability of matching NOCD is very important because it is a prediction of the risk of the pathology stage. Specificity was defined as the probability of correctly matching OCD. NB had the highest specificity, with 93.78%, but its sensitivity was low. The proposed method showed 93.56% higher performance than those of the other models. The accuracy was defined as the probability of predicting both NOCD and OCD. The proposed model had the highest accuracy, at 81.27%. The AUCs are shown in Figure 3 and Table 5.

The ROC curve has the highest DBN-DS of 0.777. The error of all models was about 0.01, and the  $p$  values were all 0.000, so the experimental results of the ROC curves were usable. The DBN-DS predicted each of the three classifiers constructed for each variable separately and combined them into one. In this paper, we propose a new classification method for the classifier. The proposed method is based on the classification of two classifiers. In addition, as the DS computes probability, if one classifier predicts NOCD at a

TABLE 4: Experimental results of all classification methods between the training and validation sets.

	Training set			Validation set		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Partin table	45.96%	88.44%	70.52%	52.69%	71.36%	64.14%
C4.5	64.46%	91.32%	80.46%	56.61%	85.22%	74.15%
NB	64.46%	93.30%	81.64%	58.86%	93.78%	80.27%
LR	60.65%	92.16%	79.42%	57.29%	85.64%	74.67%
BPN	63.90%	92.02%	80.60%	61.66%	85.57%	76.32%
SVM	52.13%	89.21%	74.35%	52.13%	84.87%	72.20%
RF	57.37%	86.43%	74.86%	56.73%	70.93%	65.44%
DBN	44.61	88.04	71.65%	50.56%	85.01%	71.68%
DBN-DS (proposed)	65.13%	94.29%	82.60%	61.77%	93.56%	81.27%

high number and the two classifiers predict a low number for OCD, then the NOCD is finally predicted based on the belief value of the DS algorithm.

Next, the DBN-DS was evaluated. The result of the confusion matrix for DBN-DS is shown in Table 6. In addition, the results of the ROC curve analysis are shown in Figure 4 and Table 7. DBN#1 learned the initial PSA. DBN#2 learned the Gleason score, while DBN#3 learned the clinical T stage.

Among the three variables, the initial PSA level had the highest prediction rate. The PSA level is closely related to pathologic stage and is the most important parameter in prostate cancer. Variables combined with PSA showed a high prediction rate. In other words, the reason for the high prediction rate was that the Gleason score and clinical T stage also affect the pathology. However, the combination of Gleason score and clinical T stage had a lower accuracy than that predicted by the initial PSA level alone. The two variables are uncertain because they are diagnosed according to the doctor's experience. However, when combined with PSA level, the performance was much higher. In this study, we found that initial PSA was the most important predictor, and that the Gleason score and clinical T stage were also important predictors.

#### 4. Discussion and Conclusion

Prediction models for pathology staging of prostate cancer are based on clinical tests and can be used to predict the spread of cancer. It is possible to diagnose cancer more precisely at the postoperative, pathological stage and to determine the degree of metastasis of prostate cancer.

We proposed a DBN-DS-based multiclassifier approach to predict the pathologic stage of prostate cancer. The proposed method provides a predictive model to improve accuracy through deep learning and information fusion based on the relationship between data measured using clinical tests. The inputs include initial PSA level, Gleason scores, and clinical T stage variables. The output can be OCD or NOCD in pathological staging (pT). This approach was evaluated using an existing validated patient dataset that included 6345 patient records from the KPCR database, which collected data from six tertiary medical institutions.

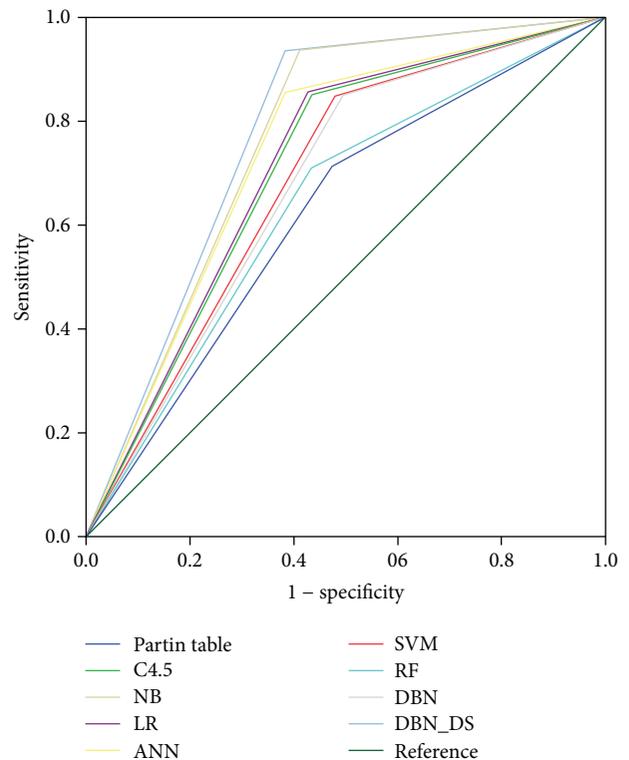


FIGURE 3: ROC curve results of all classification methods using the validation set.

The performance of the proposed DBN-DS was compared with that of the NB, LR, BPN, SVM, RF, DBN, and Partin tables. The results showed that the proposed DBN-DS had better sensitivity and accuracy than all other methods.

In a recent pathological staging methodology study, Cosma et al. [4] use a neuro-fuzzy model, with an approach similar to ours. The results also indicated that the neural network-fuzzy-based computational intelligence learning approach is suitable for prostate cancer staging and exceeds the performance of the Partin tables. The neuro-fuzzy model and our proposed method aim to predict whether a patient has OCD (pT2) or NOCD (pT3+). All methods use the initial PSA level, Gleason scores, and clinical T stage to predict the pathologic stage of prostate cancer, but the

TABLE 5: Results of a DBN-DS confusion matrix comparing the training and validation sets.

Variable	Training set			Validation set		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
DBN#1	38.57%	91.39%	70.78%	41.93%	88.68%	70.60%
DBN#2	32.51%	89.18%	67.26%	37.00%	88.47%	68.56%
DBN#3	21.19%	94.20%	65.96%	26.23%	93.85%	67.69%
DBN#1, DBN#2	41.48%	93.71%	73.50%	41.48%	93.00%	73.07%
DBN#1, DBN#3	40.02%	94.55%	73.46%	40.02%	93.85%	73.03%
DBN#2, DBN#3	34.19%	94.91%	71.42%	34.19%	93.49%	70.56%
DBN#1, DBN#2, DBN#3 (proposed)	65.13%	94.29%	82.60%	61.77%	93.56%	81.27%

TABLE 6: Detailed ROC curve analysis results of all classification methods using the validation set.

	AUC	<i>p</i> value	95% confidence interval	
			Lower bound	Upper bound
Partin table	0.620 ± 0.012	0.000	0.597	0.644
C4.5	0.709 ± 0.012	0.000	0.686	0.731
NB	0.763 ± 0.011	0.000	0.741	0.785
LR	0.715 ± 0.012	0.000	0.692	0.737
ANN	0.736 ± 0.012	0.000	0.714	0.758
SVM	0.685 ± 0.012	0.000	0.662	0.708
RF	0.638 ± 0.012	0.000	0.615	0.662
DBN	0.678 ± 0.012	0.000	0.655	0.701
DBN-DS	0.777 ± 0.011	0.000	0.755	0.798

TABLE 7: Detailed ROC curve result of DBN-DS using validation set.

	AUC	<i>p</i> value	95% confidence interval	
			Lower bound	Upper bound
DBN#1	0.653 ± 0.012	0.000	0.629	0.677
DBN#2	0.627 ± 0.012	0.000	0.603	0.651
DBN#3	0.600 ± 0.012	0.000	0.576	0.625
DBN#1, DBN#2	0.672 ± 0.012	0.000	0.649	0.696
DBN#1, DBN#3	0.669 ± 0.012	0.000	0.646	0.693
DBN#2, DBN#3	0.638 ± 0.012	0.000	0.614	0.663
DBN#1, DBN#2, DBN#3 (proposed)	0.777 ± 0.011	0.000	0.755	0.798

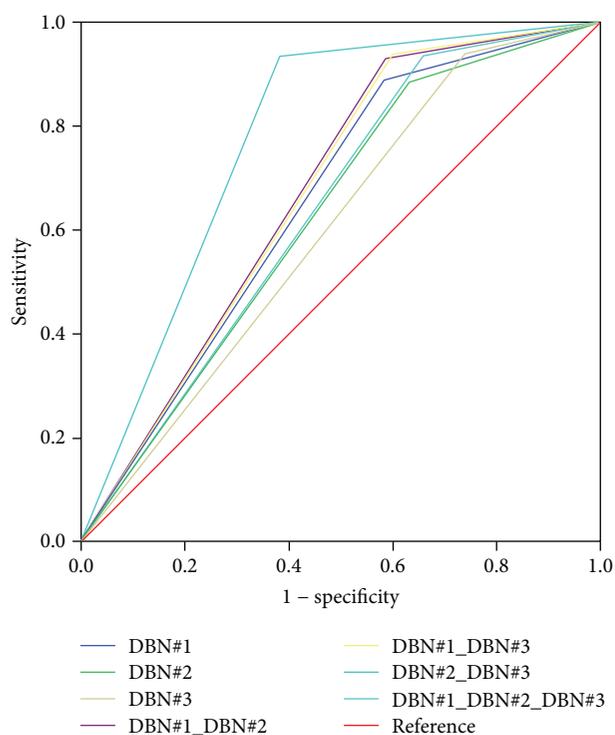


FIGURE 4: ROC curve results of DBN-DS using a validation set.

DBN-DS predicts more patient data than other studies. In addition, it is possible to learn more deeply through the DBN-DS in order to improve the prediction performance in the existing DBN. The neuro-fuzzy model obtained an area under the curve (AUC) of 0.812, while the nomogram of the AJCC achieved an AUC of 0.582. Our proposed DBN-DS achieved an AUC of 0.777, compared to 0.620 for the Partin tables. This result is similar to that reported by Cosma et al. [4], although different data sets were used for each study; however, they show a high consistency with the results of the present study.

Currently, the proposed DBN-DS method is implemented as a research tool. Once the clinical evaluation is completed, the proposed tool will be developed as an easy-to-use clinical decision support system that can be accessed by clinicians.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (NRF-2016R1A2B4015922).

## References

- [1] "International Agency for Research on Cancer—GLOBOCAN," 2008, December 2017, <http://globocan.iarc.fr/>.
- [2] C. Pound, A. W. Partin, M. A. Eisenberger, D. W. Chan, J. D. Pearson, and P. C. Walsh, "Natural history of progression after PSA elevation following radical prostatectomy," *JAMA*, vol. 281, no. 17, pp. 1591–1597, 1999.
- [3] J. E. Oesterling, C. B. Brendler, J. I. Epstein, A. W. Kimball Jr, and P. C. Walsh, "Correlation of clinical stage, serum prostatic acid phosphatase and preoperative Gleason grade with final pathological stage in 275 patients with clinically localized adenocarcinoma of the prostate," *The Journal of Urology*, vol. 138, no. 1, pp. 92–98, 1987.
- [4] G. Cosma, G. Acampora, D. Brown, R. C. Rees, M. Khan, and A. G. Pockley, "Prediction of pathological stage in patients with prostate cancer: a neuro-fuzzy model," *PLoS One*, vol. 11, no. 6, article e0155856, 2016.
- [5] Y. Matsui, S. Egawa, C. Tsukayama et al., "Artificial neural network analysis for predicting pathological stage of clinically localized prostate cancer in the Japanese population," *Japanese Journal of Clinical Oncology*, vol. 32, no. 12, pp. 530–535, 2002.
- [6] J. Epstein, P. Walsh, M. Carmichael, and C. Brendler, "Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1 c) prostate cancer," *JAMA*, vol. 271, no. 5, pp. 368–374, 1994.
- [7] M. L. Blute, O. Nativ, H. Zincke, G. M. Farrow, T. Therneau, and M. M. Lieber, "Pattern of failure after radical retropubic prostatectomy for clinically and pathologically localized adenocarcinoma of the prostate: influence of tumor deoxyribonucleic acid ploidy," *The Journal of Urology*, vol. 142, no. 5, pp. 1262–1265, 1989.
- [8] J. I. Epstein, G. Pizov, and P. C. Walsh, "Correlation of pathologic findings with progression after radical retropubic prostatectomy," *Cancer*, vol. 71, no. 11, pp. 3582–3593, 1993.
- [9] A. W. Partin, M. W. Kattan, E. N. Subong et al., "Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer: a multi-institutional update," *JAMA*, vol. 277, no. 18, pp. 1445–1451, 1997.
- [10] D. V. Makarov, B. J. Trock, E. B. Humphreys et al., "Updated nomogram to predict pathologic stage of prostate cancer given prostate-specific antigen level, clinical stage, and biopsy Gleason score (Partin tables) based on cases from 2000 to 2005," *Urology*, vol. 69, no. 6, pp. 1095–1101, 2007.
- [11] C. W. Tsao, C. Y. Liu, T. L. Cha et al., "Artificial neural network for predicting pathological stage of clinically localized prostate cancer in a Taiwanese population," *Journal of the Chinese Medical Association*, vol. 77, no. 10, pp. 513–518, 2014.
- [12] A. Tewari, C. Porter, J. Peabody et al., "Predictive modeling techniques in prostate cancer," *Molecular Urology*, vol. 5, no. 4, pp. 147–152, 2001.
- [13] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: a review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.
- [14] R. Salakhutdinov and G. E. Hinton, "Deep Boltzmann machines," in *Proceedings of the twelfth international conference on artificial intelligence and statistics, AIS-TATS 2009*, pp. 448–455, Clearwater Beach, FL, USA, 2009.
- [15] L. Honglak, G. Roger, R. Rajesh, and Y. N. Andrew, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, pp. 609–616, Montreal, QC, Canada, 2009.
- [16] H. Z. Wang, G. B. Wang, G. Q. Li, J. C. Peng, and Y. T. Liu, "Deep belief network based deterministic and probabilistic wind speed forecasting approach," *Applied Energy*, vol. 182, pp. 80–93, 2016.
- [17] R. R. Yager, "On the Dempster-Shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93–137, 1987.
- [18] V. Khatibi and G. A. Montazer, "A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8536–8542, 2010.
- [19] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The Annals of Mathematical Statistics*, vol. 38, no. 2, pp. 325–339, 1967.
- [20] G. Shafer and R. Logan, "Implementing Dempster's rule for hierarchical evidence," *Artificial Intelligence*, vol. 33, no. 3, pp. 271–298, 1987.
- [21] H. A. Moghaddam and S. Chodratnama, "Toward semantic content-based image retrieval using Dempster-Shafer theory in multi-label classification framework," *International Journal of Multimedia Information Retrieval*, vol. 6, no. 4, pp. 317–326, 2017.
- [22] J. Y. Lee, "Clinical research using smart prostate cancer database system (SPC-DB)," *Translational Andrology and Urology*, vol. 3, p. AB18, 2014.
- [23] S. B. Edge and C. C. Compton, "The American Joint Committee on Cancer: the 7th edition of the *AJCC Cancer Staging Manual* and the future of TNM," *Annals of Surgical Oncology*, vol. 17, no. 6, pp. 1471–1474, 2010.
- [24] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [25] M. A. Salama, A. E. Hassanien, and A. A. Fahmy, "Deep belief network for clustering and classification of a continuous data," in *The 10th IEEE International Symposium on Signal Processing and Information Technology*, pp. 473–477, Luxor, Egypt, 2010.
- [26] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [27] R. Ranawana and V. Palade, "Multi-classifier systems: review and a roadmap for developers," *International Journal of Hybrid Intelligent Systems*, vol. 3, no. 1, pp. 35–61, 2006.
- [28] A. Al-Ani and M. Deriche, "A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence," *Journal of Artificial Intelligence Research*, vol. 17, pp. 333–361, 2002.
- [29] G. Shafer, "The Dempster-Shafer theory," in *Encyclopedia of artificial intelligence (second Ed.)*, S. C. Shapiro, Ed., pp. 330–331, Wiley, New York, NY, USA, 1992.
- [30] O. Basir and X. Yuan, "Engine fault diagnosis based on multi-sensor information fusion using Dempster-Shafer evidence theory," *Information Fusion*, vol. 8, no. 4, pp. 379–386, 2007.
- [31] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote Sensing of Environment*, vol. 62, no. 1, pp. 77–89, 1997.
- [32] J. A. Swets, *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*, Psychology Press, New York, 2014.

## Research Article

# An Evolutionary Computation Approach for Optimizing Multilevel Data to Predict Patient Outcomes

Sean Barnes <sup>1</sup>, Suchi Saria,<sup>2</sup> and Scott Levin<sup>3</sup>

<sup>1</sup>Department of Decision, Operations & Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD, USA

<sup>2</sup>Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

<sup>3</sup>Department of Emergency Medicine, Department of Civil Engineering, Johns Hopkins University, Baltimore, MD, USA

Correspondence should be addressed to Sean Barnes; [sbarnes@rhsmith.umd.edu](mailto:sbarnes@rhsmith.umd.edu)

Received 26 August 2017; Accepted 31 January 2018; Published 18 March 2018

Academic Editor: Weide Chang

Copyright © 2018 Sean Barnes et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Widespread adoption of electronic health records (EHR) and objectives for meaningful use have increased opportunities for data-driven predictive applications in healthcare. These decision support applications are often fueled by large-scale, heterogeneous, and multilevel (i.e., defined at hierarchical levels of specificity) patient data that challenge the development of predictive models. Our objective is to develop and evaluate an approach for optimally specifying multilevel patient data for prediction problems. We present a general evolutionary computational framework to optimally specify multilevel data to predict individual patient outcomes. We evaluate this method for both flattening (single level) and retaining the hierarchical predictor structure (multiple levels) using data collected to predict critical outcomes for emergency department patients across five populations. We find that the performance of both the flattened and hierarchical predictor structures in predicting critical outcomes for emergency department patients improve upon the baseline models for which only a single level of predictor—either more general or more specific—is used ( $p < 0.001$ ). Our framework for optimizing the specificity of multilevel data improves upon more traditional single-level predictor structures and can readily be adapted to similar problems in healthcare and other domains.

## 1. Introduction

Rapid accumulation of electronic health record (EHR) data and emphasis on meaningful use of health information technology (HIT) [1] has given rise to many modeling applications that attempt to predict individual patient outcomes. The majority of these prognostic models target clinical outcomes (e.g., mortality, acute myocardial infarction, and septic shock); however, others aim at predicting service-oriented outcomes that span operations (e.g., wait times and length of stay), cost, quality, and patient satisfaction [2–10]. Regardless of outcome, these models aim at improving healthcare delivery by supporting provider and organizational decision-making.

EHRs are a valuable source of input data commonly leveraged for these predictive applications. However, the heterogeneity, large-scale nature, and variability in data entry create challenges with respect to how to optimally specify these data for predictive models. Multilevel data describing patients' clinical conditions and medical interventions are commonly hypothesized predictors available in EHRs, but present unique challenges for model specification.

Multilevel data describes individual patient characteristics at multiple levels of specificity (see Table 1). For example, the International Classification of Diseases (e.g., 9th Revision, Clinical Modification or ICD-9-CM) contains more than 14,000 diagnosis codes and 3900 procedure codes used to classify the conditions of patients and the services

TABLE 1: Common multilevel predictor data available in electronic health records.

Multilevel predictors	Description	Examples
Reasons for visit	Descriptors of the reason for the healthcare system encounter	Ambulatory care chief complaints; inpatient admission diagnoses
Diagnoses	Descriptors of patients' differential or final diagnosis departing the healthcare system	International classification of disease codes (e.g., ICD-10); read codes
Medical history	Descriptors of previous medical history and chronic conditions	EHR problem lists (e.g., diabetes, previous coronary artery bypass graft (CABG), hypertension)
Diagnostic and therapeutic procedures	Descriptors of diagnostic and therapeutic courses of action taken	Procedure coding system (ICD-10-PCS), surgical procedures, rehabilitation
Diagnostic exams	Descriptors of medical tests conducted	Laboratory exams, imaging exams, physical exams
Medication	Descriptors of medications administered	US Food and Drug Administration Drug Class (e.g., opioids and hydrocodone)
Administrative	Descriptor of the administrative status of patients	Inpatient, outpatient, observation

they receive [11]. Diagnoses and procedure codes have inherent hierarchical structure represented by digits and decimals. For example, ICD-9-CM code 038.12 may be deconstructed from the lowest-to-highest level of specificity in the following manner:

- (i) Level 1: 001–139 infectious and parasitic diseases
- (ii) Level 2: 030–041 other bacterial diseases
- (iii) Level 3: 038 septicemia
- (iv) Level 4: 038.1 staphylococcal septicemia
- (v) Level 5: 038.12 methicillin-resistant *Staphylococcus aureus* septicemia

Tools such as the U.S. Agency for Healthcare Research and Quality's Clinical Classifications Software (CCS) may similarly introduce their own conceptual structure [12]. Documentation of medical history and chronic conditions is also defined by a multilevel structure (see Table 1), for example, "diabetes" (low specificity) or type I, type II, or gestational diabetes (high specificity). Medications provide additional examples, for which definitions can be more general classes (e.g., antibiotics), more specific subclasses (e.g., penicillin), or somewhere in between (e.g., broad versus narrow spectrum).

Hypotheses may be generated about the level of specificity needed to best differentiate patients with respect to the outcome predicted. However, often, it is unclear which level will be most effective. Further, the optimal level of specificity may change for different outcomes or even the same outcome in different populations. For example, there is a substantial body of work involving the prediction of readmission for patients who undergo coronary artery bypass graft (CABG) surgery [13–16]. In much of this work, there are risk factors for comorbidities, medications, and complications that could be defined more generally or more specifically, and minimal rationale was provided about how these modeling decisions affected model performance. In addition, the optimal levels of specification for these risk factors for predicting readmission rates for CABG patients may not translate to predicting a different outcome such as mortality.

In many cases, the specification of multilevel data is hypothesis driven, in that an initial judgment on the appropriate level of specificity is made and that specification is retained throughout the modeling process. We propose a framework for learning the appropriate level(s) of specificity from data, and we evaluate the trade-offs of flattening or retaining the hierarchical structure of these multilevel predictor data. In the first case (i.e., flattening), general and specific categories are collapsed into a single mutually exclusive level. Patients are initially placed in their most general category, and then patients with indications for more specific categories are extracted from their general categories. In this case, there is a fundamental change in the structure of the multilevel data, as patients with the same general category are now distinct from one another (i.e., some patients in the general category will retain that category, while others will convert to a more specific category). This redefinition differentiates this problem from a simple feature selection problem, whereby categories that contribute to the predictive performance (with respect to the desired outcome) are retained and others are excluded. In the case for which the hierarchical structure is retained, patients with indications for more specific categories will also retain indications for their general category.

There has been previous research focused on modeling with multilevel data structures, particularly in the areas of political science, psychology, sociology, public health, and education [17–23]. In this work, the notion of multilevel data relates to predictors that are collected at multiple hierarchical levels, for example, at individual and group levels (e.g., class, school, department, organization, and district). For example, Burstein [22] proposes a structure in which background, educational process, and outcome variables are measured at the individual (i.e., student) and group (e.g., community, school, and district) levels. Similar types of research exists in the healthcare space, with much of it falling within the health service research subfield [24–28]. For example, Sjetne et al. [28] developed a model to explain the variation in patient satisfaction (measured as percentage ratings across 10 categories) as a function of both individual patient (e.g., age, gender, education level, and length of stay) and hospital (size and teaching status) characteristics. The bulk of the

existing research on multilevel data follows this approach and is inherently different from the problem that we present. In our approach, we focus only on data specified at the individual (patient) level, albeit at varying levels of specificity.

In the computer science field, there have been some recent works that are more closely related [29, 30]. Schulam and Saria [29] developed a learning framework to predict clinical trajectories using information measured at multiple levels of specificity (i.e., population, subpopulation, and individual). This general approach is similar to the aforementioned research, but the key difference is that their proposed method learns the relative importance of each level of the hierarchical structure, based on its ability to predict the desired outcome. In Choi et al. [30], the authors develop a graph-based attention model (GRAM) that leverages an existing hierarchical system (such as ICD or CCS) to predict diagnosis and heart failure outcomes. The attention mechanism primarily balanced the need for specificity of information with the observed sample size of that predictor in the training data. This approach was designed to address a specific limitation of deep learning models (in healthcare) that typically lack the requisite sample size for accurate training. Overall, our objective is similar in that we develop a learning framework for adapting hierarchical data structures for individual patient predictions, and this previous work underscores the need to develop such methods. However, we believe that our approach is more easily applied and more flexible and preserves the hierarchical predictors for interpretation by practitioners.

In the next section of this article, we define the general evolutionary computation (EC) framework. Then, we demonstrate the performance of this approach in predicting critical outcomes for emergency patients across five patient populations. After that, we discuss the implications of this approach and how it can be applied more broadly. Lastly, we conclude with some final thoughts and some proposals for future development.

## 2. Methods

We present a general EC framework for optimizing multilevel data for predictive modeling. This framework is suitable for both classification and regression problems. First, we introduce the reader to a case study of predicting critical outcomes for emergency department patients, which provides a specific context for which to present the framework. Then, we describe the framework itself, which can be readily adapted to other applications within healthcare and other domains.

*2.1. Case Study: Predicting Critical Outcomes for Emergency Department Patients.* Emergency Departments (EDs) have experienced a surge of patient volume to over 136 million visits annually in the United States (US) [31]. This has exacerbated the ED crowding crisis and places patients at undue risk of adverse events associated with delays in care [32, 33]. EDs are required to see all comers, thus patients must be quickly evaluated at presentation to determine the urgency of care needs. This process is called triage and has

standards in place that require the provider to record the patient's demographics (age, gender), elicit a chief complaint (i.e., reason for visit), and measure vital signs (heart rate, respiratory rate, temperature, blood pressure, and oxygen saturation). Triage standards in the US require clinicians to apply the Emergency Severity Index (ESI), an algorithm used to assign patients to a 5-level scale from 1 (high severity; need for immediate treatment) to 5 (low severity; nonurgent) [34]. ESI relies heavily on provider judgment, is subject to high variation [35], and poorly differentiates a large majority group (ESI level 3), counter to the true objective of the triage [36, 37].

Thus, an alternative, outcome-based approach for conducting triage has been developed and is being used in several EDs in the US [37, 38]. A key component of this data-driven approach involves predicting critical care events for ED patients based on the information collected at presentation. Here, we define a critical care event as a composite and binary outcome that includes in-hospital mortality, direct admission to a hospital intensive care unit, or emergent surgery or catheterization for the same patient stay. These outcomes are analogous to the types of outcomes that would require immediate action on the part of care providers when the patient arrives in the ED and correspond to the most urgent ESI levels (i.e., one and two). This critical care event is the outcome that we aim to predict with our model.

In this study, we apply our EC framework to optimize multilevel predictors—specifically chief complaints—for predicting critical care events for ED patients. This prediction model utilizes the same information that is collected for the traditional triage process and includes the age, gender, and arrival mode of the patient, along with the aforementioned vital signs and the chief complaints that will be optimized using our EC framework. The vital signs were (nonuniformly) discretized into clinically meaningful categories, including a dedicated category for missing information [37, 38]. We summarize the categorical predictor variables in Table 2.

We apply our method across five patient populations, including a large, urban academic medical center (ACAD), a medium-sized community hospital (COMM), international hospitals in Brazil (BRAZIL) and the United Arab Emirates (UAE), and the nationally representative National Hospital Ambulatory Medical Care Survey (NAT). We provide summary characteristics of these five patient populations in Table 3.

*2.2. General Evolutionary Computation Framework.* Evolutionary computation is a class of metaheuristic algorithms that mimic biological processes to solve difficult optimization problems [39]. Relative to exact algorithms, evolutionary algorithms are stochastic and are not guaranteed to find global optima; however, they work well in practice and can provide good solutions within manageable computation times. In addition, evolutionary algorithms provide a flexible framework that can be readily adapted to different types of problems or variations of similar problems.

Specifically, we utilize a genetic algorithm (GA) to search for the optimal combination of complaints and complaint

TABLE 2: Summary of categorical predictor variables (abnormal ranges indicated in bold).

Predictor	Categories	Ranges/categories
Age	8	18–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, >90
Gender	2	Male, female
Arrival mode	2	Via ambulance, walk in
Temperature (°F)*	6	< <b>94.8</b> , 94.8–96.1, 96.1–99.2, 99.2–100.4, > <b>100.4</b>
Pulse (bpm)*	8	< <b>49</b> , 49–59, 59–105, 105–109, 109–119, <b>119–129</b> , > <b>129</b>
Respiratory rate (bpm)*	6	< <b>13</b> , 13–14, 14–19, 19–23, > <b>23</b>
Blood pressure (mmHG)*	6	< <b>99</b> , 99–106, 106–176, 176–199, > <b>199</b>
Oxygen saturation (%)*	4	< <b>93</b> , 93–94, >94

\*Each vital sign also includes an additional category for missing data.

TABLE 3: Patient population summary.

	ACAD	COMM	BRAZIL	UAE	NAT
Sample size	104.5 K	144.9 K	94.8 K	103.5 K	74.6 K
Unique complaints	686	616	358	288	649
Critical outcome prevalence	3.45%	3.48%	3.00%	1.68%	3.05%

categories, for which the complaints represent more specific information on each patient’s reason for visit and the complaint categories combine specific complaints into clinically meaningful groups. GAs imitate the process of natural selection, whereby stronger candidate solutions survive and weaker candidate solutions are eliminated [40]. We implemented our GA using the distributed evolutionary algorithms in Python package [41].

For this application of a GA, candidate solutions in the population are represented by binary bit strings of length  $n$ —where  $n$  represents the number of specific complaints—for which each bit  $b_i$  represents whether a specific complaint is excluded (0) or selected (1) as a predictor in the classification model for the critical care outcome in ED patients. We include all aforementioned age, gender, arrival mode, and complaint categories in the prediction model and therefore do not need to include them in the search process. The population contains  $N$  candidate solutions, each of which is initialized with randomly generated 0 and 1 values (i.e., a random selection of specific complaints). For each generation, a subset of the population is selected via a tournament selection scheme for crossover operations. Uniform crossover is ideal for this application (as opposed to other common crossover operations such as single- or multipoint crossover) because there is minimal advantage in preserving contiguous blocks of chromosomes (i.e., each selected complaint is essentially independent from the others). Once crossover is completed, a subset of candidate solutions in the new generation is selected for mutation. We utilize a simple bit flip operation for mutation, which inverts a subset of complaint bits within each candidate solution. For example, complaints selected for mutation that are currently excluded become selected, and complaints selected for mutation that are currently selected become excluded. We summarize the representation of candidate solutions and the crossover and mutation operations in Figure 1. Control parameters for the

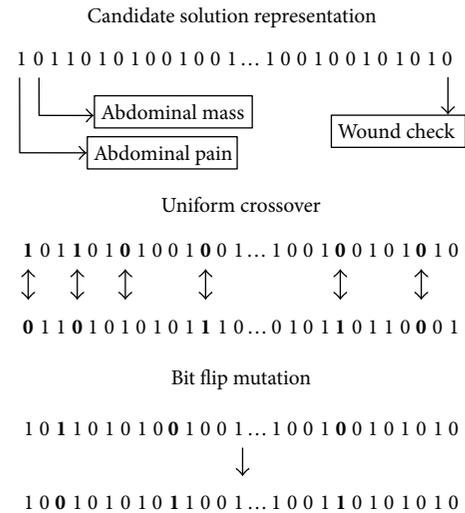


FIGURE 1: Genetic algorithm representation and recombination operators.

GA (summarized in Table 4) were selected via experimentation to maintain the diversity of the population and prevent premature convergence toward a suboptimal solution. In general, there is evidence that a broad range of control parameters leads to good performance [42]; therefore, it was determined that comprehensive experimentation with these parameters would add little value and be computationally prohibitive for this application.

We model the fitness of each candidate solution using 5-fold cross-validated area under the receiver operating characteristics curve (AUC, also commonly referred to as the C statistic), which is a standard measure of predictive performance for classification models [43]. We use logistic regression for the classification estimator for two reasons.

TABLE 4: Summary table of genetic algorithm control parameters and operators.

Parameter	Setting
Population size ( $N$ )	40
Number of generations	100
Selection	Tournament ( $k = 3$ )
Crossover operation	Uniform
Crossover rate	0.6
Mixing ratio	0.2
Mutation operation	Bit flip
Mutation rate	0.2
Bit flip rate	0.05

First, logistic regression is a deterministic algorithm and therefore does not confound the performance of the GA as would a stochastic ensemble approach such as a random forest or boosting algorithm. Second, logistic regression is computationally efficient and therefore allows the GA to explore more generations of candidate solutions for a fixed computation budget.

The specific calculation of fitness depends on the modeling approach for the multilevel data. For the flattening approach, patients only have a positive indication for either a selected complaint or a selected complaint category. Patients with a selected complaint are removed from their corresponding complaint category before the classification model is trained. For example, suppose a complaint for abdominal cramping is selected, which belongs to the more general abdominal pain category. Therefore, patients with the specific abdominal cramping complaint will be removed from the more general abdominal pain category. Patients with complaints that are not selected (e.g., abdominal mass in Figure 1) retain positive indications for the corresponding complaint category (i.e., abdominal pain for this example). This structure maintains a single, mutually exclusive, level for the chief complaint predictor. By contrast, the hierarchical approach retains positive indications for the corresponding complaint category regardless of whether a specific complaint is selected or not. For example, patients with abdominal cramping will have positive indications for both the specific complaint and the corresponding complaint category (abdominal pain). Once the chief complaint specification has been updated for each candidate solution (based on the selected complaints), we calculate the 5-fold cross-validated AUC for the critical care outcome using logistic regression as the classification algorithm and the age, gender, arrival mode, complaint categories, and selected complaints as predictors. The top  $N$  candidate solutions with respect to fitness are retained for the next generation, and the process terminates when it reaches the prespecified number of generations.

**2.3. Model Evaluation.** We run our GA using the flattened and hierarchical fitness functions for each of the five patient populations and compare the performance of the best-found solutions with two baseline models. The first

baseline model only includes the specific complaints for the classification model, whereas the second baseline model only uses complaint categories. We utilize DeLong’s method to evaluate the statistical differences in fitness function values between our EC approach and the baseline models [44]. In addition, we evaluate the performance of each model for specific subgroups of patients using a bullseye analogy, in order to characterize any performance differences across relevant subsets of the population. We define the inner region as patients who are directly affected because their specific complaint is selected by the GA. The middle region contains patients who are indirectly affected by a change in their complaint category. Although their specific complaint is not selected, the composition of their complaint category is altered because some patients within the complaint category are treated differently. Finally, the outer region contains patients with no direct connection to patients with selected complaints and is only affected by the overall classification model. We also compare differences in the predicted probabilities for each subgroup between the GA and the baseline models.

In addition to overall model performance, we explore the selected and excluded complaints themselves, which can provide valuable insight as to which complaints are meaningful in this specific context. An advantage of an EC approach is that each candidate solution—and particularly the strongest candidate solutions—provides feedback about the importance of specific complaints. We compare the selected complaints between the flattened and hierarchical approaches for a given population, and we also attempt to draw comparisons across the five populations.

### 3. Results

We first present detailed results for the academic hospital and then summarize the results for the other ED populations. In Figures 2 and 3, we summarize the bullseye performance for the flattened and hierarchical approaches, respectively, relative to the two baseline models. We note here that separate figures are required for the comparison due to the distinct selection of complaints by each approach and therefore distinct specifications of the inner, middle, and outer subpopulations.

Overall, both GA approaches demonstrate a statistically significant improvement in the overall 5-fold cross-validated AUC relative to the baseline models ( $p < 0.001$  for the both cases), so there is a benefit to including both specific and categorized complaint information for this application. In addition, statistically significant improvements were observed for all subgroups relative to the baseline model with complaints only and for the inner and middle subgroups relative to the baseline model with categorized complaints only. These results suggest that the GAs achieved improvements for multiple subgroups in the population without sacrificing the model’s performance on other subgroups.

In Figure 4, we summarize the differences in predicted probabilities for the hierarchical approach relative to the baseline models. The results for the flattened approach are very similar (not shown). One notable difference is that the

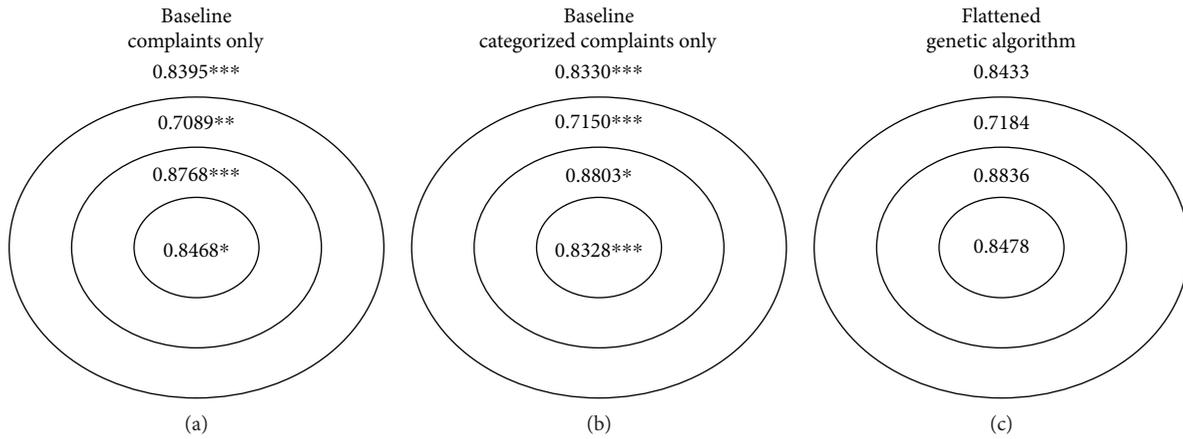


FIGURE 2: Bullseye performance for baseline models (with specific complaints only and complaint categories only, resp.) and flattened genetic algorithm for the academic hospital. Overall performance is indicated outside of the bullseye. Statistical significance for the difference in 5-fold cross-validated AUC (using DeLong’s method) between the flattened genetic algorithm approach and the corresponding baseline models is indicated by \*\*\* for  $p < 0.001$ , \*\* for  $p < 0.01$ , and \* for  $p < 0.05$ .

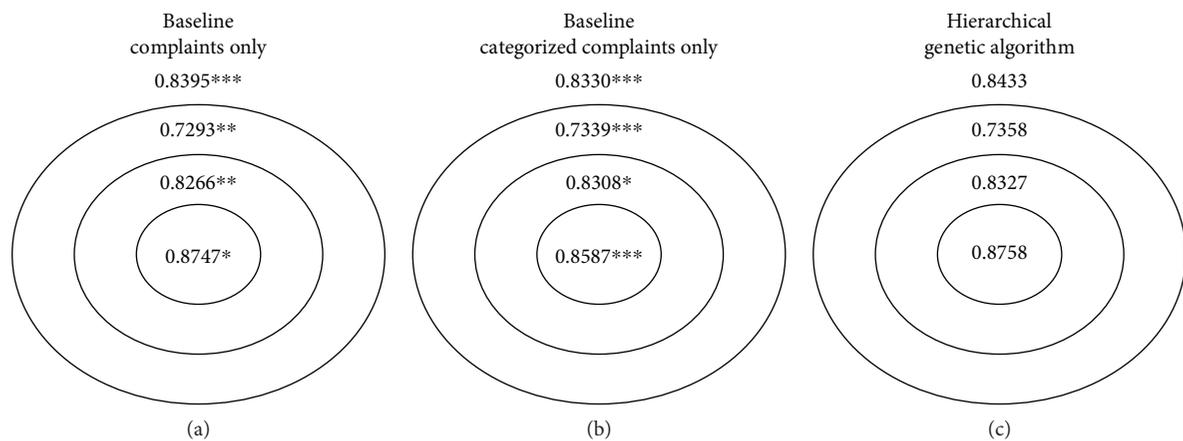


FIGURE 3: Bullseye performance for baseline models (with specific complaints only and complaint categories only, resp.) and hierarchical genetic algorithm for the academic hospital. Overall performance is indicated outside of the bullseye. Statistical significance for the difference in 5-fold cross-validated AUC (using DeLong’s method) between the flattened genetic algorithm approach and the corresponding baseline models is indicated by \*\*\* for  $p < 0.001$ , \*\* for  $p < 0.01$ , and \* for  $p < 0.05$ .

predicted probabilities for patients in the inner subgroup are frequently adjusted relative to the baseline model with categorized complaints only. These adjustments are the direct effect of including more specific information (i.e., complaints) in addition to the categorized complaints. Therefore, some patients are shifted toward being a higher risk of a critical care outcome, and others are shifted toward a lower risk, depending on their specific (rather than categorized) complaint. The other notable difference is the significant frequency of adjustments to predicted probabilities for the middle subgroup of patients relative to the baseline model with specific complaint information. The predicted probabilities for these patients are adjusted by augmenting specific complaint information with categorized complaints. Minimal changes are made to the predicted probabilities for the outer subgroups that are not directly affected by the selection of specific complaints.

The performances of the flattened and hierarchical approaches on the four other patient populations were quite similar to their performances on the academic hospital patient population. Specifically, both approaches achieved a statistically significant improvement in overall 5-fold cross-validated AUC relative to the baseline models. In addition, both approaches consistently achieved statistically significant improvements relative to the baseline model with categorized complaints only for the inner bullseye subgroup and relative to the baseline model with complaints only for the middle bullseye subgroups (i.e., the subgroups most directly affected by the EC approach). Statistically significant improvements were observed for other bullseye subgroups, but these improvements were not consistent across all populations. Finally, there were more adjustments for predicted probabilities relative to the baseline model with categorized complaints only than

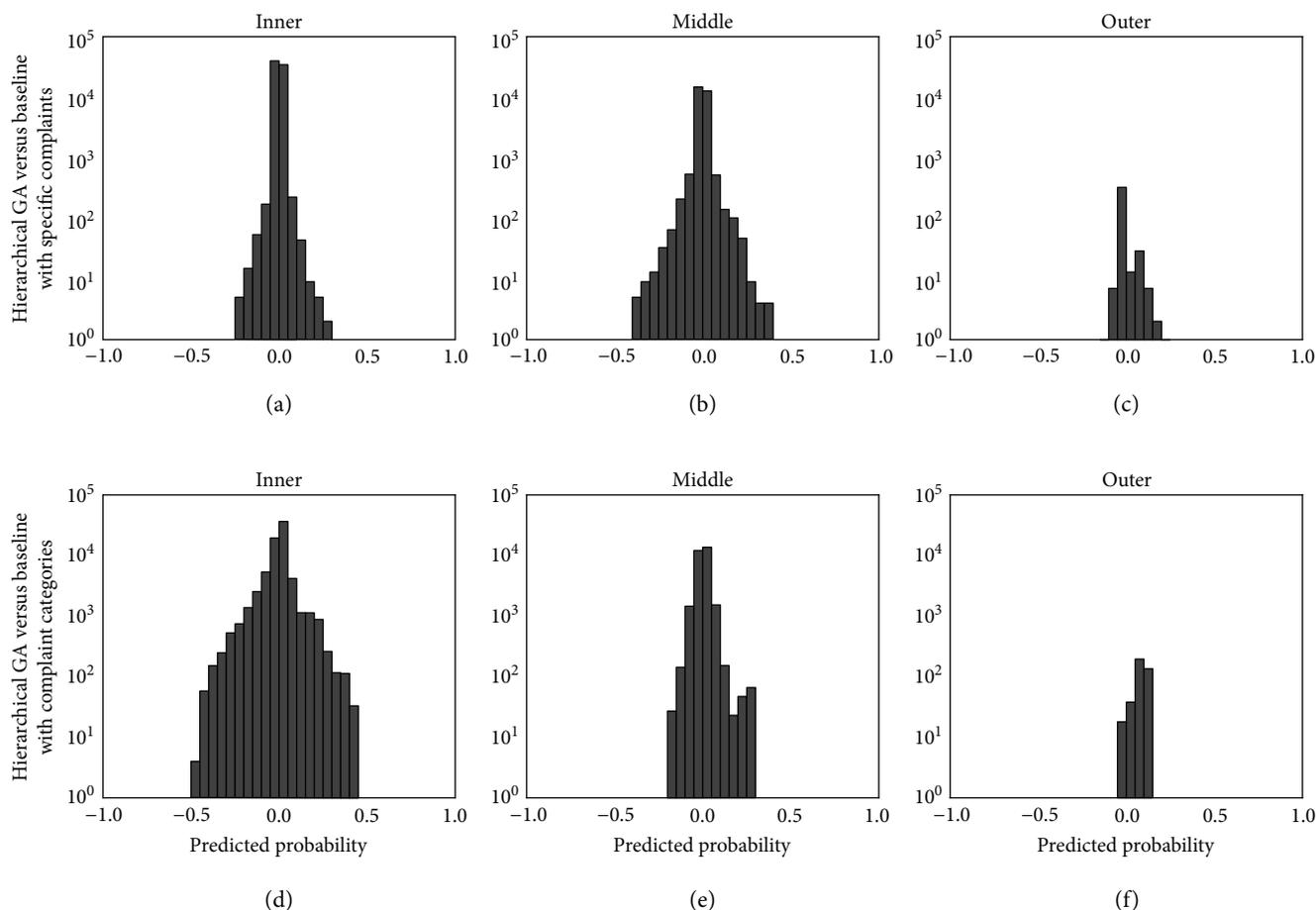


FIGURE 4: Histogram summary of differences in predicted probabilities of the hierarchical approach relative to baseline models. (a, b, c) Baseline model with complaints only. (d, e, f) Baseline model with categorized complaints only for the inner (a, d), middle (b, e), and outer subgroups (c, f) of patients. Note that the y-axis displays the frequency (count) of patients on a logarithmic scale.

the baseline model with complaints only, particularly for the inner bullseye subgroup.

Overall, there is minimal difference between the performance of the flattened and hierarchical approaches, and there is no significant difference across any of the five populations (see Table 5). In addition, there are similar effects on predicted probabilities (as in Figure 4), in that the most substantial effects are differences for the inner subgroup relative to the baseline model with categorized complaints only and for the middle subgroup relative to the baseline model with complaints only.

Despite the similarities in performance, there are some key differences between the two approaches. Training times are much faster for the hierarchical approach (see Table 5), as there is no restructuring of the multilevel data as for the flattened approach. On the other hand, the flattened approach has the advantage of reducing the dimensionality of the data into a single level, which could improve run times once the multilevel structure has been reduced into a flattened format. We note, however, that prediction times using either trained model would be very fast. Finally, there is significant disagreement among the

selected complaints for a given population (see Table 5). In general, the two approaches only agree on approximately 55–60% of the complaints to either exclude or select in the predictive model for critical outcomes for ED patients. The remaining complaints were uniquely selected by only one approach.

#### 4. Discussion

This EC approach demonstrates a statistically significant improvement over single-level models that use only complaint or categorized complaint information. These improvements are significant not only for the overall population, but for directly affected subgroups within the population without sacrificing performance on others. It is important to note that these improvements, although seemingly small in magnitude, would have a significant impact over the large volume of patients who visit the ED. Similar (in magnitude) improvements were observed in previous work relative to their selected baseline models [29, 30], although their performance was only evaluated at the overall (not the subgroup) level.

TABLE 5: Comparison of results generated from flattened and hierarchical approaches across five patient populations.

(a)

	ACAD	COMM	Flattened BRAZIL	UAE	NHAMCS
Overall AUC	0.8431	0.8361	0.8261	0.8820	0.8429
Training time (hr)	42.47	78.67	19.89	15.00	29.06
Selected complaints (%)	48.3	52.8	53.4	59.0	49.9

(b)

	ACAD	COMM	Hierarchical BRAZIL	UAE	NHAMCS
Overall AUC	0.8433	0.8364	0.8260	0.8819	0.8436
Training time (hr)	4.93	8.91	3.46	3.27	3.09
Selected complaints (%)	49.3	64.6	55.6	55.6	46.4

(c)

	Comparison				
Difference in overall AUC ( $p$ value)	0.6144	0.2210	0.7022	0.3622	0.2579
Jointly selected complaints (%)	28.1	33.1	32.4	37.5	27.5
Jointly excluded complaints (%)	30.6	27.6	23.5	22.9	31.2

In addition to the performance improvements, this approach reduces the dimensionality of multilevel features. For the flattened approach, multilevel data is collapsed into a single mutually exclusive level. This is advantageous when population size may limit the number of predictor variables that can be meaningfully included. For the hierarchical approach, excluded complaints are pruned from the multilevel data structure. Once enforced, these reductions in dimensionality can facilitate faster development of prediction models, including algorithm selection, parameter tuning, cross-validation, testing, and prediction. The output from these feature selection approaches also provides practitioners with important feedback about the relevance of specific information in the context of a particular outcome. We believe that feature selection—as opposed to using an attention mechanism—has advantages in interpretation over previous approaches.

It is unclear whether the uniquely selected complaints are meaningful in the context of a specific complaint structure (i.e., flattened or hierarchical), or if they are simply insignificant artifacts of the stochastic GA. However, jointly selected complaints have strong support that they are meaningful for a particular outcome, and jointly unselected complaints have strong support that they are not meaningful. Potential improvements to this approach may involve a hybrid solution that leverages output from both approaches. For example, select complaints for the prediction model only if they are jointly selected by both approaches. Or alternatively, select complaints for the prediction model only if they are selected by at least one approach and they meet some minimum sample size requirement.

The stochastic nature of the evolutionary approach may raise questions about its reliability. However, the top candidate solutions for a given run consistently select the same complaints to exclude or include in the prediction model. Very few complaints (<10% for each population) are inconsistently excluded or selected in the prediction model, and for many of these cases, the complaints lean strongly toward being excluded or included (e.g., bladder pain was included in 19 of 20 of the top candidate solutions for the academic hospital).

## 5. Conclusion

In this study, we propose an EC framework for the specification of multilevel data for predictive models. This framework is easy to implement, leverages readily available open-source software, and can be adapted to optimize specification of multilevel data for many predictive applications. This includes the flexibility to accommodate other evolutionary algorithms (e.g., random mutation hill climbing and simulated annealing). The representation of candidate solutions (i.e., binary bit strings) would most likely be similar, and selection, crossover, and mutation operations (and associated control parameters) can be adjusted according to performance. Further, alternative fitness functions may be applied in place of the 5-fold cross-validated AUC. For example, a different cross-validation scheme (e.g., train-test split and stratified cross validation), estimator (e.g., classification tree and regression estimator), or performance measure (e.g., classification accuracy,  $R^2$ ) could readily be substituted into the framework. In addition, alternate types

of preprocessing—similar to the dynamic restructuring of multilevel data for the flattening approach—can be inserted prior to the computation of the fitness function, which is a noted advantage over other feature selection approaches.

We focus here on the specific application of specifying complaint information for predicting critical outcomes for ED patients; however, this approach is generalizable to many types of multilevel data within healthcare. For example, the other key component of the electronic triage algorithm requires prediction of admission outcomes for ED patients. We have applied this framework to this prediction problem as well, and the results are quite similar to those reported here for the critical care outcome.

## Disclosure

This work was presented at the INFORMS International Conference in 2016 [45]. It was an oral presentation.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

## Acknowledgments

The authors would like to acknowledge Sara de Ramirez (Johns Hopkins University), Paulo Schmitz (Hospital Moinhos de Vento), and Xavier Anton (Abu Dhabi Health Services Company) for their assistance in providing access to individual (and deidentified) patient data for this study.

## References

- [1] S. S. Jones, R. S. Rudin, T. Perry, and P. G. Shekelle, "Health information technology: an updated systematic review with a focus on meaningful use," *Annals of Internal Medicine*, vol. 160, no. 1, pp. 48–54, 2014.
- [2] K. C. Johnston, A. F. Connors, D. P. Wagner et al., "A predictive risk model for outcomes of ischemic stroke," *Stroke*, vol. 31, no. 2, pp. 448–455, 2000.
- [3] A. Green and S. Davis, "Toward a predictive model of patient satisfaction with nurse practitioner care," *Journal of the American Academy of Nurse Practitioners*, vol. 17, no. 4, pp. 139–148, 2005.
- [4] C. A. Powers, C. M. Meyer, M. C. Roebuck, and B. Vaziri, "Predictive modeling of total healthcare costs using pharmacy claims data: a comparison of alternative econometric cost modeling techniques," *Medical Care*, vol. 43, no. 11, pp. 1065–1072, 2005.
- [5] P. P. Tekkis, A. J. Senagore, and C. P. Delaney, "Conversion rates in laparoscopic colorectal surgery: a predictive model with 1253 patients," *Surgical Endoscopy*, vol. 19, no. 1, pp. 47–54, 2005.
- [6] S. Harbarth, H. Sax, I. Uckay et al., "A predictive model for identifying surgical patients at risk of methicillin-resistant *Staphylococcus aureus* carriage on admission," *Journal of the American College of Surgeons*, vol. 207, no. 5, pp. 683–689, 2008.
- [7] N. Tangri, L. A. Stevens, J. Griffith et al., "A predictive model for progression of chronic kidney disease to kidney failure," *JAMA*, vol. 305, no. 15, pp. 1553–1559, 2011.
- [8] J. Billings, I. Blunt, A. Steventon, T. Georghiou, G. Lewis, and M. Bardsley, "Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30)," *BMJ Open*, vol. 2, no. 4, article e001667, 2012.
- [9] S. Barnes, E. Hamrock, M. Toerper, S. Siddiqui, and S. Levin, "Real-time prediction of inpatient length of stay for discharge prioritization," *Journal of the American Medical Informatics Association*, vol. 23, no. e1, pp. e2–e10, 2016.
- [10] A. L. Hill, D. I. S. Rosenbloom, E. Goldstein et al., "Real-time predictions of reservoir size and rebound time during antiretroviral therapy interruption trials for HIV," *PLoS Pathogens*, vol. 12, no. 4, article e1005535, 2016.
- [11] World Health Organization, *ICD-9-CM: International Classification of Diseases, 9th Revision: Clinical Modification*, Medicode, Salt Lake City, UT, USA, 1998.
- [12] A. Elixhauser, C. Steiner, and L. Palmer, *Clinical Classifications Software (CCS)*, 2014, April 2016, <http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>.
- [13] M. Engoren, R. H. Habib, J. J. Dooner, and T. A. Schwann, "Use of genetic programming, logistic regression, and artificial neural nets to predict readmission after coronary artery bypass surgery," *Journal of Clinical Monitoring and Computing*, vol. 27, no. 4, pp. 455–464, 2013.
- [14] E. L. Hannan, Y. Zhong, S. J. Lahey et al., "30-day readmissions after coronary artery bypass graft surgery in New York state," *JACC: Cardiovascular Interventions*, vol. 4, no. 5, pp. 569–576, 2011.
- [15] J. D. Price, J. L. Romeiser, J. M. Gnerre, A. L. W. Shroyer, and T. K. Rosengart, "Risk analysis for readmission after coronary artery bypass surgery: developing a strategy to reduce readmissions," *Journal of the American College of Surgeons*, vol. 216, no. 3, pp. 412–419, 2013.
- [16] R. D. Stewart, C. T. Campos, B. Jennings, S. S. Lollis, S. Levitsky, and S. J. Lahey, "Predictors of 30-day hospital readmission after coronary artery bypass," *The Annals of Thoracic Surgery*, vol. 70, no. 1, pp. 169–174, 2000.
- [17] M. R. Steenbergen and B. S. Jones, "Modeling multilevel data structures," *American Journal of Political Science*, vol. 46, no. 1, pp. 218–237, 2002.
- [18] H. Goldstein and R. P. McDonald, "A general model for the analysis of multilevel data," *Psychometrika*, vol. 53, no. 4, pp. 455–467, 1988.
- [19] G. Affleck, A. Zautra, H. Tennen, and S. Armeli, "Multilevel daily process designs for consulting and clinical psychology: a preface for the perplexed," *Journal of Consulting and Clinical Psychology*, vol. 67, no. 5, pp. 746–754, 1999.
- [20] B. Muthén, "10. Latent variable modeling of longitudinal and multilevel data," *Sociological Methodology*, vol. 27, no. 1, pp. 453–480, 1997.
- [21] A. V. Diez-Roux, "Multilevel analysis in public health research," *Annual Review of Public Health*, vol. 21, no. 1, pp. 171–192, 2000.
- [22] L. Burstein, "Chapter 4: the analysis of multilevel data in educational research and evaluation," *Review of Research in Education*, vol. 8, no. 1, pp. 158–233, 1980.
- [23] J. B. Schreiber and B. W. Griffin, "Review of multilevel modeling and multilevel studies in *The Journal of Educational*

- Research (1992-2002),” *The Journal of Educational Research*, vol. 98, no. 1, pp. 24–34, 2004.
- [24] N. Rice and A. Leyland, “Multilevel models: applications to health data,” *Journal of Health Services Research & Policy*, vol. 1, no. 3, pp. 154–164, 1996.
- [25] A. Aiello, A. Garman, and S. B. Morris, “Patient satisfaction with nursing care: a multilevel analysis,” *Quality Management in Health Care*, vol. 12, no. 3, pp. 187–190, 2003.
- [26] A. Woods, “Multilevel modelling in primary care research,” *The British Journal of General Practice*, vol. 54, no. 504, pp. 560–561, 2004.
- [27] Y.-W. B. Wu and P. J. Wooldridge, “The impact of centering first-level predictors on individual and contextual effects in multilevel data analysis,” *Nursing Research*, vol. 54, no. 3, pp. 212–216, 2005.
- [28] I. S. Sjetne, M. Veenstra, and K. Stavem, “The effect of hospital size and teaching status on patient experiences with hospital care: a multilevel analysis,” *Medical Care*, vol. 45, no. 3, pp. 252–258, 2007.
- [29] P. Schulam and S. Saria, “A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure,” in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, vol. 1, pp. 748–756, MIT Press, Cambridge, MA, USA, 2015.
- [30] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun, “GRAM: graph-based attention model for healthcare representation learning,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 787–795, Halifax, NS, Canada, 2017.
- [31] U.S. Centers for Disease Control and Prevention, *National Hospital Ambulatory Medical Care Survey: Emergency Department Summary Tables*, 2011, October 2016, [http://www.cdc.gov/nchs/data/ahcd/nhamcs\\_emergency/2011\\_ed\\_web\\_tables.pdf](http://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2011_ed_web_tables.pdf).
- [32] *Hospital-Based Emergency Care: At the Breaking Point*. Institute of Medicine, December 2016, <http://www.nationalacademies.org/hmd/Reports/2006/Hospital-Based-Emergency-Care-At-the-Breaking-Point.aspx>.
- [33] S. L. Bernstein, D. Aronsky, R. Duseja et al., “The effect of emergency department crowding on clinically oriented outcomes,” *Academic Emergency Medicine*, vol. 16, no. 1, pp. 1–10, 2009.
- [34] N. Gilboy, P. Tanabe, D. Travers, and A. M. Rosenau, *Emergency Severity Index (ESI): A Triage Tool for Emergency Department Care, Version 4*, 2012, December 2016, <http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/systems/hospital/esi/esihandbk.pdf>.
- [35] M. Christ, F. Grossmann, D. Winter, R. Bingisser, and E. Platz, “Modern triage in the emergency department,” *Deutsches Ärzteblatt International*, vol. 107, no. 50, pp. 892–898, 2010.
- [36] R. Arya, G. Wei, J. V. McCoy, J. Crane, P. Ohman-Strickland, and R. M. Eisenstein, “Decreasing length of stay in the emergency department with a split emergency severity index 3 patient flow model,” *Academic Emergency Medicine*, vol. 20, no. 11, pp. 1171–1179, 2013.
- [37] A. F. Dugas, T. D. Kirsch, M. Toerper et al., “An electronic emergency triage system to improve patient distribution by critical outcomes,” *The Journal of Emergency Medicine*, vol. 50, no. 6, pp. 910–918, 2016.
- [38] S. Levin, M. Toerper, E. Hamrock et al., “Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index,” *Annals of Emergency Medicine*, 2017.
- [39] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Springer, 2003, April 2016, <http://link.springer.com/content/pdf/10.1007/978-3-662-44874-8.pdf>.
- [40] D. Whitley, “A genetic algorithm tutorial,” *Statistics and Computing*, vol. 4, no. 2, pp. 65–85, 1994.
- [41] F.-A. Fortin, F.-M. De Rainville, M.-A. Gardner, M. Parizeau, and C. Gagné, “DEAP: evolutionary algorithms made easy,” *Journal of Machine Learning Research*, vol. 13, pp. 2171–2175, 2012.
- [42] J. Grefenstette, “Optimization of control parameters for genetic algorithms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 16, no. 1, pp. 122–128, 1986.
- [43] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [44] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, “Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach,” *Biometrics*, vol. 44, no. 3, pp. 837–845, 1988.
- [45] S. Barnes, S. Saria, and S. Levin, “An evolutionary computation framework for optimizing multi-level data for predicting individual patient outcomes,” in *Presented at the 2016 INFORMS International Conference*, Hawaii, HI, USA, 2016.

## Research Article

# Outlier Removal in Model-Based Missing Value Imputation for Medical Datasets

Min-Wei Huang <sup>1,2</sup>, Wei-Chao Lin <sup>3,4</sup> and Chih-Fong Tsai <sup>5</sup>

<sup>1</sup>School of Medicine, China Medical University, Taichung, Taiwan

<sup>2</sup>Department of Psychiatry, Chiayi Branch, Taichung Veterans General Hospital, Chiayi, Taiwan

<sup>3</sup>Department of Information Management, Chang Gung University, Taoyuan, Taiwan

<sup>4</sup>Department of Thoracic Surgery, Chang Gung Memorial Hospital, Linkou, Taoyuan, Taiwan

<sup>5</sup>Department of Information Management, National Central University, Zhongli, Taoyuan, Taiwan

Correspondence should be addressed to Wei-Chao Lin; [viclin@gap.cgu.edu.tw](mailto:viclin@gap.cgu.edu.tw)

Received 11 July 2017; Revised 17 September 2017; Accepted 1 October 2017; Published 4 February 2018

Academic Editor: Weide Chang

Copyright © 2018 Min-Wei Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Many real-world medical datasets contain some proportion of missing (attribute) values. In general, missing value imputation can be performed to solve this problem, which is to provide estimations for the missing values by a reasoning process based on the (complete) observed data. However, if the observed data contain some noisy information or outliers, the estimations of the missing values may not be reliable or may even be quite different from the real values. The aim of this paper is to examine whether a combination of instance selection from the observed data and missing value imputation offers better performance than performing missing value imputation alone. In particular, three instance selection algorithms, DROP3, GA, and IB3, and three imputation algorithms, KNNI, MLP, and SVM, are used in order to find out the best combination. The experimental results show that that performing instance selection can have a positive impact on missing value imputation over the numerical data type of medical datasets, and specific combinations of instance selection and imputation methods can improve the imputation results over the mixed data type of medical datasets. However, instance selection does not have a definitely positive impact on the imputation result for categorical medical datasets.

## 1. Introduction

The first step in the data mining or knowledge discovery in databases (KDD) process is to collect a certain amount of data for a specific defined problem. However, in practice, it is usually the case that the medical dataset collected for later data mining steps is not complete due to problems such as manual data entry procedures, incorrect measurements, and equipment errors. As a result, the collected datasets generally contain some missing (attribute) values or missing data [9, 21].

For many data mining algorithms, it is not possible to develop learning models when used over incomplete medical datasets. Despite the fact that some algorithms, such as decision trees, can handle incomplete datasets without any pre-processing support [24], the final analysis or mining results

can be greatly affected by the incomplete datasets. In other words, the prediction performance of the constructed model trained by an incomplete dataset is questionable.

There are two types of solutions used to solve the problem of missing values or incomplete datasets. The first solution, which is the simplest and most straightforward solution, is based on case deletion. With the case deletion approach, the (incomplete) data having missing values are discarded directly. However, this method is generally appropriate only when the chosen dataset contains a very small amount of missing data, for example, 5% missing rate. The second solution is based on missing value imputation. It can provide estimations for missing values by reasoning from the observed data (i.e., complete data) [13, 14, 20].

In the literature, the efficacy of some different missing value imputation algorithms used with different kinds of

datasets containing various missing data rates has been compared. The experimental results have shown that missing value imputation is a better choice than case deletion when the incomplete datasets contain a certain amount of missing values. Model-based missing value imputation algorithms based on machine learning techniques, such as  $k$ -nearest neighbor, multilayer perceptron neural networks, and support vector machines, have recently lately been widely considered [14, 16, 21].

Since all of the model-based imputation algorithms require observed data without missing values in the incomplete dataset, as the training set to provide estimations of the missing values, the imputation results are directly affected by the observed data. From the view point of instance selection [12, 17], a given training set generally contains noisy data or outliers that can degrade the final performance of a learning model. The aim of instance selection is to filter out unrepresentative data from a given training set, and a learning model trained by the reduced training set is likely to perform better than the one trained by the original training set.

In other words, from the instance selection perspective, there would be some noisy data that exist in the observed dataset for missing value imputation. As a result, it is important to examine the performances of performing instance selection over the observed dataset before missing value imputation. In Tsai and Chang [22], different priorities of combining instance selection and imputation algorithms over various domain datasets were studied. Although they conclude that performing instance selection first and imputation second is the better combination process, they only use one specific instance selection algorithm combined with one specific imputation method for the experiments.

Therefore, the research objective of this paper is to examine whether methods combining instance selection and missing value imputation can outperform those using missing value imputation alone for incomplete medical datasets. The combination process is discussed below. Given an incomplete dataset, which contains complete data without missing values and incomplete data having missing values, the complete data for the imputation algorithms are selected by the instance selection process, and then missing value imputation is performed over the reduced set of complete data. For the medical domain classification problem, we aim to demonstrate that through different instance selection algorithms, the new imputation results by different imputation models can make the learning classifier performs better than the one using the original imputation results obtained without performing instance selection.

The contribution of this research is twofold. For missing value imputation, we show that using all of the observed data to produce the estimations to replace the missing values based on the baseline approach may not be the best imputation solution. That is, the quality of the observed data should be carefully considered. On the other hand, for instance selection, we demonstrate that it can be combined with the imputation process for incomplete medical datasets, which has never been done before.

In our experimental setup, three different instance selection algorithms and three different model-based imputation

algorithms are combined interchangeably in order to find the best combination for the incomplete medical datasets. The real-world medical datasets can contain categorical (i.e., discrete), numerical (i.e., continuous), or both types of data. Here, three types of datasets with different missing rates ranging from 10% to 50% are used to assess the imputation performance.

The rest of this paper is organized as follows. Section 2 overviews related literature including the missingness mechanisms, missing value imputation, and instance selection. Section 3 describes the two imputation processes that are examined in this paper, which are the baseline imputation process and the process of combining instance selection and missing value imputation. Section 4 presents the experimental setup and results. Finally, Section 5 concludes the paper.

## 2. Literature Review

*2.1. The Missingness Mechanisms.* Missing data randomness can be divided into three categories, namely, missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR) [19].

MCAR is the highest level of randomness. Let  $X$  be the random attribute. If  $P(X|x \text{ missing}) = P(X|x \text{ observed})$ , then the distribution of  $X$  is not affected by missing values. Therefore, MCAR refers to data where the missingness mechanism does not depend on the attribute of interest, or any other attribute, which is observed in the data. In other words, it occurs when the probability of an instance (case) having a missing value for an attribute does not depend on either the known values or the missing data. Any missing data treatment method can be applied at this level of randomness without risk of introducing bias into the data.

On the other hand, for MAR, let  $X$  be the random attribute, and let  $Z$  be a set of predictor attributes. If  $P(X|x \text{ missing}, Z) = P(X|x \text{ observed}, Z)$ , then the distribution of  $X$  is not affected by missing values for  $X \in Z$ . In other words, MAR occurs when the probability of an instance having a missing value for an attribute may depend on the value of that attribute.

NMAR occurs when the probability of an instance having a missing value for an attribute may depend on the value of that attribute. This is the most difficult condition to model. However, in practice, it is difficult to judge the missing data mechanism, as the values for the missing data are unknown.

*2.2. Missing Value Imputation.* Missing value imputation can be regarded as a pattern classification task. In pattern classification, each data sample is represented by a  $d$ -dimensional feature vector where  $d$  is the number of features or attributes. In addition, each feature vector usually belongs to one of  $c$  classes or categories. To develop a classifier, a given training set, composed of a number of training data, is used to train the chosen classification technique. For classification, a given testing set composed of a number of testing data having the same number of features that are unknown or new to the classifier, and the classifier, are used to classify each testing data sample into one of the learned  $c$  classes.

For missing value imputation, the given incomplete  $d$ -dimensional dataset can be divided into complete and incomplete subsets, to be used as the training and testing sets, respectively. For example, when its  $k$ th attribute of the  $i$ th incomplete data sample is missing (where  $k \in d$ ), the  $k$ th attribute of the training data is used as the final classification output, and the other  $d-1$  attributes, with the exception of the  $k$ th one, are used as the input features. The classifier is trained to classify the incomplete data having the  $k$ th missing attribute values [13].

The various missing value imputation methods can be classified into statistical and machine learning methods [5, 13]. Different imputation methods have been compared in the literature (e.g., [1, 5, 11]), and some novel imputation methods have been proposed in recent studies (e.g., [26, 27]).

In particular,  $k$ -nearest neighbor imputation (KNNI) [10] is one of the most popular approaches. It is based on the  $k$ -nearest neighbor classification principle where missing values are imputed using values calculated from the  $k$ -nearest neighbors. An important parameter for the KNNI method is the value of  $k$ , which is typically set to 1, but is sensitive to outliers. According to Jonsson and Wohlin [15], the performance is relatively unaffected by the value of  $k$  while Batista and Monard [5] report that  $k=10$  for large datasets.

For more detailed information about the other imputation methods, please refer to de Leeuw [8] and Garcia-Laencina et al. [13].

**2.3. Instance Selection.** The aim of instance selection is to filter out some noisy or unrepresentative data from a given (training) dataset. In practice, the collected data may not all be equally informative, and some data points are considered noisy points or outliers. Using the original dataset without excluding the outliers could lead to significant degradation in degradation [2], but performing instance selection is likely to increase generalization accuracy and the dataset size can also be reduced [25].

Instance selection can be defined as follows. Let  $X_i$  be an instance where  $X_i = (X_{i1}, X_{i2}, \dots, X_{im}, X_{ic})$ , meaning that  $X_i$  is represented by  $m$ -dimensional features and  $X_i$  belongs to class  $c$  given by  $X_{ic}$ . Then, assume that there is a target set TA that consists of  $M$  instances, which is used for instance selection. Consequently, a subset of selected samples  $S$  is produced, where  $S \subseteq TA$ . Given a testing set TS, we can classify a new pattern  $T$  from TS over the instances of  $S$  and TA. If the instance selection algorithm has been chosen appropriately, the classifier performance trained by  $S$  should be better than that of TA.

A number of related studies proposing instance selection methods for obtaining better mining quality appear in the literature. Recently, Garcia et al. [12] compared fifty related instance selection algorithms over various datasets in terms of classification accuracy. They divided related algorithms into three types of techniques, which are edition, condensation, and hybrid methods. Generally speaking, edition methods aim to remove noisy data samples from a given (training) set in order to increase classifier accuracy. Condensation methods aim to remove redundant data samples where the classifier's performance trained by the reduced

training set will not be affected. Hybrid methods focus on searching for a small subset by simultaneously eliminating both noisy and redundant data samples.

Although there is no exact winner for all of the problem datasets, they found that, on average, hybrid methods, such as genetic algorithms [6], IB3 [3], and DROP3 [25], are able to provide the largest data reduction rates and can make the trained classifiers outperform the ones without instance selection. In addition, since there is no generally agreed definition of outliers for different domain problems, the determination of outliers is based on the chosen instance selection method to filter out unrepresentative data samples from a given dataset.

### 3. The Two Imputation Processes

**3.1. The Baseline Imputation Process.** The baseline imputation process is described below. Given a dataset  $D$  with some missing values where each data sample is composed of a number of attributes and their associated class labels, the data with and without missing values can be denoted as complete ( $D_{\text{complete}}$ ) and incomplete subsets ( $D_{\text{incomplete}}$ ), where  $D \in D_{\text{complete}} + D_{\text{incomplete}}$ .

To impute the  $i$ th missing attribute of the  $j$ th data sample in  $D_{\text{incomplete}}$ , the  $i$ th attribute of  $D_{\text{complete}}$  is used as the output class for classification or prediction, and the other attributes, except for the original output class of  $D$ , are used as the input attributes (or variables). The resultant training set for estimating the  $i$ th missing attribute in  $D_{\text{incomplete}}$  is generated, while the data samples having the  $i$ th missing attribute in  $D_{\text{incomplete}}$  are used as the testing data.

In this paper, three different model-based imputation methods based on supervised learning techniques are considered for comparison: the KNNI ( $k=1$ ), multilayer perceptron (MLP) (the parameters of MLP are based on the default values of the Weka software), and support vector machine (SVM) (the parameters of SVM are based on the default values of LIBSVM [7]). The reason of choosing KNNI and MLP is because Garcia-Laencina et al. [13] compared KNNI, MLP, SOM (self-organizing map), and EM (expectation maximization) for missing value imputation and they found that KNNI and MLP perform similar and can provide better performances than SOM and EM. Note that for the numerical data type of datasets, support vector regression is used. The imputation result is based on the output of each method (or classifier) over the testing dataset.

After the imputation process is completed, which means that the original dataset  $D$  is imputed, denoted as  $D'$ , as a *pseudo-complete* dataset, then,  $D'$  is used as the training set and  $T$  for the testing set to train and test the SVM classifier, respectively. The final classification result is regarded as the evaluation metric and is used to examine the imputation performance of these three imputation methods.

**3.2. The Process of Combining Instance Selection and Imputation.** Differing from the baseline imputation process, in this process, instance selection is performed first, and then the output is used for missing value imputation. The first step

is to choose a specific instance selection method for removing some of the noisy data from the complete subset  $D_{\text{complete}}$ . The resultant reduced subset, denoted as  $D_{\text{complete\_reduced}}$ , is produced. Next,  $D_{\text{complete\_reduced}}$  and  $D_{\text{incomplete}}$  are combined (now denoted by  $D_{\text{reduced}}$ ) for missing value imputation by the three chosen imputation methods (i.e., KNNI, MLP, and SVM) individually. Note that the number of data samples in  $D_{\text{reduced}}$  is smaller than in  $D$  (and  $D'$ ). Finally, after performing imputation, the reduced dataset  $D_{\text{reduced}}$  becomes a *pseudo-complete* reduced dataset, denoted as  $D'_{\text{reduced}}$ .

In particular, during the instance selection step, three instance selection methods are employed for comparison, namely, IB3, DROP, and genetic algorithms (GA). They have been widely used as the baseline instance selection algorithms in related studies [18, 23]. There are nine different combinations of instance selection and imputation methods for each dataset: IB3/DROP3/GA + KNNI, IB3/DROP3/GA + MLP, and IB3/DROP3/GA + SVM.

Similar to the final step of the baseline imputation process, the SVM classifier is trained and tested by  $D'_{\text{reduced}}$  and  $T$ , respectively. Consequently, the classification accuracy of SVM trained by  $D'$  and  $D'_{\text{reduced}}$  over the testing set  $T$  and the results are compared to examine the instance selection effect.

## 4. Experiments

**4.1. Experimental Setup.** Three different attribute types of medical datasets are chosen from the UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/>), comprising categorical, numerical, and mixed attribute types of data, containing 4, 5, and 6 datasets, respectively. Moreover, each type of dataset contains different numbers of attributes, samples, and classes, which are helpful in determining the instance selection effect of using different types of datasets with different missing rates on the final classification accuracy. Table 1 lists the basic information for these datasets.

Each medical dataset is divided into a 90% training ( $D$ ) and 10% testing ( $T$ ) set based on the 10-fold cross-validation strategy. In addition, to examine the effect of performing instance selection on missing value imputation, five different missing rates for each dataset  $D$ , which are 10%, 20%, 30%, 40%, and 50% at 10% intervals, are simulated. Particularly, they are simulated by the MCAR (missing completely at random) mechanism, which is the most widely considered in related studies because MCAR is easy to be empirically tested [4]. Since larger missing rates by MCAR may cause each data sample containing at least one or more missing values, that is, there is no complete data sample in  $D$  for the imputation model, the criterion of simulating different missing rates is that at least five complete data samples containing no missing values should exist in  $D$ .

Moreover, in order to reduce the likelihood of obtaining biased results by randomly introducing missing values, each missing rate calculation is performed 10 times over each  $D$ . As a result, ten incomplete datasets are generated from each  $D$  with one specific missing rate. Then, the two imputation

TABLE 1: Dataset information.

Dataset	Number of instances	Number of attributes	Number of classes
<i>Categorical datasets</i>			
Lymphography	148	18	4
Nursery	12960	8	11
Promoters	106	58	2
SPECT	267	22	2
<i>Numerical datasets</i>			
Blood	748	5	2
Breast cancer	286	9	2
<i>E. coli</i>	336	8	8
Pima	768	8	2
Yeast	1484	8	10
<i>Mixed datasets</i>			
Abalone	4177	8	29
Acute	120	6	2
Contraceptive	1473	9	3
Liver_disorders	345	7	2
Statlog	270	13	2
Statlog_German	1000	20	2

processes are executed over the incomplete datasets for performance comparison.

**4.2. Results on Categorical Medical Datasets.** Figure 1 shows the average classification accuracy of SVM obtained using different imputation processes over the categorical medical datasets with different missing rates. We can see that there is a gradual degradation in the classification accuracy as the missing rates increase. This indicates that datasets with more missing values (i.e., larger missing rates) limit the complete data samples used as the training set for imputation, which is likely to make the classifier provide lower classification accuracy. This finding can be applied to two types of datasets.

Specifically, lowest rates of classification accuracy are obtained with the SVM classifier, combining DROP3 for instance selection with the three imputation methods. In particular, performing instance selection by IB3 and GA can provide significantly better imputation results than using DROP3 ( $p < 0.01$ ).

On the other hand, we found that when the missing rates fall below 30% (i.e., 10% and 20%), there is no need to consider performing instance selection before imputation over categorical datasets. This is because the SVM classifier based on the baseline imputation methods significantly outperforms the one based on the combined methods ( $p < 0.01$ ). In this case, the SVM imputation method performs the best and MLP the second best.

However, it is interesting that when the missing rates are larger than 30%, performance is slightly better when instance selection is combined with imputation than with the baseline imputation methods. In particular, combining the IB3 instance selection method with the imputation methods (i.e., KNNI, MLP, and SVM) significantly performs better

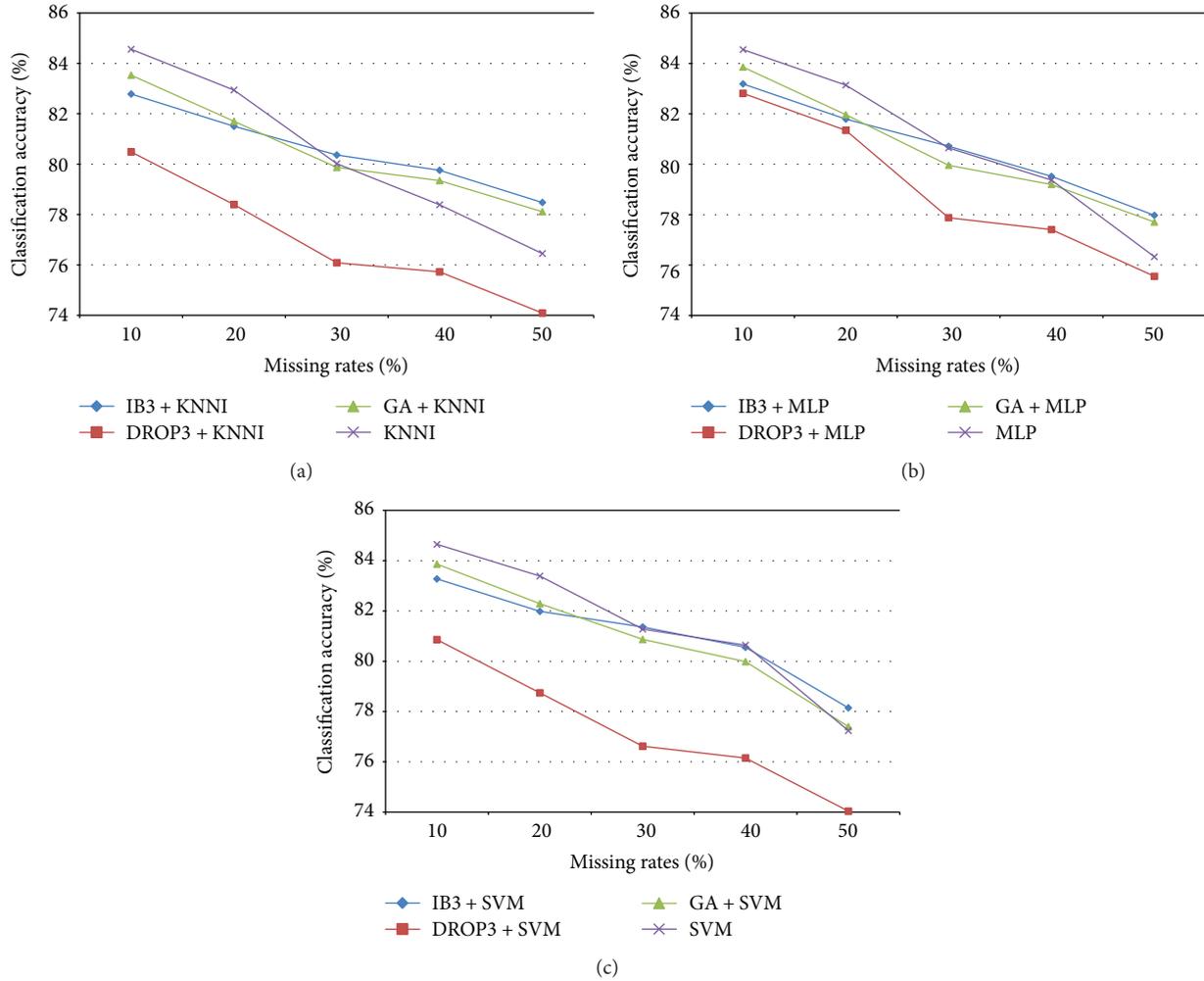


FIGURE 1: Classification results of imputation and instance selection combined with imputation over the categorical medical datasets.

than the baseline imputation methods ( $p < 0.05$ ). Therefore, performing instance selection could have some positive impact on missing value imputation over categorical datasets with the missing rates larger than 30%.

**4.3. Results on Numerical Medical Datasets.** Figure 2 shows the average classification accuracy of SVM produced by the different imputation processes over the numerical medical datasets with different missing rates. The results indicate that, in most cases, the combination of instance selection and imputation performs better than the baseline imputation methods (i.e., with different missing rates).

Although the differences in performance between most of the combinations are very small, that is, below 2% of classification accuracy, we still can find out that the best combination is based on GA + MLP for the 10% missing rate and IB3 + KNNI for the 20% to 50% missing rates, which significantly outperform the other combinations and the baseline imputation methods ( $p < 0.01$ ).

In contrast, comparison of the performance degradation with missing rates from 10% to 50% shows that the most stable classification performance is obtained with the SVM classifier based on IB3 + SVM. Specifically, the

performance degradation of SVM based on IB3 + SVM with 10% to 50% missing rates is 8%, whereas DROp3 + SVM is the second best (i.e., 8.13%) and IB3 + KNNI is the third best (i.e., 8.18%).

These results demonstrate that performing instance selection has a positive impact on missing value imputation over most numerical datasets. Furthermore, the choice of instance selection method does not significantly affect the imputation results.

**4.4. Results on Mixed Medical Datasets.** Figure 3 shows the average classification accuracy of the SVM obtained by different imputation processes over mixed medical datasets with different missing rates. We can see that performing instance selection can improve the imputation results, except for the MLP imputation method combination. In addition, it should be noted that the smallest performance improvement is obtained using DROp3 for instance selection when compared with IB3 and GA. The best combination is based on GA + MLP for 10% missing rate and IB3 + SVM for 20% to 50% missing rates, which significantly outperform the other combinations and the baseline imputation methods ( $p < 0.01$ ). However, the performance

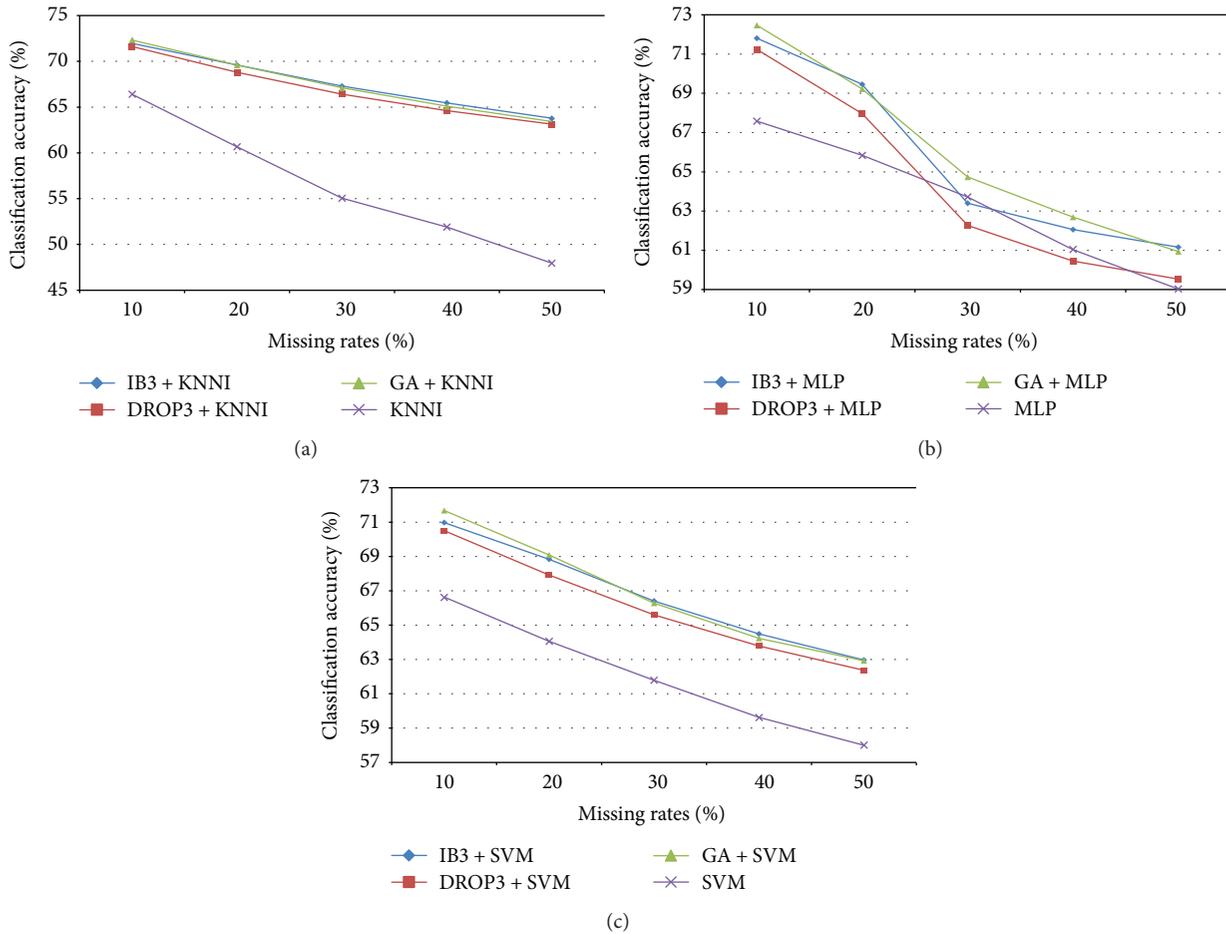


FIGURE 2: Classification results of imputation and instance selection combined with imputation over the numerical medical datasets.

differences between them are very small, that is, below 1% of classification accuracy.

Therefore, it can be concluded that with this type of datasets, performing instance selection may have a positive impact on the imputation results, if the instance selection and imputation methods are carefully chosen. The classifier performs better when using IB3 or GA for instance selection and KNNI or SVM for missing value imputation than performing the imputation step alone.

**4.5. Further Comparisons.** The best imputation and combined methods that significantly outperform the other methods ( $p < 0.05$ ) used over each medical dataset with different missing rates are listed in Table 2. There is no exact winner for different medical domain datasets with different missing rates. The findings obtained based on the dataset characteristics, such as the dimensionalities (i.e., number of attributes), the dataset sizes (i.e., number of instances), and the number of classes, are discussed below.

In categorical medical datasets, when their dimensionalities are low (e.g., lower than 22) and they belong to two-class classification problems, such as the SPECT dataset, meaning that when the dataset contains a relatively low complex problem, it is sufficient to use the baseline imputation process for most of the different missing rates. However,

when the categorical medical datasets contain very high dimensionalities, such as Promoters, performing instance selection may improve the imputation result.

On the other hand, for most numerical datasets, performing instance selection to filter out some noisy data can improve the imputation result. Similar to the numerical datasets, better results can be obtained by combining instance selection and imputation over most of the mixed medical datasets. The exception is the Acute dataset, which contains a small number of data samples, attributes, and classes (i.e., 160, 6, and 2, resp.), so better performance is obtained with the baseline imputation process than the combined process.

In summary, it is difficult to conclude whether combining instance selection and imputation is the better choice by only looking at one specific dataset characteristic, such as numbers of attributes. The three dataset characteristics usually relate to each other for each specific domain problem dataset. In spite of this, these experimental results show that performing instance selection before missing value imputation is recommended for most of the cases.

## 5. Conclusion

The incomplete dataset problem is usually approached by missing value imputation. In the past, many different types

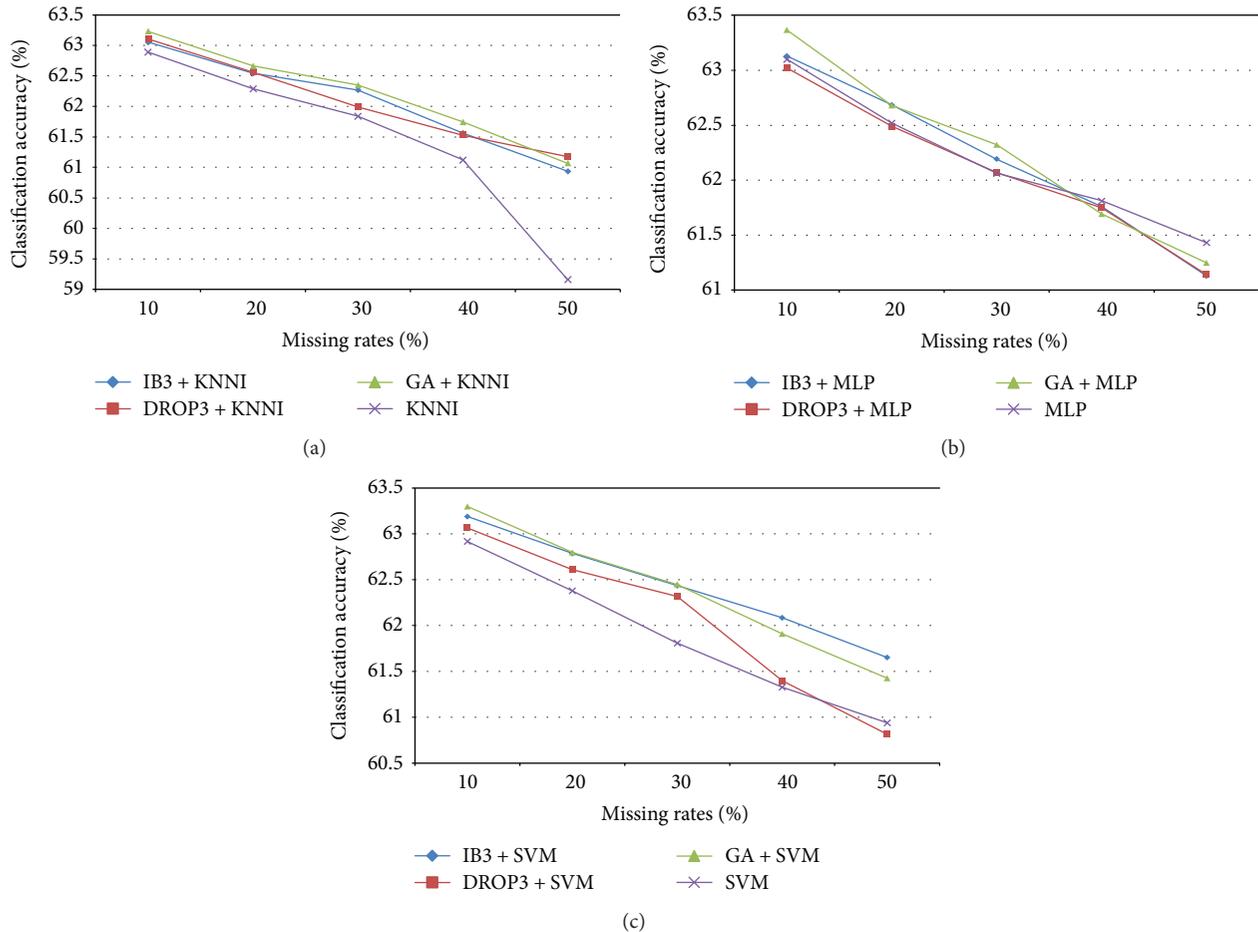


FIGURE 3: Classification results of imputation and instance selection combined with imputation over the mixed medical datasets.

of imputation algorithms have been studied. Model-based algorithms based on machine learning techniques have been applied recently. The imputation result is heavily dependent on the reasoning process used to process the observed data or training data, and the quality of the data that the imputation algorithms use to produce estimations to replace missing is an important issue.

In this paper, we focus on examining whether performing instance selection to filter out some noisy data from a given training medical dataset has a positive impact on the final imputation results. Specifically, the aim is to compare the classification performance obtained through the processes combining instance selection and imputation and the baseline imputation process. Three types of medical datasets including categorical, numerical, and mixed types of data are used. This allows us to identify the effect of performing instance selection on missing value imputation and understand when we should consider instance selection before imputation.

Three different instance selection methods and three model-based imputation algorithms are compared. Our experimental results show that performing instance selection first mostly improves the imputation result over these three types of medical data. In particular, we found that the negative impact is to consider instance selection before

imputation when the dataset contains lower dimensionalities and numbers of classes. However, for numerical datasets, the combined instance selection and imputation process performs better than the baseline imputation process for most datasets with different missing rates. Finally, for mixed datasets, the instance selection effect is between that for categorical and numerical datasets, which means that combining instance selection and imputation could be a better choice if both kinds of algorithms are carefully chosen.

Several issues could be considered in future work. First, since we only focus on the missing completely at random (MCAR) mechanism, other mechanisms should be considered with different types of medical data to fully examine the instance selection effect. Second, as there is no generally agreed upon definition of outliers, different instance selection algorithms will usually filter out different data samples from the same dataset. We think that fusing multiple instance selection results by the union or intersection strategy may produce a better quality of the observed (training) data for missing value imputation. Third, similar to the second issue, the imputation result could be improved if multiple imputation results were combined by different imputation algorithms. Last but not the least, some real-world big medical datasets containing very large volumes of data samples with high dimensionalities should be used for further study in

TABLE 2: The best imputation process over each dataset.

Dataset	Missing rate				
	10%	20%	30%	40%	50%
<i>Categorical datasets</i>					
Lymphography	GA + SVM	IB3 + KNNI	DROP3 + SVM	IB3 + KNNI	DROP3 + KNNI
Nursery	KNNI	GA + MLP	IB3 + MLP	MLP	MLP
Promoters	DROP3 + MLP	IB3/DROP3 + KNNI	IB3/DROP3 + MLP	IB3/DROP3 + SVM	IB3/DROP3 + SVM
SPECT	KNNI	MLP	MLP	MLP	KNNI
<i>Numerical datasets</i>					
Blood	GA + KNNI	GA + MLP	GA + MLP	GA + KNNI	DROP3 + MLP
Breast cancer	IB3 + SVM	IB3 + MLP	IB3 + SVM	IB3 + SVM	IB3 + SVM
<i>E. coli</i>	IB3 + KNNI	IB3 + KNNI	IB3 + KNNI	IB3 + KNNI	IB3 + KNNI
Prima	IB3/DROP3/GA + KNNI/MLP/SVM	IB3 + MLP	IB3 + KNNI/MLP DROP3 + KNNI GA + MLP	IB3 + KNNI/MLP	IB3 + MLP
Yeast	IB3 + SVM	IB3 + KNNI	IB3 + SVM	IB3 + SVM	GA + SVM
<i>Mixed datasets</i>					
Abalone	IB3 + SVM	GA + MLP	GA + MLP	IB3 + SVM	GA + MLP
Acute	GA + SVM KNNI/MLP/SVM	MLP/SVM	KNNI/SVM	SVM	MLP
Contraceptive	KNNI	SVM	SVM	IB3 + SVM	MLP
Liver_disorders	IB3 + KNNI	IB3 + KNNI	IB3 + KNNI	IB3/GA + SVM	IB3 + KNNI
Statlog	IB3 + KNNI/MLP	IB3 + KNNI/MLP/SVM GA + MLP	GA + SVM	IB3 + SVM GA + MLP/SVM	IB3 + MLP
Statlog_German	IB3/DROP3/GA + KNNI/MLP/SVM	IB3/DROP3/GA + KNNI/MLP/SVM	IB3/DROP3/GA + KNNI/MLP/SVM	IB3/DROP3/GA + KNNI/MLP/SVM	IB3/DROP3/GA + KNNI/MLP/SVM

order to conclude whether performing instance selection has a positive impact on missing value imputation.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research is partially supported by the Ministry of Science and Technology of Taiwan (MOST 103-2410-H-008-034-MY2).

## References

- [1] E. Acuna and C. Rodriguez, "The treatment of missing values and its effect on classifier accuracy," in *Classification, Clustering, and Data Mining Applications. Studies in Classification, Data Analysis, and Knowledge Organisation*, D. Banks, F. R. McMorris, P. Arabie and W. Gaul, Eds., pp. 639–647, Springer, Berlin, Heidelberg, 2004.
- [2] C. C. Aggarwal and P. S. Yu, "Outlier detection for high dimensional data," in *Proceedings of the 2001 ACM SIGMOD international conference on Management of data - SIGMOD '01*, vol. 30, pp. 37–46, Santa Barbara, CA, USA, 2001.
- [3] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, 1991.
- [4] A. N. Baraldi and C. K. Enders, "An introduction to modern missing data analyses," *Journal of School Psychology*, vol. 48, no. 1, pp. 5–37, 2010.
- [5] G. Batista and M. Monard, "An analysis of four missing data treatment methods for supervised learning," *Applied Artificial Intelligence*, vol. 17, no. 5–6, pp. 519–533, 2003.
- [6] J. R. Cano, F. Herrera, and M. Lozano, "Using evolutionary algorithms as instance selection for data reduction in KDD: an experimental study," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 6, pp. 561–575, 2003.
- [7] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
- [8] E. de Leeuw, "Reducing missing data in surveys: an overview of methods," *Quality and Quantity*, vol. 35, no. 2, pp. 147–160, 2001.
- [9] M. C. P. de Souto, P. A. Jaskowiak, and I. G. Costa, "Impact of missing data imputation methods on gene expression clustering and classification," *BMC Bioinformatics*, vol. 16, no. 1, pp. 1–9, 2015.
- [10] J. K. Dixon, "Pattern recognition with partly missing data," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 10, pp. 617–621, 1979.
- [11] A. Farhangfar, L. Kurgan, and J. Dy, "Impact of imputation of missing values on classification error for discrete data," *Pattern Recognition*, vol. 41, no. 12, pp. 3692–3705, 2008.
- [12] S. Garcia, J. Derrac, J. R. Cano, and F. Herrera, "Prototype selection for nearest neighbor classification: taxonomy and empirical study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 417–435, 2012.
- [13] P. J. Garcia-Laencina, J.-L. Sancho-Gomez, and A. R. Figueiras-Vidal, "Pattern classification with missing data: a review," *Neural Computing and Applications*, vol. 19, no. 2, pp. 263–282, 2010.
- [14] J. Huang, J. W. Keung, F. Sarro et al., "Cross-validation based  $K$  nearest neighbor imputation for software quality datasets: an empirical study," *Journal of Systems and Software*, vol. 132, pp. 226–252, 2017.
- [15] P. Jonsson and C. Wohlin, "An evaluation of  $k$ -nearest neighbor imputation using Likert data," in *10th International Symposium on Software Metrics, 2004. Proceedings*, pp. 108–118, Chicago, IL, USA, 2004.
- [16] M. A. Kiasari, G.-J. Jang, and M. Lee, "Novel iterative approach using generative and discriminative models for classification with missing features," *Neurocomputing*, vol. 225, pp. 23–30, 2017.
- [17] E. Leyva, A. Gonzalez, and R. Perez, "Three new instance selection methods based on local sets: a comparative study with several approaches from a bi-objective perspective," *Pattern Recognition*, vol. 48, no. 4, pp. 1523–1537, 2015.
- [18] W.-C. Lin, C.-F. Tsai, S.-W. Ke, C.-W. Hung, and W. Eberle, "Learning to detect representative data for large scale instance selection," *Journal of Systems and Software*, vol. 106, pp. 1–8, 2015.
- [19] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, John Wiley and Sons, Hoboken, NJ, USA, 1987.
- [20] J. Shao, W. Meng, and G. Sun, "Evaluation of missing value imputation methods for wireless soil datasets," *Personal and Ubiquitous Computing*, vol. 21, no. 1, pp. 113–123, 2017.
- [21] E.-L. Silva-Ramirez, R. Pino-ejias, and M. Lopez-Coello, "Single imputation with multilayer perceptron and multiple imputation combining multilayer perceptron and  $k$ -nearest neighbours for monotone patterns," *Applied Soft Computing*, vol. 29, pp. 65–74, 2015.
- [22] C.-F. Tsai and F.-Y. Chang, "Combining instance selection for better missing value imputation," *Journal of Systems and Software*, vol. 122, pp. 63–71, 2016.
- [23] C.-F. Tsai and Z.-Y. Chen, "Towards high dimensional instance selection: an evolutionary approach," *Decision Support Systems*, vol. 61, pp. 79–92, 2014.
- [24] H. C. Valdiviezo and S. Van Aelst, "Tree-based prediction on incomplete data using imputation or surrogate decisions," *Information Sciences*, vol. 311, pp. 163–181, 2015.
- [25] D. R. Wilson and T. R. Martinez, "Reduction techniques for instance-based learning algorithms," *Machine Learning*, vol. 38, no. 3, pp. 257–286, 2000.
- [26] S. Zhang, "Parimputation: from imputation and null-imputation to partially imputation," *IEEE Intelligent Informatics Bulletin*, vol. 9, no. 1, pp. 32–38, 2008.
- [27] X. Zhu, S. Zhang, Z. Jin, Z. Zhang, and Z. Xu, "Missing value estimation for mixed-attribute data sets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 1, pp. 110–121, 2011.

## Research Article

# Comparative Analysis of Hybrid Models for Prediction of BP Reactivity to Crossed Legs

Gurmanik Kaur,<sup>1</sup> Ajat Shatru Arora,<sup>2</sup> and Vijender Kumar Jain<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, SBBSU, Khiala, District Jalandhar, Punjab 144030, India

<sup>2</sup>Department of Electrical and Instrumentation Engineering, SLIET, Deemed University (Established by Govt. of India), Longowal, District Sangrur, Punjab 148106, India

Correspondence should be addressed to Gurmanik Kaur; [mannsliet@gmail.com](mailto:mannsliet@gmail.com)

Received 14 August 2017; Revised 5 October 2017; Accepted 16 October 2017; Published 26 November 2017

Academic Editor: Ming-Yuan Hsieh

Copyright © 2017 Gurmanik Kaur et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Crossing the legs at the knees, during BP measurement, is one of the several physiological stimuli that considerably influence the accuracy of BP measurements. Therefore, it is paramount to develop an appropriate prediction model for interpreting influence of crossed legs on BP. This research work described the use of principal component analysis- (PCA-) fused forward stepwise regression (FSWR), artificial neural network (ANN), adaptive neuro fuzzy inference system (ANFIS), and least squares support vector machine (LS-SVM) models for prediction of BP reactivity to crossed legs among the normotensive and hypertensive participants. The evaluation of the performance of the proposed prediction models using appropriate statistical indices showed that the PCA-based LS-SVM (PCA-LS-SVM) model has the highest prediction accuracy with coefficient of determination ( $R^2$ ) = 93.16%, root mean square error (RMSE) = 0.27, and mean absolute percentage error (MAPE) = 5.71 for SBP prediction in normotensive subjects. Furthermore,  $R^2$  = 96.46%, RMSE = 0.19, and MAPE = 1.76 for SBP prediction and  $R^2$  = 95.44%, RMSE = 0.21, and MAPE = 2.78 for DBP prediction in hypertensive subjects using the PCA-LSSVM model. This assessment presents the importance and advantages posed by hybrid computing models for the prediction of variables in biomedical research studies.

## 1. Introduction

Accurate measurement of blood pressure (BP) is indispensable for the diagnosis of hypertension at its early stage. Hypertension appears as a top risk factor for life-threatening conditions such as coronary artery disease, stroke, and kidney failure [1]. However, according to a recent editorial in the Hypertension journal of the American Heart Association (AHA), “few measurements in medicine are done as poorly and consistently as BP measurement. Though, there is clear recognition of biological variability, we continue to make decisions largely on measurements taken at random times under poorly controlled conditions” [2]. This observation supports the need to develop novel methods for accurate prediction of BP.

Recommendations of several international organisations including the AHA [3], British Hypertension Society (BHS) [4], and European Society of Hypertension (ESH) [5] revealed that BP is influenced by numerous biological and analytical sources of variation. Biological variations are relative to changes in the individual and are induced by, for instance, emotions, day and night rhythm, seasons, meals, and postures. Analytical variations are derived from the variability of the instrument used, observer bias, and so forth. However, it is not always feasible to control all the factors, but we can minimize their effect by taking them into account in reaching a decision [5].

Correct positioning of a subject's legs is often neglected during BP measurement. As it seems a comfortable position, subjects spontaneously cross their legs at the knees. Several

clinical and research studies have been proved that crossing the legs at knee level during BP measurement has a potential effect on the accuracy of measurements. Foster-Fitzpatrick et al. demonstrated a significant increase in BP taken with the legs crossed at the knee level in hypertensive subjects [6]. Peters et al. reported that crossed legs during BP measurement significantly increased systolic BP (SBP) and diastolic BP (DBP) in hypertensive subjects. In healthy volunteers, SBP and DBP increased when legs were crossed at knee level, but the effect was nonsignificant on DBP [7]. Keele-Smith and Price-Daniel, demonstrated that BP was significantly higher when legs were crossed versus uncrossed in a well-senior population [8]. Pinar et al. showed that crossing legs at knee level increased BP readings in hypertensive subjects [9]. Adiyaman et al. found significant increases in BP readings when the legs were crossed at knee level [10]. van Groningen et al. measured BP using a Finometer; they found an increase in BP readings with the legs crossed at knee level [11]. Pinar et al. reported that in hypertensive subjects, BP increased significantly when they crossed their legs [12].

Despite studies confirming the importance of leg position on BP measurement, it is likely that leg position varies markedly in clinical practice and also in published studies [2] and it may result in the misdiagnosis of hypertension or in overestimation of the severity of hypertension and may lead to overly aggressive therapy. Antihypertensive treatment may be unnecessary in the absence of concurrent cardiovascular risk factors [13].

Moreover, there is growing evidence that anthropometric indices are a major determinant of BP. Several studies have been conducted in the past to identify anthropometric characteristics that can be used as markers of BP [14–16]. These studies have explored a significant correlation between BP and anthropometric characteristics of a subject. Therefore, anthropometric characteristics should be considered to attain an accurate measurement of BP reactivity. However, multicollinearity between anthropometric characteristics has also been reported, which may result in “overfitting” of the prediction model [17–19].

The various methods utilized for prediction of biological variables range from the traditional statistical models to the complicated artificial intelligence-based models [20–25]. Recent studies on prediction of BP are as follows: Montemoreno presented a system for simultaneous noninvasive estimate of the blood glucose level (BGL), SBP, and DBP using a photoplethysmograph (PPG) and machine learning techniques. Physiological properties including blood viscosity, vessel compliance, hemodynamics, metabolic syndrome, demographic characteristics, and emotional state were used as input variables. The machine learning techniques tested were as follows: ridge linear regression, multilayer perceptron artificial neural network (ANN), support vector machine (SVM), and random forest. The best results were obtained with the random forest technique [26]. Genc proposed a linear stochastic model that integrated a known portion of the cardiovascular system and unknown portion through a parameter estimation to predict evolution of the mean arterial pressure (MAP). The performance of the model was

tested on a case study of acute hypotensive episodes (AHEs) on PhysioNet data. They concluded that true positive rates (TPRs) and false positive rates (FPRs) were improved during the prediction period [27]. Forouzanfar et al. presented a novel feature-based ANN for estimation of BP from wrist oscillometric measurements. Unlike previous methods that used the raw oscillometric waveform envelope (OMWE) as input to the ANN, in this paper, they proposed to use features extracted from the envelope. The OMWE was mathematically modeled as a sum of two Gaussian functions. The optimum parameters of the Gaussian functions were found by minimizing the least squares error (LSE) between the model and the OMWE using the Levenberg Marquardt algorithm and were used as input features. The performance of ANN was compared with that of the conventional maximum amplitude algorithm (MAA), adaptive neuro fuzzy inference system (ANFIS), and already-published ANN-based methods. It was found that the proposed approach achieved lower values of mean absolute error (MAE) and standard deviation ( $\sigma$ ) of error (SDE) in the estimation of BP [28]. Kurylyak et al. estimated the BP from the PPG signal using ANN. Training data were extracted from the multiparameter intelligent monitoring in an intensive care waveform database for better representation of possible pulse and pressure variation. The comparison between estimated and reference values showed better accuracy than the linear regression method [29]. Golino et al. compared the classification tree technique with traditional logistic regression for prediction of BP. Body mass index (BMI), waist circumference (WC), hip circumference (HC), and waist-hip ratio (WHR) were used as predictor variables. Finally, the comparison of the classification tree technique with traditional logistic regression indicated that the former outperformed the latter in terms of predictive power [30].

Hsin-Hsiuang et al. compared logistic regression, SVM, and permanental classification methods in predicting hypertension by using the genotype information. They used logistic regression analysis in the first step to detect significant single-nucleotide polymorphisms (SNPs). In the second step, they used the significant SNPs with logistic regression, SVM, and permanental classification methods for prediction purposes. The results showed that SVM and permanental classification both outperformed logistic regression [31]. Khan et al. proposed SVM for performing the prediction of BP with primary emotions using Facebook status. Current human BP and those belonging to up to six previous primary emotions and BP values with respect to human emotion were given as input variables. The outcome showed that SVM can be prosperously applied for prediction of BP through primary emotions. On the contrary, validations signified that the error statistics of the SVM model marginally outperformed [32]. Barbe et al. developed a logistic regression model to calibrate and correct an oscillometric monitor such that the device better corresponds to the Korotkoff method regardless of the health status of the patient. The model eliminated the systematic errors caused by patients suffering from hyper- or hypotension. They reported that systematic error was reduced by nearly 50% corresponding to the performance specifications of the device [33].

TABLE 1: Descriptive statistics of anthropometric characteristics of study samples.

Anthropometric characteristics	Normotensives		Hypertensives	
	Mean	SD	Mean	SD
Age (years)	23.1	1.24	42.83	6.665
Height (cm)	1.61	0.03	1.583	0.035
Weight (kg)	55.96	7.29	62.48	10.89
BMI (kg/m <sup>2</sup> )	21.55	2.504	23.57	3.497
MUAC (cm)	26.56	2.45	26.72	2.4

To perform a better training process and improve the forecasting accuracy, hybrid computing models in medical diagnosis are being developed to support physicians in successful decision making regarding clinical admission, early prevention, early clinical diagnosis, and application of clinical therapies by allowing calculation of disease likelihood based on known subject characteristics and clinical test results [34]. The main premise behind developing a hybrid computing model is to exploit the synergy between two or more models, leveraging their benefits and overcoming their respective limitations. The past few years have seen a vast interest in the hybrid computing models that seem to have completely replaced the traditional unisystem approaches. The rationale of using hybrid modeling in biomedical research studies is mainly to obtain fewer important predictor variables, and the selected predictor variables can serve as inputs for the designed prediction model. Hence, hybrid approach can improve the diagnostic accuracy with reduction in complexity of the prediction model [35].

The present study is a continuation of our previous studies [36, 37] dealing with the development of hybrid computing techniques for prediction of BP reactivity to talking and unsupported back. This research work focuses on the development of principal component analysis-(PCA-) based forward stepwise regression (FSWR), ANN, ANFIS, and least squares SVM (LS-SVM) hybrid computing models for prediction of BP reactivity to crossed legs by taking into account the anthropometric markers of BP in normotensive and hypertensive subjects. The prediction accuracy of the developed models was assessed using coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute percentage error (MAPE).

## 2. Materials and Methods

**2.1. Participants.** A total of 40 normotensive and 30 hypertensive subjects among the students, staff, and faculty of Sant Longowal Institute of Engineering and Technology, Deemed University, Longowal, Distt. Sangrur, Punjab, INDIA, were included in this study. Participants were aged over 18 years. Exclusion criteria were pregnant subjects, arrhythmic subjects, and the subjects who had a history of any condition that would interfere with positioning of lower extremity of the subjects. The institutional research committee approved the research

protocol and all participants gave written informed consent before participation.

**2.2. Data Collection.** A standard questionnaire was administered for the collection of anthropometric data including age, height, weight, BMI, and mid-upper arm circumference (MUAC) of the participants. The mean and standard deviation (SD) of the collected anthropometric data is given in Table 1.

A specially separated room was used to conduct this study. This ensured minimal interference within the room while the tests were being carried out. The observers involved in the study were trained using the BHS's BP measurement training materials [38].

To eliminate the observer bias, BP was measured using a validated, newly purchased, and fully automated sphygmomanometer OMRON HEM-7203 (OMRON HEALTHCARE Co. Ltd., Kyoto, Japan) that uses the oscillometric method of measurement. The BP monitor is available with a small cuff (17–22 cm), medium cuff (22–32 cm), and large cuff (32–42 cm). BP measurement was preceded by selection of the appropriate size cuff according to the MUAC of the subjects.

Subjects were advised to avoid alcohol, cigarette smoking, coffee/tea intake, and exercise for at least 30 minutes prior to their BP measurement. They were instructed to empty their bladder prior to measurements. Subjects were also instructed to sit upright on a chair with a supported back, kept the feet flat on the floor and the upper arm (under measurement) at heart level, as they are the potential confounding factors. Moreover, they were asked not to talk and move during measurement [3].

After a rest period of 5 minutes [3], the measurements were performed four times repeatedly at an interval of one minute. First measurement was discarded and the average of the last three measurements was taken into account. Subsequently, the legs were crossed at the knees and after four minutes, the same measurement protocol was repeated. All measurements were obtained under similar measurement conditions except for the different leg positions. And the measurement protocol was repeated for 7 days.

### 2.3. Experimental Methods

**2.3.1. PCA.** PCA is the first step of counteracting multicollinearity. It is a dimension reduction technique that does not

take the correlation between the input variables into account. Thus, PCA is considered as an unsupervised dimension reduction method [39–41]. To evaluate the influence of each input variable in the PCA, varimax rotation was used to obtain values of rotated factor loadings. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett’s test of sphericity were used to check the suitability of data for application of PCA [42–45].

**2.3.2. FSWR.** FSWR is a traditional statistical modeling technique used for developing an optimum prediction model by extracting the best anthropometric characteristics or predictor variables depending upon their statistical significance or probability ( $p$ ) value. It starts with an empty prediction model and adds one anthropometric predictor variables at a time. The first predictor variable included in the model has the highest correlation with the independent variable  $y$ . The second variable included is the one which has the highest correlation with  $y$ , after  $y$  has been adjusted for the effect of the first predictor variable. This process terminates when the last variable entering the model has insignificant regression coefficient [46].

**2.3.3. ANN.** To achieve the best architecture of ANN, various structures of feed-forward ANN with different numbers of hidden layers and neurons in each hidden layer were investigated. Finally, in light of the performance indices obtained from investigations, an ANN structure with two hidden layers and six nodes in each hidden layer was selected for further analysis. In addition, the architecture of ANN also consisted of one input layer with four input nodes (representing four PCs) and one output layer with one output node (representing BP reactivity to crossed legs). The choice of hyperbolic tangent sigmoid activation function for hidden layer and linear activation function for output layer trained the network in lesser number of epochs with better performance criteria and also yielded the best outcome predictions. The back propagation learning algorithm based on the Levenberg-Marquardt technique was used to find the local minimum of the error function. It blends the steepest descent method and the Gauss-Newton algorithm and inherits the speed advantage of the Gauss-Newton algorithm and the stability of the steepest descent method. It is more powerful and faster than the conventional gradient descent technique [47, 48].

**2.3.4. ANFIS.** A Sugeno-type FIS model was developed using “genfis1” with grid partitioning on data for prediction of BP reactivity to crossed legs. Different ANFIS parameters including numbers of membership functions (MFs) and types of input and output MF were tested to achieve the perfect training and maximum prediction accuracy. Input membership function “psigmf” and output membership function “linear” were used to develop the prediction model [49].

Other parameters of the trained ANFIS model were as follows: number of MFs=16, number of nodes=55, number of linear parameters=80, number of nonlinear

TABLE 2: Pearson’s correlation coefficients between each pair of anthropometric characteristics in normotensive and hypertensive subjects.

Anthropometric characteristics	Height	Weight	BMI	MUAC
Age (years)	0.535 <b>0.113</b>	0.784* <b>0.598</b>	0.701* <b>0.509</b>	0.668* <b>0.585</b>
Height (cm)		0.543 <b>0.165</b>	0.237 <b>0.305</b>	0.619* <b>0.021</b>
Weight (kg)			0.934* <b>0.885*</b>	0.743* <b>0.767*</b>
BMI (kg/m <sup>2</sup> )				0.617* <b>0.691*</b>

\* indicates  $p < 0.001$ ; bold values indicate correlations between anthropometric characteristics of hypertensive subjects.

parameters=32, total number of parameters=112, and number of fuzzy rules=16.

**2.3.5. LS-SVM.** The most important steps to develop a LS-SVM model are as follows: selection of a kernel and its parameters. After many experimental observations, radial basis function (RBF) kernel and grid search optimization algorithm (with 2-fold cross-validation) were selected to obtain the optimal combination of regularization parameter ( $\gamma$ ) and squared bandwidth ( $\sigma^2$ ) [50, 51].

### 3. Results

**3.1. Effect of Crossed Legs on BP.** The results of the paired  $t$ -test demonstrated a statistically significant higher SBP with crossed legs (mean difference  $\pm$  SD =  $5.838 \pm 2.5919$ ,  $p < 0.001$ ) in normotensive subjects, but there was no significant difference between DBP measurements (mean difference  $\pm$  SD =  $0.0037 \pm 0.0126$ ,  $p = 0.0737$ ). In hypertensive subjects, both SBP (mean difference  $\pm$  SD =  $10.3524 \pm 4.5844$ ,  $p < 0.001$ ) and DBP (mean difference  $\pm$  SD =  $6.1704 \pm 1.8531$ ,  $p < 0.001$ ) were significantly different when legs were crossed at knee level. These results are consistent with the recommendations of the AHA council for BP measurement in humans and experimental animals [3].

**3.2. Multicollinearity Diagnostic.** A visual inspection of the Pearson’s correlation coefficients revealed the existence of multicollinearity, as correlation coefficient  $> 0.6$  [52], between pairs of anthropometric characteristics, in normotensive and hypertensive individuals, as shown in Table 2.

**3.3. Application of PCA on BP Data.** In the next step, PCA was used to omit the multicollinearity between pairs of anthropometric characteristics and simplify the complexity of the relationship between them [53].

To verify the applicability of PCA, Bartlett’s test of sphericity was applied [54]. A high value of chi square ( $\chi^2$ ), for normotensive ( $\chi^2 = 231.012$ , DF = 10,  $p < 0.0001$ ) and hypertensive ( $\chi^2 = 119.48$ , DF = 10,  $p < 0.0001$ ) individuals

TABLE 3: Pearson's correlation coefficient between each pair of PCs in normotensive and hypertensive subjects.

PC	PC2	PC3	PC4
PC1	-0.0000225 <b>0.00000878</b>	0.000000798 <b>0.00000423</b>	-0.0000167 <b>0.00000659</b>
PC2		-7.237e-016 <b>0.00000919</b>	5.808e-016 <b>0.0000142</b>
PC3			-7.557e-017 <b>0.0000175</b>

Bold values indicate correlation in anthropometric characteristics of hypertensive subjects.

implied that PCA is applicable to our data set. The value of KMO was also greater than 0.6 for normotensive (0.63) and hypertensive (0.75) individuals, which indicates that our sample size is enough to apply PCA [55].

Out of 5 PCs, only the first four PCs (PC1–PC4), explaining more than 5% of variations, were retained for further analysis. In normotensive subjects, the selected PCs explained 99.8% of the total variation. Variance proportions explained by PC1, PC2, PC3, and PC4 were found as 71.84%, 16.58%, 6.34%, and 5.04%, respectively. In hypertensive subjects, the selected PCs explained 98.04% of the total variation. Variance proportion accounted for by PC1, PC2, PC3, and PC4 was estimated to be 61.10%, 22.5%, 8.78%, and 5.66%, respectively. Loadings of anthropometric characteristics after varimax rotation give an indication of the extent to which the original variables are influential in forming new variables. For both normotensive and hypertensive subjects, weight and BMI were the characteristics having the highest correlation with PC1 and height had the highest correlation with PC2.

Moreover, Pearson's correlation between pairs of PCs, as shown in Table 3, indicates that the problem of multicollinearity presented in Table 2 is solved as there is no significant relationship between any pair of PCs in the correlation table (correlation coefficient < 0.6).

To develop PCA-based prediction models, principal score values obtained from the principle score coefficients were used as independent variables and BP reactivity was used as dependent variable. Moreover, 80% data were used for training while the entire data set was used for testing. Data were normalized before training to achieve more accurate predictions. MATLAB (version 7.5) was used to develop the prediction models.

**3.4. PCA-Based FSWR (PCA-FSWR).** When probabilities were taken into consideration, the regressions of standardized SBP reactivity on PC1 (composed of weight and BMI) were found statistically significant in normotensive subjects. Whereas, PC3 (composed of age) was found statistically significant for SBP and DBP reactivity in hypertensive subjects. Figures 1(a)–1(c) show the scatter plot between the observed and predicted values of BP reactivity from the PCA-FSWR model in normotensive and hypertensive subjects.

The final model equations for prediction of BP reactivity in normotensive and hypertensive subjects are given as follows:

- (a) Model equation obtained for prediction of SBP reactivity in normotensive subjects:

$$\text{SBP reactivity} = 5.8381 - 1.8514(\text{PC1}). \quad (1)$$

- (b) Model equation obtained for prediction of SBP reactivity in hypertensive subjects:

$$\text{SBP reactivity} = 10.3524 - 1.6246(\text{PC3}). \quad (2)$$

- (c) Model equation obtained for prediction of DBP reactivity in hypertensive subjects:

$$\text{DBP reactivity} = 6.1704 - 0.6467(\text{PC3}). \quad (3)$$

**3.5. PCA-Based ANN (PCA-ANN).** The scatter plots between the observed and predicted values of BP reactivity from the PCA-ANN model, as illustrated in Figures 2(a)–2(c), although revealed marked deviations, but they were smaller than those from the PCA-FSWR model.

**3.6. PCA-Based ANFIS (PCA-ANFIS).** As presented in Figures 3(a)–3(c), the scatter plots plotted between observed and predicted values of BP reactivity from the PCA-ANFIS model clearly demonstrate improvements in predicted values as compared to those of the performance of the PCA-FSWR and PCA-ANN prediction models.

**3.7. PCA-Based LS-SVM (PCA-LS-SVM).** The optimal values of regularization parameter ( $\gamma$ ) and squared bandwidth ( $\sigma^2$ ) obtained from the developed PCA-LS-SVM model are as follows:

- (1)  $\gamma = 200$ ,  $\sigma^2 = 0.53$  (for prediction of SBP reactivity in normotensive subjects)
- (2)  $\gamma = 253.0920$ ,  $\sigma^2 = 0.0782$  (for prediction of SBP reactivity in hypertensive subjects)
- (3)  $\gamma = 1.0635e + 004$ ,  $\sigma^2 = 0.0148$  (for prediction of DBP reactivity in hypertensive subjects)

The scatter plots between the observed and predicted values of BP reactivity from PCA-LS-SVM as shown in Figures 4(a)–4(c) revealed the best predicted values when compared to predictions of the PCA-FSWR, PCA-ANN, and PCA-ANFIS models.

The comparison of statistical indices of the models, as shown in Table 4, reveals that the PCA-LS-SVM model has the highest value of  $R^2$  and lowest value of RMSE and MAPE for prediction of BP reactivity to crossed legs in normotensive and hypertensive subjects.

## 4. Discussion

Accurate prediction of BP is integral to successful decision making and leads to better patient care. Overestimation of

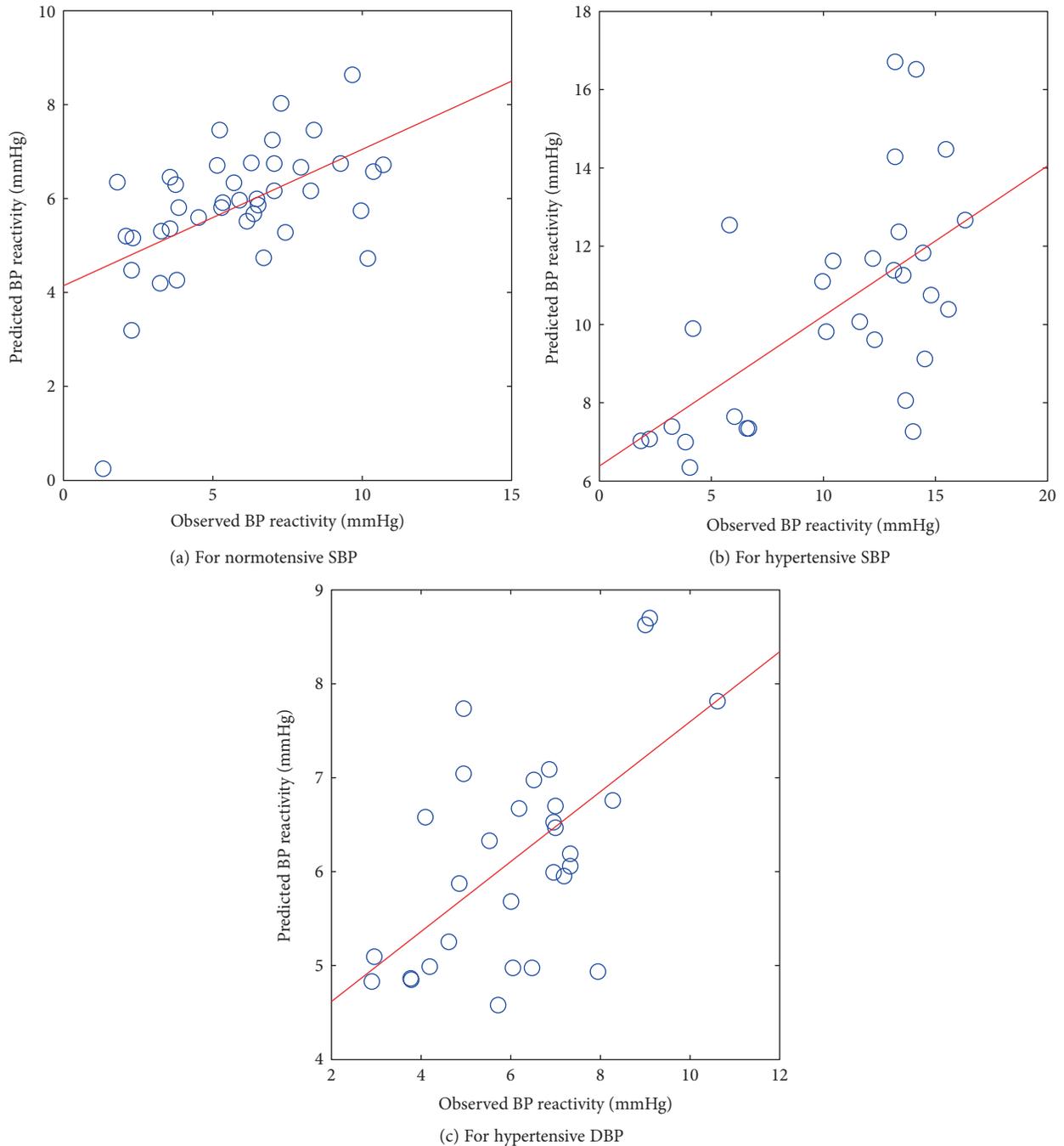


FIGURE 1: Scatter plot between observed and predicted values of BP reactivity using the PCA-FSWR model.

BP would increase the number of patients with hypertension. They may experience adverse effects of medication and have increased insurance and treatment cost. Furthermore, the inaccurate labeling leads to an increased perception of disease and absenteeism from work [56].

The marked elevation in BP with the crossed leg position may be due to isometric activity of the leg muscles. Isometric activity increases vascular resistance or total peripheral resistance (TPR) and BP [57]. Another explanation for the significant rise in BP with the crossed legs is

translocation of blood volume from the dependent vascular beds in the legs to the central thoracic compartment that causes a high stroke volume, as cardiac output is determined by the stroke volume multiplied by heart rate. Therefore, an increase in stroke volume causes an increase in cardiac output [6].

Evidently, this work demonstrates that crossed legs in sitting position significantly elevated SBP of normotensive subjects and SBP and DBP of hypertensive subjects. Similar conclusions were found by previous studies [6–12].

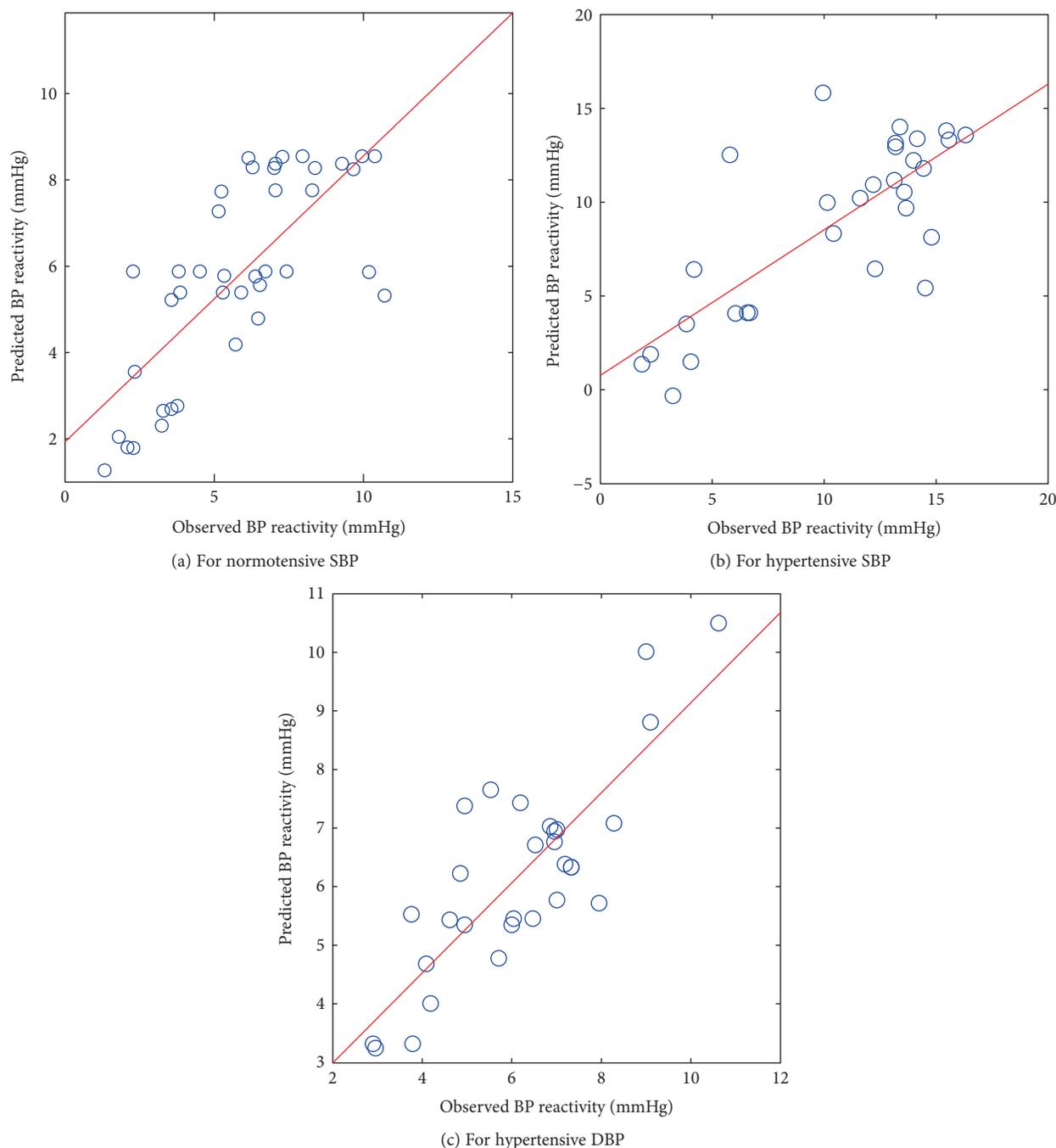


FIGURE 2: Scatter plot between observed and predicted values of BP reactivity using the PCA-ANN model.

Furthermore, PCA-based hybrid computing models for predictions of BP reactivity to crossed legs are proposed in this paper. To the best of our knowledge, this is the first study that focused specifically on prediction of BP reactivity to crossed legs using the PCA-FSWR, PCA-ANN, PCA-ANFIS, and PCA-LS-SVM models. Therefore, the results were compared with indirectly related prediction studies, as shown in Table 5.

In all studies, the higher performance of the soft computing models was sourced from a greater degree of robustness and fault tolerance than traditional models. The results of present research work illustrated that the

PCA-LS-SVM hybrid model obtained the best prediction results because LS-SVM is firmly based on the theory of statistical learning; therefore, it can attain a global optimal solution and has good generalization ability and low dependency on sample data.

The present study has a number of merits. We used small, medium, and large size cuffs to cover the entire MUAC range demanded by participants. Inappropriate cuff size results in underestimation or overestimation of BP. Moreover, to strengthen the accuracy of measurements, we took the mean of three readings per leg position for seven days [3].

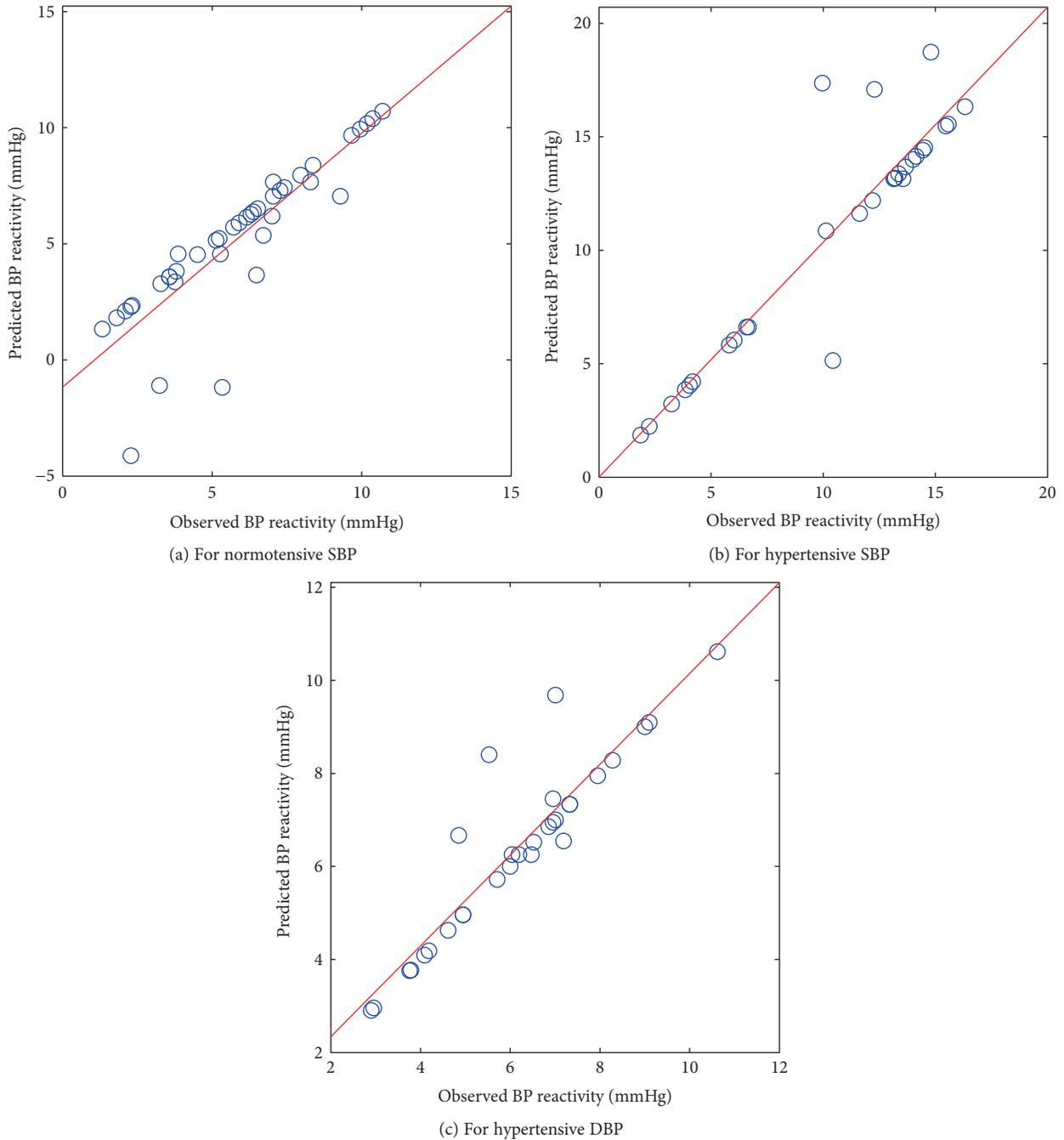


FIGURE 3: Scatter plot between observed and predicted values of BP reactivity using the PCA-ANFIS model.

However, any single comparison between the prediction models might not reliably represent the true end results. It is essential to assess the performance of prediction models in external validation studies using larger database.

### 5. Conclusions

This paper has detailed an examination of hybrid computing models in an effort to predict BP reactivity to crossed legs using anthropometric predictor variables. By eliminating the multicollinearity problem, PCA provided more objective

interpretation of anthropometric predictor variables used for prediction. Then, the PCA-FSWR, PCA-ANN, PCA-ANFIS, and PCA-LS-SVM models were tested for prediction of BP from PCs. It was found that the PCA-LS-SVM model achieves substantial improvements in terms of  $R^2$ , RMSE, and MAPE compared with all the other models. This research work may provide valuable reference for researchers and engineers who apply hybrid computing approaches for modeling biological variables. The results may also be helpful to physicians in making more accurate diagnosis of hypertension in clinical practice. Our future research is targeted

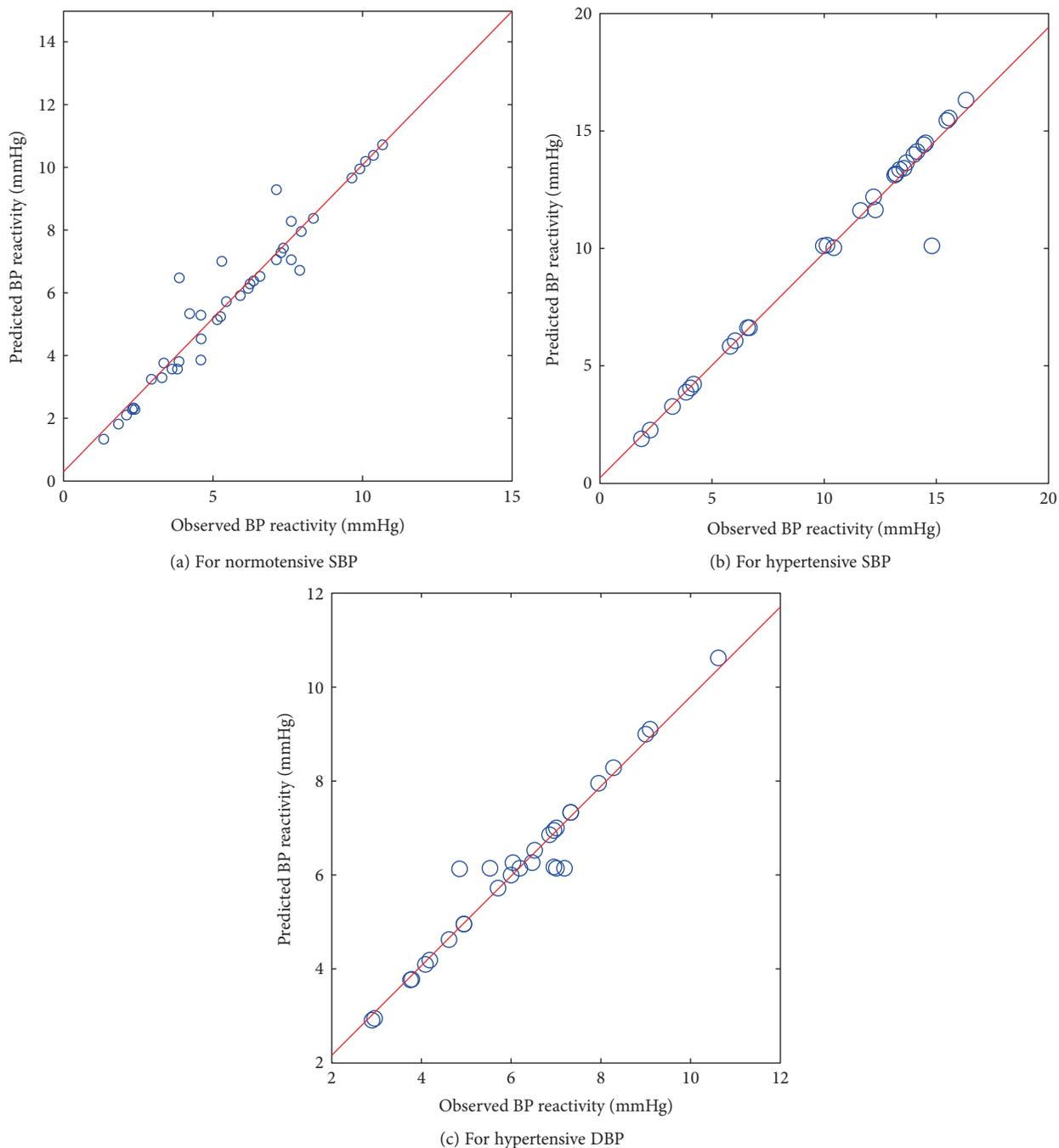


FIGURE 4: Scatter plot between observed and predicted values of BP reactivity using the PCA-LS-SVM model.

TABLE 4: Statistical indices for the proposed models.

Model	Normotensive subjects			Hypertensive subjects					
	SBP			SBP			DBP		
	$R^2$ (%)	RMSE	MAPE (%)	$R^2$ (%)	RMSE	MAPE (%)	$R^2$ (%)	RMSE	MAPE (%)
PCA-FSWR	29.05	2.21	40.33	38.35	3.66	48.35	37.21	1.49	22.72
PCA-ANN	55.67	0.67	26.25	60.11	0.74	30.39	67.91	0.57	14.63
PCA-ANFIS	75.42	0.67	17.39	84.81	0.44	6.74	84.26	0.44	5.06
PCA-LS-SVM	93.16	0.27	5.71	96.46	0.19	1.76	95.44	0.21	2.78

TABLE 5: Comparison of results with other studies.

Ref.	Model developed	Predicted parameter	Results
[26]	Ridge linear regression, ANN, SVM, and random forest	BGL, BP	Random forest technique outperformed ridge linear regression, ANN, and SVM. $R^2 = 0.91\%$ (SBP), $R^2 = 0.89\%$ (DBP), and $R^2 = 0.90\%$ (BGL)
[28]	ANN (raw input), ANN (feature based), MAA, and ANFIS (feature based)	SBP, DBP	ANN (feature based) achieved the best performance compared to other models. For SBP predictions: MAE = 6.28, SDE = 8.58. For DBP predictions: MAE = 5.73, SDE = 7.33
[29]	ANN	SBP, DBP	The experimental results confirmed the correctness of the ANN when compared with the linear regression model. Mean $\pm \sigma$ : SBP: $3.80 \pm 3.46$ , DBP: $2.21 \pm 2.09$ . Relative error: SBP: $3.48 \pm 3.19$ . DBP: $3.90 \pm 3.51$
[32]	SVM with RBF and polynomial kernel	SBP, DBP	SVM (RBF kernel) outperformed SVM (polynomial kernel). Coefficient of correlation ( $R$ ) = 0.97 (SBP), 0.96 (DBP). RMSE = 6.94 (SBP), and 5.78 (DBP). Scatter index (SI) = 22.34 (SBP), 22.79 (DBP)
[36]	PCA-ANN, PCA-ANFIS, and PCA-LS-SVM	SBP, DBP	PCA-LS-SVM outperformed PCA-ANN and PCA-ANFIS. For normotensive subjects: SBP: $R^2 = 95.42\%$ , RMSE = 0.21, and MAPE = 5.88%. DBP: $R^2 = 94.22\%$ , RMSE = 0.24, and MAPE = 4.05%. For hypertensive subjects: SBP: $R^2 = 98.76\%$ , RMSE = 0.11, and MAPE = 0.88%. DBP: $R^2 = 98.78\%$ , RMSE = 0.11, and MAPE = 0.84%
[37]	PCA-SWR, PCA-ANN, PCA-ANFIS, and PCA-LS-SVM	DBP	PCA-LS-SVM outperformed PCA-FSWR, PCA-ANN, and PCA-ANFIS. For normotensive subjects: $R^2 = 98.49\%$ , RMSE = 0.1243, and MAPE = 3.01%. For hypertensive subjects: $R^2 = 95.95\%$ , RMSE = 0.2013, and MAPE = 2.9%
[58]	ANN, ANFIS, and SVM	River flow in the semiarid mountain region	In comparing the results of the ANN, ANFIS, and SVM models, it was seen that the values of $R$ , RMSE, mean absolute relative error (MARE), and Nash-Sutcliffe (NS) of the SVM model were higher than those of ANN and ANFIS for all combinations of input data
[59]	ANN, ANFIS	To predict depths-to-water table one month in advance, at three wells located at different distances from the river	Both models can be used with a high level of precision to the model water tables without a significant effect of the distance of the well from the river, as model precision expressed via RMSE was roughly the same in all three cases (0.14154–0.15248). $R$ varied from 0.91973 to 0.9623 and coefficient of efficiency (COE) from 0.84588 to 0.92586
[60]	ANN, ANFIS, and SVM	Longitudinal dispersion coefficient (LDC)	The SVM model was found to be superior ( $R^2 = 90\%$ ) in predicting LDC due to low uncertainty as compared with those in the ANN ( $R^2 = 82\%$ ) and ANFIS ( $R^2 = 83\%$ ) models, while the ANFIS model performed better than the ANN model

TABLE 5: Continued.

Ref.	Model developed	Predicted parameter	Results
[61]	Multilayer perceptron (MLP), ANN, fuzzy genetic (FG), LS-SVM, multivariate adaptive regression spline (MARS), ANFIS, multiple linear regression (MLR), and Stephens and Stewart models (SS)	Evaporation in different climates	The accuracies of the applied models were rank as: MLP, GRNN, LSSVM, FG, ANFIS-GP, MARS, and MLR
Present study	PCA-FSWR, PCA-ANN, PCA-ANFIS, and PCA-LS-SVM	BP reactivity to crossed legs	PCA-LS-SVM outperformed PCA-FSWR, PCA-ANN, and PCA-ANFIS. For normotensive subjects: SBP: $R^2 = 93.16\%$ , RMSE = 0.27, and MAPE = 5.71%. For hypertensive subjects: SBP: $R^2 = 96.46\%$ , RMSE = 0.19, and MAPE = 1.76%. DBP: $R^2 = 95.44\%$ , RMSE = 0.21, and MAPE = 2.78%

to study an ensemble approach by combining the outputs of different hybrid techniques with more predictor variables. In addition, future research work will address using an ensemble approach by combining the outputs of different hybrid models with more predictor variables.

### Ethical Approval

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5).

### Consent

Informed consent was obtained from all participants for being included in the study.

### Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

### References

- [1] S. S. George and A. B. Loanni, "Home blood pressure monitoring in the diagnosis and treatment of hypertension: a systematic review," *American Journal of Hypertension*, vol. 24, no. 2, pp. 123–134, 2011.
- [2] D. W. Jones and J. E. Hall, "Hypertension: pathways to success," *Hypertension*, vol. 51, no. 5, pp. 1249–1251, 2008.
- [3] T. G. Pickering, J. E. Hall, L. J. Appel et al., "Recommendations for blood pressure measurement in humans and experimental animals: part 1: blood pressure measurement in humans: a statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Council on high blood pressure research," *Circulation*, vol. 111, no. 5, pp. 697–716, 2005.
- [4] B. Williams, N. R. Poulter, M. J. Brown et al., "Guidelines for management of hypertension: report of the fourth working party of the British hypertension society 2004-BHS IV," *Journal of Human Hypertension*, vol. 18, no. 3, pp. 139–185, 2004.
- [5] E. O'Brien, R. Asmar, L. Beilin et al., "European Society of Hypertension recommendations for conventional, ambulatory and home blood pressure measurement," *Journal of Hypertension*, vol. 21, no. 5, pp. 821–848, 2003.
- [6] L. Foster-Fitzpatrick, A. Ortiz, H. Sibilano, R. Marcantonio, and L. T. Braun, "The effects of crossed leg on blood pressure measurement," *Nursing Research*, vol. 48, no. 2, pp. 105–108, 1999.
- [7] G. L. Peters, S. K. Binder, and N. R. Campbell, "The effect of crossing legs on blood pressure: a randomized single blind cross-over study," *Blood Pressure Monitoring*, vol. 4, no. 2, pp. 97–101, 1999.
- [8] R. Keele-Smith and C. Price-Daniel, "Effects of crossing legs on blood pressure measurement," *Clinical Nursing Research*, vol. 10, no. 2, pp. 202–213, 2001.
- [9] R. Pinar, N. Sabuncu, and A. Oksay, "Effects of crossed leg on blood pressure," *Blood Pressure*, vol. 13, no. 4, pp. 252–254, 2004.
- [10] A. Adiyaman, N. Tosun, L. D. Elving, J. Deinum, J. W. M. Lenders, and T. Thien, "The effect of crossing legs on blood pressure," *Blood Pressure Monitoring*, vol. 12, no. 3, pp. 189–193, 2007.
- [11] L. F. van Groningen, A. Adiyaman, L. Elving, T. Thien, J. W. Lenders, and J. Deinum, "Which physiological mechanism is responsible for the increase in blood pressure during leg crossing?," *Journal of Hypertension*, vol. 26, no. 3, pp. 433–437, 2008.
- [12] R. Pinar and R. Watson, "The effect of crossing legs on blood pressure in hypertensive subjects," *Journal of Clinical Nursing*, vol. 19, no. 9–10, pp. 1284–1288, 2010.
- [13] D. G. Vedit, R. S. Lang, R. J. Seballos, A. Misra-Hebert, J. Campbell, and J. F. Bena, "Taking blood pressure: too important to trust to humans?," *Cleveland Clinic Journal of Medicine*, vol. 77, no. 10, pp. 683–688, 2010.
- [14] F. D. Fuchs, M. Gus, L. B. Moreira et al., "Anthropometric indices and the incidence of hypertension: a comparative analysis," *Obesity Research*, vol. 13, no. 9, pp. 1515–1517, 2005.
- [15] D. C. Moser, I. C. Giuliano, A. C. Titski, A. R. Gaya, M. J. Coelho-e-Silva, and N. Leite, "Anthropometric measures and blood pressure in school children," *Jornal de Pediatria*, vol. 89, no. 3, pp. 243–249, 2013.
- [16] I. Mariko, M. Masahide, and Y. Eiji, "Body mass index, blood pressure, and glucose and lipid metabolism among permanent

- and fixed-term workers in the manufacturing industry: a cross-sectional study," *BMC Public Health*, vol. 14, no. 207, pp. 1–8, 2014.
- [17] S. Doll, F. Paccaud, P. Bovet, M. Burnier, and V. Wietlisbach, "Body mass index, abdominal adiposity and blood pressure: consistency of their association across developing and developed countries," *International Journal of Obesity and Related Metabolic Disorders*, vol. 26, no. 1, pp. 48–57, 2002.
- [18] K. Bose, A. Ghosh, S. Roy, and S. Gangopadhyay, "Blood pressure and waist circumference: an empirical study of the effects of waist circumference on blood pressure among Bengalee male jute mill workers of Belur, West Bengal, India," *Journal of Physiological Anthropology and Applied Human Science*, vol. 22, no. 4, pp. 169–173, 2003.
- [19] R. Baba, M. Koketsu, M. Nagashima, H. Inasaka, M. Yoshinaga, and M. Yokota, "Adolescent obesity adversely affects blood pressure and resting heart rate," *Circulation Journal*, vol. 71, no. 5, pp. 722–726, 2007.
- [20] K. Gurmanik, A. Ajat Shatru, and J. K. Vijender, "Multi-class support vector machine classifier in EMG diagnosis," *WSEAS Transactions on Signal Processing*, vol. 5, no. 12, pp. 379–389, 2009.
- [21] M. R. Mohebian, H. R. Marateb, M. Mansourian, M. A. Mananas, and F. Mokarian, "A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 75–85, 2017.
- [22] J. Thomas and P. R. Theresa, "Human heart disease prediction system using data mining techniques," in *Proceedings of the International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, Nagercoil, India, March, 2016.
- [23] E. W. Steyerberg, A. J. Vickers, N. R. Cook et al., "Assessing the performance of prediction models: a framework for some traditional and novel measures," *Epidemiology*, vol. 21, no. 1, pp. 128–138, 2010.
- [24] M. Vezzoli, A. Ravaggi, L. Zanotti et al., "RERT: a novel regression tree approach to predict extrauterine disease in endometrial carcinoma patients," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.
- [25] P. Johnson, L. Vandewater, W. Wilson et al., "Genetic algorithm with logistic regression for prediction of progression to Alzheimer's disease," *BMC Bioinformatics*, vol. 15, no. 16, pp. 1–14, 2014.
- [26] E. Monte-Moreno, "Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques," *Artificial Intelligence in Medicine*, vol. 53, no. 2, pp. 127–138, 2011.
- [27] S. Genc, "Prediction of mean arterial blood pressure with linear stochastic models," in *Proceedings of the IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*, Boston, MA, USA, 2011.
- [28] M. Forouzanfar, H. R. Dajani, V. Z. Groza, M. Bolic, and S. Rajan, "Feature-based neural network approach for oscillometric blood pressure estimation," *IEEE Transaction on Instrumentation and Measurement*, vol. 60, no. 8, pp. 2786–2796, 2011.
- [29] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A neural network-based method for continuous blood pressure estimation from a PPG signal," in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Minneapolis, MN, USA, May 2013.
- [30] H. F. Golino, S. B. Amaral, S. F. P. Duarte et al., "Predicting increased blood pressure using machine learning," *Journal of Obesity*, vol. 2014, Article ID 637635, 12 pages, 2014.
- [31] H. H. Huang, T. Xu, and J. Yang, "Comparing logistic regression, support vector machines, and permanental classification methods in predicting hypertension," *BMC Proceedings*, vol. 8, Supplement 1, pp. S96.1–S96.5, 2014.
- [32] S. M. U. Khan, J. S. Manzoor, and S. U. J. Lee, "Predicting student blood pressure by support vector machine using Facebook," in *Proceedings of the IEEE World Congress on Services*, Anchorage, AK, USA, June–July 2014.
- [33] K. Barbe, Y. Kurylyak, and F. Lamonaca, "Logistic ordinal regression for the calibration of oscillometric blood pressure monitors," *Biomedical Signal Processing and Control*, vol. 11, pp. 89–96, 2014.
- [34] A. G. Floares, "Using computational intelligence to develop intelligent clinical decision support systems," in *Computational Intelligence Methods for Bioinformatics and Biostatistics*, vol. 6160 of Lecture Notes in Computer Science, pp. 266–275, Springer, Berlin, Heidelberg, 2010.
- [35] R. Kala, A. Shukla, and R. Tiwari, "Comparative analysis of intelligent hybrid systems for detection of PIMA Indian diabetes," in *Proceedings of the IEEE World Congress on Nature & Biologically Inspired Computing (NaBIC)*, Coimbatore, India, 2009.
- [36] K. Gurmanik, A. Ajat Shatru, and J. K. Vijender, "Prediction of BP reactivity to talking using hybrid soft computing approaches," *Computational and Mathematical Methods in Medicine*, vol. 2014, Article ID 762501, 13 pages, 2014.
- [37] K. Gurmanik, A. Ajat Shatru, and J. K. Vijender, "Using hybrid models to predict blood pressure reactivity to unsupported back based on anthropometric characteristics," *Frontiers of Information Technology & Electronic Engineering*, vol. 16, no. 6, pp. 474–485, 2015.
- [38] British Hypertension Society, *Blood Pressure Measurement (CD-ROM)*, BMJ Books, London, 1998.
- [39] H. Beatriz, P. Rafeal, V. Marisol, B. Enrique, F. J. Manuel, and F. Luis, "Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis," *Water Research*, vol. 34, no. 3, pp. 807–816, 2000.
- [40] M. A. Pett, N. R. Lackey, and J. J. Sullivan, *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*, SAGE Publication, Inc, California, 2003.
- [41] C. Handan, D. Nilsun, K. Arzu, and K. Syddyk, "Use of principal component scores in multiple linear regression models for prediction of *Chlorophyll-a* in reservoirs," *Ecological Modelling*, vol. 181, no. 4, pp. 581–589, 2005.
- [42] D. A. Jackson, "Stopping rules in principal components analysis: a comparison of heuristic and statistical approaches," *Ecology*, vol. 74, no. 8, pp. 2204–2214, 1993.
- [43] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, Pearson Education, London, 2006.
- [44] R. Noori, M. A. Abdoli, A. A. Ghasrodashti, and M. J. Ghazizade, "Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: a case study of Mashhad," *Environmental Progress & Sustainable Energy*, vol. 28, no. 2, pp. 249–258, 2009.

- [45] H. Wackernagel, *Multivariate Geostatistics: An Introduction with Applications*, Springer-Verlag, New York and London, 2010.
- [46] S. Chatterjee, A. S. Hadi, and B. Price, *Regression Analysis by Example*, Wiley, New York, 2004.
- [47] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.
- [48] I. A. Basheer and M. Hajmeer M., "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3–31, 2000.
- [49] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.
- [50] A. J. Smola and B. Scholkopf, "On a kernel based method for pattern recognition, regression, approximation and operator inversion," *Algorithmica*, vol. 22, no. 1-2, pp. 211–231, 1998.
- [51] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific Publishing, London, UK, 2002.
- [52] J. Benesty, J. Chen, and Y. Huang, "On the importance of the Pearson correlation coefficient in noise reduction," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 4, pp. 757–765, 2008.
- [53] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2002.
- [54] P. R. Pares-Neto, D. A. Jackson, and K. M. Somers, "How many principal components? Stopping rules for determining the number of non-trivial axes revisited," *Computational Statistics & Data Analysis*, vol. 49, no. 4, pp. 974–997, 2005.
- [55] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 141–151, 1960.
- [56] J. Handler, "The importance of accurate blood pressure measurement," *The Permanente Journal*, vol. 13, no. 3, pp. 51–54, 2009.
- [57] J. Hisdal, K. Toska, T. Flatebo, B. Waaler, and L. Walloe, "Regulation of arterial blood pressure in humans during isometric muscle contraction and lower body negative pressure," *European Journal of Applied Physiology*, vol. 91, no. 2-3, pp. 336–341, 2004.
- [58] H. Zhibin, W. Xiaohu, L. Hu, and D. Jun, "A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region," *Journal of Hydrology*, vol. 509, pp. 379–386, 2014.
- [59] D. Nevenka, D. Milka, S. Ruzica et al., "Comparison of groundwater level models based on artificial neural networks and ANFIS," *The Scientific World Journal*, vol. 2015, Article ID 742138, 13 pages, 2015.
- [60] N. Roohollah, D. Zhiqiang, K. Amin, and T. K. Fatemeh, "How reliable are ANN, ANFIS, and SVM techniques for predicting longitudinal dispersion coefficient in natural rivers?," *Journal of Hydraulic Engineering*, vol. 142, no. 1, pp. 1–8, 2016.
- [61] W. Lunche, K. Ozgur, Z. K. Mohammad, and G. Yiqun, "Comparison of six different soft computing methods in modeling 2 evaporation in different climates," *Hydrology and Earth System Sciences Discussions*, pp. 1–51, 2016.

## Research Article

# Clinical Decision Support System for Diabetes Based on Ontology Reasoning and TOPSIS Analysis

Rung-Ching Chen,<sup>1</sup> Hui Qin Jiang,<sup>1,2</sup> Chung-Yi Huang,<sup>1,3</sup> and Cho-Tsan Bau<sup>4</sup>

<sup>1</sup>Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan

<sup>2</sup>College of Computer and Information Engineering, Xiamen University of Technology, No. 600, Ligong Rd., Jimei District, Xiamen, Fujian, China

<sup>3</sup>Library, Chienkuo Technology University, Changhua, Taiwan

<sup>4</sup>Taichung Hospital, Ministry of Health and Welfare, Executive Yuan, Taichung, Taiwan

Correspondence should be addressed to Rung-Ching Chen; [rcching@cyut.edu.tw](mailto:rcching@cyut.edu.tw)

Received 2 July 2017; Accepted 10 September 2017; Published 26 October 2017

Academic Editor: Weide Chang

Copyright © 2017 Rung-Ching Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Introduction.** Although a number of researchers have considered the positive potential of Clinical Decision Support System (CDSS), they did not consider that patients' attitude which leads to active treatment strategies or *HbA1c* targets. **Materials and Methods.** We adopted the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) published to propose an *HbA1c* target and antidiabetic medication recommendation system for patients. Based on the antidiabetic medication profiles, which were presented by the American Association of Clinical Endocrinologists (AACE) and American College of Endocrinology (ACE), we use TOPSIS to calculate the ranking of antidiabetic medications. **Results.** The endocrinologist set up ten virtual patients' medical data to evaluate a decision support system. The system indicates that the CDSS performs well and is useful to 87%, and the recommendation system is suitable for outpatients. The evaluation results of the antidiabetic medications show that the system has 85% satisfaction degree which can assist clinicians to manage T2DM while selecting antidiabetic medications. **Conclusions.** In addition to aiding doctors' clinical diagnosis, the system not only can serve as a guide for specialty physicians but also can help nonspecialty doctors and young doctors with their drug prescriptions.

## 1. Introduction

The Institute of Medicine [1] defines patient-centered care strategy as "providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions." Clinical practitioners need to select different drugs to meet the needs of patients. However, patients with type 2 diabetes mellitus exhibit tremendous differences in phenotypes resulting in significant heterogeneity in clinical results. Consequently, clinical practitioners cannot be certain whether a prescription for a particular patient is the best.

Clinical decision support system (CDSS) may help clinicians, patients, and others to suggest patient-appropriate evidence-based treatment options. Ontologies are essential tools for the organization and representation of knowledge [2–7]. Ontologies contain the collection of patients, symptoms, diseases, diagnoses, treatments, and drug information, thereby creating a healing strategy according to patient's requirements to reconfigure a clinical decision support system [8]. Some of the studies suggested using ontologies to build clinical guidelines and care plans [5, 9–12].

In most of the knowledge ontologies, there is a design by the experience of domain experts. For example, Bau et al. [2] used domain ontology and rule reasoning to construct a

TABLE 1: Utilization of antidiabetic medications from 1998 to 2013.

Authors	(Publication year) title	Source period (year)	Country	Antidiabetic medications
Chang et al. [15]	(2012) National trends in antidiabetic treatment in Taiwan, 2000–2009	2000–2009	Taiwan	Biguanides, SU, Meglitinides, TZDs, $\alpha$ -glucosidase, DPP-4, insulin
Abdelmoneim et al. [20]	(2013) Use patterns of antidiabetic regimens by patients with type 2 diabetes	1998–2010	Canada	Biguanides, SU, Meglitinides, TZDs, $\alpha$ -glucosidase, insulin
Kohro et al. [17]	(2013) Trends in antidiabetic prescription patterns in Japan from 2005 to 2011—impact of the introduction of dipeptidyl peptidase-4 inhibitors	2005–2011	Japan	Biguanides, SU, Meglitinides, TZDs, $\alpha$ -glucosidase, DPP-4, GLP-1, insulin
Hsu et al. [21]	(2015) Utilization of oral antidiabetic medications in Taiwan following strategies to promote access to medicines for chronic diseases in community pharmacies	2001–2010	Taiwan	Biguanides, SU, Meglitinides, TZDs, $\alpha$ -glucosidase, DPP-4, GLP-1
Rafaniello et al. [18]	(2015) Trends in the prescription of antidiabetic medications from 2009 to 2012 in a general practice of Southern Italy: a population-based study	2009–2012	Italy	Biguanides, SU, Meglitinides, TZDs, $\alpha$ -glucosidase, DPP-4, GLP-1, insulin
Ko et al. [19]	(2016) Trends of antidiabetic drug use in adult Type 2 diabetes in Korea in 2002–2013: nationwide population-based cohort study	2002–2013	Korea	Biguanides, SU, Meglitinides, TZDs, $\alpha$ -glucosidase, DPP-4, insulin
Ou et al. [16]	(2016) Recent trends in the use of antidiabetic medications from 2008 to 2013: a nationwide population-based study from Taiwan	2008–2013	Taiwan	Biguanides, SU, Meglitinides, TZDs, $\alpha$ -glucosidase, DPP-4, GLP-1, insulin

SU: sulfonylureas; TZDs: thiazolidinediones; DPP-4: dipeptidyl peptidase-4; SGLT2: sodium-glucose cotransporter 2; GLP-1: glucagon-like peptide 1.

CDSS for diabetic patients undergoing surgery. They have three main classes in this ontology: disease, management, and patient. The disease class consists of diabetes and comorbidity information. The management class consists of anesthesia, capillary glucose tests, control of DM, medication, no medication, and water restriction information. The patient class consists of the patient clinical profile. The system constructs a clinical decision support system (CDSS) for undergoing surgery based on domain ontology and rules reasoning in the setting of hospitalized diabetic patients. However, the ontology knowledge is built on the experience of clinical practitioners, so it is hard to update these ontologies knowledge when there is a new clinical guideline.

Sherimon and Krishnan [11] had proposed an OntoDiabetic system which an ontology-based clinical decision support system for risk analysis and prediction of diabetes mellitus. The system consists of two main ontologies: the diabetic patient clinical analysis ontology and the semantic profile. The diabetic patient clinical analysis ontology and reasoning rules encapsulate the NICE (National Institute for Health and Care Excellence) guidelines. The OntoDiabetic system calculates the score and predicts the risk of diabetic patients due to smoking, alcohol, physical activity,

and sexual and cardiovascular diseases that mainly affect diabetes. Lots of effort was made on OntoDiabetic. What seems to be lacking, however, is that the system cannot provide antidiabetic medications suggestion.

Because there are many types of antidiabetic medications, they need to obtain permission from the government. Although the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) statements [13, 14] provide 12 types of antidiabetic medications, not every drug can be used. Table 1 shows eight antidiabetic medications which are commonly used in Japan, Korea, Canada, Italy, and Taiwan from 1998 to 2013 [15–21].

Although a number of researchers have considered the positive potential of CDSS, they did not consider patients' characteristics. For example, at the "patients' attitude" factor, if the patient has highly motivated or excellent self-care capacities, it can use active treatment strategies or *HbA1c* targets. To solve this problem, we proposed a solution in our previous research which adopted the ADA and the EASD standards who published an updated position statement on the management of hyperglycemia in type 2 diabetes to build *HbA1c* target inference module as well as drug knowledge ontology [22]. The system

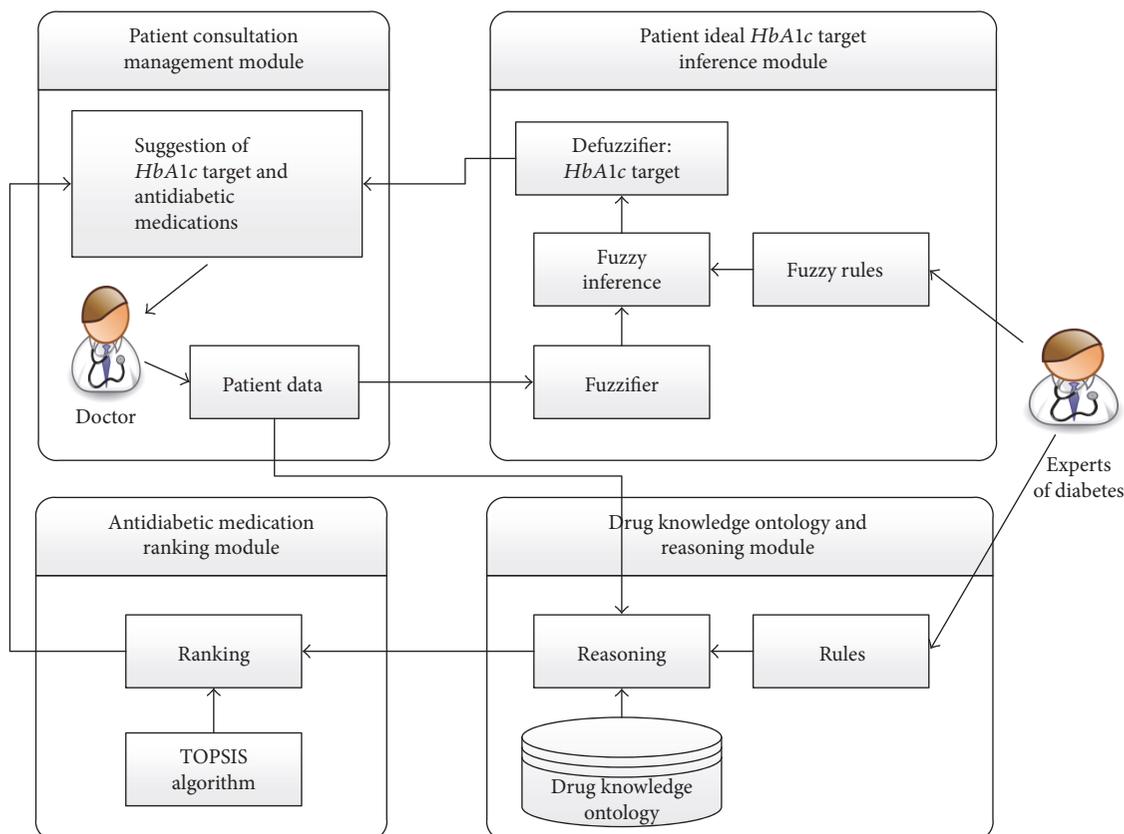


FIGURE 1: The recommendation system.

combines fuzzy logic and ontology reasoning to propose an antidiabetic medication recommendation system for patients with T2DM.

In this paper, we will further consider the safety and positivity of *HbA1c* target, and the priority of antidiabetic medication. We use the antidiabetic medication profiles, which are presented by the American Association of Clinical Endocrinologists (AACE) and American College of Endocrinology (ACE) in 2016 [23]. Based on the antidiabetic medication profiles, we used the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) to calculate the relative closeness to the ideal solution and thus determined the ranking of antidiabetic medications.

## 2. Materials and Methods

The recommendation system consists of four modules: the patient consultation management, the patient perfect *HbA1c* target inference, the drug knowledge ontology and reasoning, and antidiabetic medication ranking modules. The framework of the recommendation system is shown in Figure 1.

The first step, the “patient consultation management module,” provides a user interface to the clinical doctor. So, the clinical doctor can input patient’s data in the user interface. Those patient’s data will be provided to the other two modules. The second step, the “patient ideal *HbA1c* target inference module,” will use fuzzy technology to infer the patient’s individualization *HbA1c* target. The third step, the

“drug knowledge ontology and reasoning module,” will recommend antidiabetic medications for the patient. The fourth step, the “antidiabetic medications ranking module,” will use TOPSIS technology to calculate the relative closeness to the ideal solution and thus determine the ranking of antidiabetic medications. In this system, the experts of diabetes decided the fuzzy rules and ontology reasoning rules.

**2.1. Patient Consultation Management Module.** The patient consultation management module requires the patients’ data which is also necessary for the other modules. The modulation of the intensiveness of glucose lowering therapy in T2DM is according to the ADA and the EASD position statement [14]. The sufficient communication between the clinical doctor and the patient is also necessary to evaluate seven factors. They are (1) the risks associated with hypoglycemia and other drug adverse effects, (2) disease duration, (3) life expectancy, (4) important comorbidities, (5) established vascular complications, (6) patient attitude and expected treatment efforts, and (7) resources and support system. Each of the seven factors has five levels measured by integers 0 to 4. The clinical doctor also needs to record adverse drug reactions (ADRs) and individual history of diseases.

**2.2. Patient Ideal *HbA1c* Target Inference Module.** The main functional modules include fuzzifier, fuzzy rules, fuzzy inference, and defuzzier. There are seven inputs, namely,  $x_1, \dots, x_7$ , for fuzzy logic and the input factors are divided into five

TABLE 2: Names of membership functions, input, and output variables.

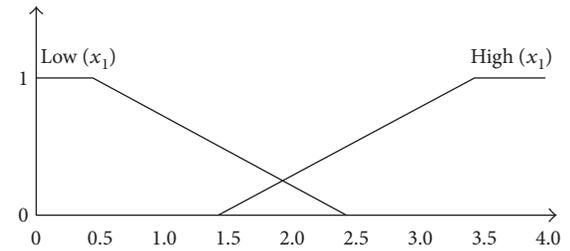
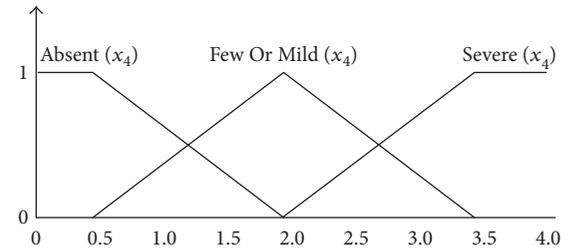
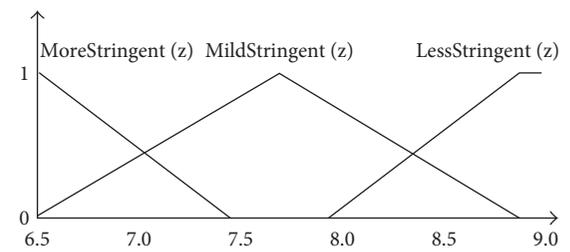
Variable	Name	Function 1	Function 2	Function 3
$x_1$	Risks potentially associated with hypoglycemia and other drug adverse effects	Low	High	—
$x_2$	Disease duration	Newly Diagnosed	Long Standing	—
$x_3$	Life expectancy	Long	Short	—
$x_4$	Important comorbidities	Absent	FewOrMild	Severe
$x_5$	Established vascular complications	Absent	FewOrMild	Severe
$x_6$	Patient attitude and expected treatment efforts	Highly Motivated	Less Motivated	—
$x_7$	Resources and support system	Readily Available	Limited	—
$z$	<i>HbA1c</i>	More Stringent	Mild Stringent	Less Stringent

levels, ranging from integers 0 to 4. The output value  $z$  is the ideal patient *HbA1c* target level which considers individual differences. The American Association of Clinical Endocrinologists (AACE) and American College of Endocrinology (ACE) suggested *HbA1c* below 6.5% [24], but patient-centered care is needed to consider the patient’s characteristics to set the patient’s *HbA1c* target. So, the output  $z$  represents the ideal *HbA1c* target, which varies between 6.5% and 9.0%.

The definition of membership functions is according to the ADA and EASD position statement [14], for example, the “Risks potentially associated with hypoglycemia and other drug adverse effects” can have two levels: “Low” or “High.” So,  $x_1$  has two membership functions:  $Low(x_1)$  and  $High(x_1)$ . The names of the membership functions as well as input and output variables are shown in Table 2. To get the acceptable results, the endocrinologist who works in Taichung Hospital, Ministry of Health and Welfare, has established virtual patients’ data and use the FuzzyLite [25] to trial and adjust the parameters of membership functions. Through the sufficient experience of the clinician, the system has better results.

Figure 2 shows the corresponding membership function for  $x_1$  factor, and the membership function for  $Low(x_1)$  and  $High(x_1)$  is a trapezoid. Because the  $x_2$ ,  $x_3$ ,  $x_6$ , and  $x_7$  also can be divided into two functions, their membership functions are the same as that for  $x_1$ . Figure 3 shows the corresponding membership function for  $x_4$  factor, and the membership function for  $Absent(x_4)$  and  $Severe(x_4)$  is trapezoidal while  $FewOrMild(x_4)$  is triangular. Because the  $x_5$  also can be divided into three functions,  $x_5$  membership functions are the same as  $x_4$ . The domain knowledge of the defining membership functions is derived from doctors’ reports. Figure 4 shows the corresponding membership function for  $z$  and the membership functions for  $MoreStringent(z)$  and  $LessStringent(z)$  are trapezoidal while that for  $MildStringent(z)$  is triangular.

The second step is to apply inputs to the fuzzy rules. The fuzzy inference will then stipulate what action will be taken for each combination of sets of memberships. To evaluate the effectiveness of the system, we developed two kinds of fuzzy rule methods. The primary consideration for the first method is relative safety of treatment so we label it “fuzzy safety rules.” The second method is to consider the

FIGURE 2: Membership functions of  $x_1$  factor.FIGURE 3: Membership functions of  $x_4$  factor.FIGURE 4: Membership functions of  $z$  factor.

performance of more positive treatment; we mark it “fuzzy positivity rules.”

**2.2.1. Method 1: Fuzzy Safety Rules.** The number of fuzzy rules depends on several input factors. For example, if the clinical doctor inputs  $\times 1$ ,  $\times 2$ , and  $\times 4$  values, the fuzzy rules will consist of 12 individual rules. Because  $\times 1$  has two membership functions (low and high),  $\times 2$  has two membership functions (Newly Diagnosed and Long Standing), and  $\times 4$

TABLE 3: Example of fuzzy safety rules table.

Rule	$x_1$	$x_2$	$x_4$	$z$
1	Low	Newly Diagnosed	Absent	More Stringent
2	Low	Newly Diagnosed	FewOrMild	Mild Stringent
3	Low	Newly Diagnosed	Severe	Less Stringent
4	Low	Long Standing	Absent	Less Stringent
5	Low	Long Standing	FewOrMild	Less Stringent
6	Low	Long Standing	Severe	Less Stringent
7	High	Newly Diagnosed	Absent	Less Stringent
8	High	Newly Diagnosed	FewOrMild	Less Stringent
9	High	Newly Diagnosed	Severe	Less Stringent
10	High	Long Standing	Absent	Less Stringent
11	High	Long Standing	FewOrMild	Less Stringent
12	High	Long Standing	Severe	Less Stringent

has three membership functions (Absent, Few or Mild, and Severe). Based on individual experts' experience and intuition, the fuzzy rules table is shown in Table 3. Rule 1 indicates that if  $x_1$  is low and  $x_2$  is newly diagnosed, and  $x_4$  is absent, then  $z$  is more stringent. Rule 2 states that if  $x_1$  is low and  $x_2$  is newly diagnosed, and  $x_4$  is few or mild, then  $z$  is mild stringent. Otherwise, the output  $z$  is less stringent in rules 3–12 because  $x_1$  is high, or  $x_2$  is long standing, or  $x_4$  is severe.

**2.2.2. Method 2: Fuzzy Positivity Fuzzy Rules.** Table 4 shows the fuzzy positivity rules. Rules 1–5 indicate that if one of the  $\{x_1, x_2, x_3, x_4, x_5\}$  input variables is High/Long-Standing/Short/Severe/Severe, then *HbA1c* is less stringent. Rules 6–9 indicate that if one of the  $\{x_2, x_3, x_6, x_7\}$  input variables is Newly-Diagnosed/Long/Highly-Motivated/Readily-Available, then *HbA1c* is more stringent. Rule 10 indicates that if the “Risks-Of-Hypoglycemia-or-Drug-Effects” are low and “Important-Comorbidities” and “Established-Vascular-Complications” are absent, then *HbA1c* is More-Stringent. Rules 11 and 12 show that if “Important-Comorbidities” or “Established-Vascular-Complications” are Few-or-Mild, then *HbA1c* is Mild-Stringent. Rule 13 states that if both of “Important-Comorbidities” and “Established-Vascular-Complications” are “Few-or-Mild,” then *HbA1c* is Less-Stringent.

Finally, for both fuzzy safety rules and fuzzy positivity rules, the system uses the mean of maximum (MOM) to perform defuzzification.

**2.3. Drug Knowledge Ontology and Reasoning Module.** Protégé and WebProtégé are free software programs for building ontology knowledge solutions [6, 26]. Further, “Jena” is the Java rule-based inference engine developed by Apache Software Foundation [27]. We use WebProtégé to build drug knowledge, and the web-based interface is an easy interface with a diabetes diplomate. When the ontology build is complete, we use Jena to evaluate the antidiabetic medications reasoning module. The details are as follows.

**2.3.1. Drug Knowledge Ontology.** According to an update of the position statement published by the ADA and the EASD

[13, 14], we created a glucose-lowering agents ontology. Table 5 shows the classes and the descriptions of their concepts in the domain knowledge. Classes can contain individual objects called “instances.” Table 6 presents the defined properties in the ontology. Object properties represent relationships between two instances and each property has a domain and range. After classes and object properties are created, we build glucose-lowering agent instances based on the ADA and the EASD’s position statement on the management of hyperglycemia in type 2 diabetes. Figure 5 shows “Biguanides” instances of the “Glucose-Lowering\_Agents” class and Figure 6 shows an example of patient\_1’s instance.

**2.3.2. Antidiabetic Medications Reasoning Module.** Jena is a free and open source Java framework for building semantic web and inference applications [27]. The Jena inference engines support the use of Jena rules to infer from instance data and class descriptions.

Jena is a rule inference engine running on the Java platform. This study developed Jena rules for reasoning which Glucose-Lowering\_Agents are not suitable for patients. Table 7 shows the rules described as follows:

Rule 1: If patients have a history of disease which is related to the disadvantages of Glucose-Lowering\_Agents, the Glucose-Lowering\_Agents are not recommended.

Rule 2: If patients have adverse drug reactions (ADRs), the ADRs are not recommended.

When the system removes some antidiabetic medications, the system can determine other antidiabetic medications. For example, patient\_1 has a history of “increasing\_LDL-C” and “Edema.” TZDs has both disadvantages, “increasing\_LDL-C” and “Edema,” one of SGLT2’s disadvantages is “increasing\_LDL-C.” By Rule 1, TZDs and SGLT2 will not be recommended to patient\_1. patient\_1 also has ADRs with GLP-1. Thus, by Rule 2, GLP-1 will not be recommended to patient\_1. This system provides seven common antidiabetic medications in Taiwan which include Biguanides (Metformin), Sulfonylureas (SU), TZDs, DPP-4, SGLT2, GLP-1, and Insulin. When TZDs, SGLT2, and GLP-1 are not recommended to patient\_1, Biguanides (Metformin), DPP-4, Sulfonylureas (SU), and Insulin are recommended.

**2.4. Define Risk of Antidiabetic Medications.** The American Association of Clinical Endocrinologists (AACE) and American College of Endocrinology (ACE) published an algorithm for determining glycemic control in 2009 [28]. The comprehensive diabetes management algorithm was updated in 2013, 2015, and 2016 [23, 24, 29–31]. One significant contribution was the presentation of antidiabetic medication profiles. It shows each drug’s properties considered for patients.

According to the antidiabetic medication profiles, we convert seven traditional antidiabetic drugs to present the antidiabetic medication risk values. Table 8 shows the risks

TABLE 4: Example of fuzzy positivity rules table.

Rule	Function
1	If (Risks-Of-Hypoglycemia-or-Drug-Effects is High) then ( <i>HbA1c</i> is Less-Stringent)
2	If (Disease-Duration is Long-Standing) then ( <i>HbA1c</i> is Less-Stringent)
3	If (Life-Expectancy is Short) then ( <i>HbA1c</i> is Less-Stringent)
4	If (Important-Comorbidities is Severe) then ( <i>HbA1c</i> is Less-Stringent)
5	If (Established-Vascular-Complications is Severe) then ( <i>HbA1c</i> is Less-Stringent)
6	If (Disease-Duration is Newly-Diagnosed) then ( <i>HbA1c</i> is More-Stringent)
7	If (Life-Expectancy is Long) then ( <i>HbA1c</i> is More-Stringent)
8	If (Patient-Attitude is Highly-Motivated) then ( <i>HbA1c</i> is More-Stringent)
9	If (Resources-and-Support-System is Readily-Available) then ( <i>HbA1c</i> is More-Stringent)
10	If (Risks-Of-Hypoglycemia-or-Drug-Effects is Low) and (Important-Comorbidities is Absent) and (Established-Vascular-Complications is Absent) then ( <i>HbA1c</i> is More-Stringent)
11	If (Important-Comorbidities is Few-or-Mild) then ( <i>HbA1c</i> is Mild-Stringent)
12	If (Established-Vascular-Complications is Few-or-Mild) then ( <i>HbA1c</i> is Mild-Stringent)
13	If (Important-Comorbidities is Few-or-Mild and (Established-Vascular-Complications is Few-or-Mild) then ( <i>HbA1c</i> is Less-Stringent)

TABLE 5: Classes in the domain ontology.

Class	Description
Glucose-Lowering_Agents	Concepts are glucose-lowering drugs. Ontology content is based on the ADA/EASD's position statement on management of hyperglycemia in type 2 diabetes to be established
Glucose-Lowering_Advantages	Concepts about glucose-lowering advantages
Glucose-Lowering_Cellular_mechanisms	Concepts about glucose-lowering cellular mechanisms
Glucose-Lowering_Compounds	Concepts about glucose-lowering compounds
Glucose-Lowering_Cost	Concepts about glucose-lowering cost
Glucose-Lowering_Disadvantages	Concepts about glucose-lowering disadvantages
Glucose-Lowering_Primary_physiological_actions	Concepts about glucose-lowering primary physiological actions
Patients	Concepts about patient's profile, the properties include patient's adverse drug reactions (ADRs) and history of diseases

TABLE 6: Defined properties in the ontology.

Property name	Property type	Domain	Range
has_Advantages	Object	Glucose-Lowering_Agents	Glucose-Lowering_Advantages
has_Cellular_mechanisms	Object	Glucose-Lowering_Agents	Glucose-Lowering_Cellular_mechanisms
has_Compounds	Object	Glucose-Lowering_Agents	Glucose-Lowering_Compounds
has_Cost	Object	Glucose-Lowering_Agents	Glucose-Lowering_Cost
has_Disadvantages	Object	Glucose-Lowering_Agents	Glucose-Lowering_Disadvantages
has_Primary_physiological_actions	Object	Glucose-Lowering_Agents	Glucose-Lowering_Primary_physiological_actions
has_History_of_Diseases	Object	Patients	Glucose-Lowering_Disadvantages
has_Adverse Drug_Reactions	Object	Patients	Glucose-Lowering_Agents
Not_recommended	Object	Patients	Glucose-Lowering_Agents
ID_No	Data	Patients	xsd: string

of antidiabetic medications. The risk value of "Few adverse events of possible benefits" is 1. The risk value of "Neutral" is 3 and "Use with caution" is 5. Finally, the "Likelihood of

adverse effects" is defined as a risk value of 7. Of note, we added "cost" property values according to the position statement of the ADA and the EASD [14]. If cost is "Low," the

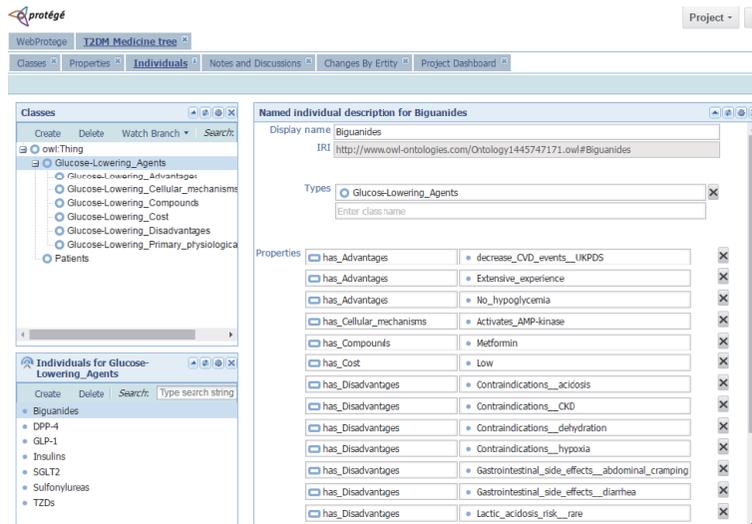


FIGURE 5: “Biguanides” instances of the “Glucose-Lowering\_Agents” class.

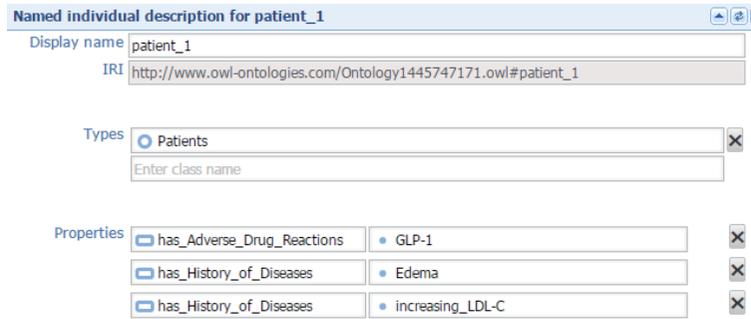


FIGURE 6: Example of “patient\_1.”

TABLE 7: Example of ontology reasoning rules table.

No.	Rule
(1)	$(?x \text{ rdf:type } \text{http://www.owl-ontologies.com/Ontology1445747171.owl\#Patients})$ $(?x \text{ http://www.owl-ontologies.com/Ontology1445747171.owl\#has\_History\_of\_Diseases } ?y)$ $(?n \text{ rdf:type } \text{http://www.owl-ontologies.com/Ontology1445747171.owl\#Glucose-Lowering\_Agents})$ $(?n \text{ http://www.owl-ontologies.com/Ontology1445747171.owl\#has\_Disadvantages } ?y) \rightarrow$ $(?x \text{ http://www.owl-ontologies.com/Ontology1445747171.owl\#Not\_recommand } ?n)$
(2)	$(?x \text{ rdf:type } \text{http://www.owl-ontologies.com/Ontology1445747171.owl\#Patients})$ $(?x \text{ http://www.owl-ontologies.com/Ontology1445747171.owl\#has\_Adverse\_Drug\_Reactions } ?n) \rightarrow$ $(?x \text{ http://www.owl-ontologies.com/Ontology1445747171.owl\#Not\_recommand } ?n)$

property value is 1. If cost is “High,” the property value is 3. In this case, there are some antidiabetic medications like “Meglitinides,” “ $\alpha$ -glucosidase inhibitors” whose cost is “Moderate” so the property value is 2. However, these antidiabetic medications are not popular in Taiwan, so they do show in Table 8.

2.5. *TOPSIS Multicriteria Decision Analysis.* When the risk of antidiabetic medications is known, we can use it to calculate the antidiabetic medication recommended priority. The Technique for Order of Preference by Similarity to

Ideal Solution (TOPSIS) implements a multicriteria decision which was developed by Hwang and Yoon in 1981 [4, 32–34]. TOPSIS was employed to decide antidiabetic medications ranking.

In previous calculations, the system recommended MET (Biguanides), DPP-4, SU (Sulfonylureas), and Insulin for patient\_1. Table 9 shows the risk of antidiabetic medications and cost for patient\_1. We will use the risk data of Table 9 as an example to explain the TOPSIS method.

The TOPSIS process of patient\_1 was carried out as follows.

TABLE 8: Risk of antidiabetic medications and cost.

Properties	Antidiabetic medications						
	MET	GLP-1	SGLT2	DPP-4	TZD	SU	Insulin
Hypo	3	3	3	3	3	7	7
Weight	1	1	1	3	5	7	7
Renal/GU	7	7	5	3	3	7	7
GI Sx	5	5	3	3	3	3	3
CHF	3	3	3	3	5	3	3
CVD	1	3	3	3	3	5	3
Bone	3	3	3	3	5	3	3
Cost	1	3	3	3	1	1	3

MET: metformin (Biguanides); SU: sulfonylureas; Hypo: hypoglycemia; GU: genitourinary; GI Sx: glycemic index symptom; CHF: congestive heart failure; CVD: cardiovascular diseases.

*Step 1: Create the Decision Matrix.* Create an evaluation matrix consisting of  $m$  alternatives and  $n$  criteria with the intersection of each alternative and criteria are given as  $A$ .

$$A = [A_{ij}]_{m \times n} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \tag{1}$$

$$i = 1, 2, \dots, m, j = 1, 2, \dots, n.$$

For example, the decision matrix of Table 9 is

$$A_{4 \times 8} = \begin{bmatrix} 3 & 1 & 7 & 5 & 3 & 1 & 3 & 1 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ 7 & 7 & 7 & 3 & 3 & 5 & 3 & 1 \\ 7 & 7 & 7 & 3 & 3 & 3 & 3 & 3 \end{bmatrix}. \tag{2}$$

*Step 2: Construct Normalized Decision Matrix.* The matrix  $A$  is then normalized to form the matrix:

$$R = [R_{ij}]_{m \times n} = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{bmatrix}, \tag{3}$$

where  $r_{ij} = a_{ij} / (\sqrt{\sum_{k=1}^m a_{kj}^2})$ ,  $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ .

$$R_{4 \times 8} = \begin{bmatrix} 0.279 & 0.096 & 0.560 & 0.693 & 0.500 & 0.151 & 0.500 & 0.224 \\ 0.279 & 0.289 & 0.240 & 0.416 & 0.500 & 0.452 & 0.500 & 0.671 \\ 0.650 & 0.674 & 0.560 & 0.416 & 0.500 & 0.754 & 0.500 & 0.224 \\ 0.650 & 0.674 & 0.560 & 0.416 & 0.500 & 0.452 & 0.500 & 0.671 \end{bmatrix}. \tag{4}$$

*Step 3: Determine the Weight.* Determine the weight  $W$  with the antidiabetic medication's risk properties and cost. The initial weight  $W = [w_1 w_2 \cdots w_n] = [1 1 1 1 1 1 1 1]$ ;  $w_1$  is weight of the Hypo property;  $w_2$  is weight of Weight property;  $w_3$  is weight of Renal/GU property;  $w_4$  is weight of GI Sx property;  $w_5$  is weight of CHF property;  $w_6$  is weight of CVD property;  $w_7$  is weight of Bone property; and  $w_8$  is weight of Cost property.

In this case, patient\_1 has a history of "Edema" and "increasing LDL-C." Because "Edema" is relative to CHF, the  $w_5$  is set to 2, and "increasing LDL-C" is relative to CVD, the  $w_6$  set to 2. The weight  $W$  of patient\_1 is shown as follows:

$$W = [w_1 w_2 w_3 w_4 w_5 w_6 w_7 w_8] = [1 1 1 1 2 2 1 1]. \tag{5}$$

*Step 4: Construct the Weighted Normalized Decision Matrix.* Calculate the weighted normalized decision matrix  $V$ :

$$V = [V_{ij}]_{m \times n} = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{bmatrix}, \tag{6}$$

where  $v_{ij} = w_j r_{ij}$ ,  $i = 1, 2, \dots, m$ , and  $j = 1, 2, \dots, n$ .

$$V_{4 \times 8} = \begin{bmatrix} 0.279 & 0.096 & 0.560 & 0.693 & 1.000 & 0.302 & 0.500 & 0.224 \\ 0.279 & 0.289 & 0.240 & 0.416 & 1.000 & 0.905 & 0.500 & 0.671 \\ 0.650 & 0.674 & 0.560 & 0.416 & 1.000 & 1.508 & 0.500 & 0.224 \\ 0.650 & 0.674 & 0.560 & 0.416 & 1.000 & 0.905 & 0.500 & 0.671 \end{bmatrix}. \tag{7}$$

*Step 5: Determine the Ideal and Negative Ideal Solutions.* Determine the ideal solution  $A^*$  and the negative ideal solution  $A^-$ :

$$A^* = \{V_1^*, V_2^*, \dots, V_n^*\}, \quad (8)$$

where  $V_j^* = \min_i V_{ij}, i = 1, 2, \dots, m$ .

$$A^- = \{V_1^-, V_2^-, \dots, V_n^-\}, \quad (9)$$

where  $V_j^- = \max_i V_{ij}, i = 1, 2, \dots, m$ .

$$\begin{aligned} A^* &= \left( \min_i v_{i1}, \min_i v_{i2}, \min_i v_{i3}, \min_i v_{i4}, \min_i v_{i5}, \min_i v_{i6}, \min_i v_{i7}, \min_i v_{i8} \right) \\ &= (0.279, 0.096, 0.240, 0.416, 1.000, 0.302, 0.500, 0.224), \\ A^- &= \left( \max_i v_{i1}, \max_i v_{i2}, \max_i v_{i3}, \max_i v_{i4}, \max_i v_{i5}, \max_i v_{i6}, \max_i v_{i7}, \max_i v_{i8} \right) \\ &= (0.650, 0.674, 0.560, 0.693, 1.000, 1.508, 0.500, 0.671). \end{aligned} \quad (10)$$

*Step 6: Calculate the Separation Measures for Each Alternative.* Calculate the distance between the target alternative  $i$  and ideal alternative  $S_i^*$  and the negative ideal alternative  $S_i^-$ :

$$\begin{aligned} S_i^* &= \sqrt{\sum_{j=1}^n (V_{ij} - V_j^*)^2}, \quad i = 1, 2, \dots, m, \\ S_i^- &= \sqrt{\sum_{j=1}^n (V_{ij} - V_j^-)^2}, \quad i = 1, 2, \dots, m. \end{aligned} \quad (11)$$

From the above formula, the system can find the values of  $S_i^*$  and  $S_i^-$  as follows:

$$\begin{aligned} S_1^* &= 0.424, \\ S_2^* &= 0.775, \\ S_3^* &= 1.424, \\ S_4^* &= 1.067, \\ S_1^- &= 1.458, \\ S_2^- &= 0.911, \\ S_3^- &= 0.526, \\ S_4^- &= 0.664. \end{aligned} \quad (12)$$

*Step 7: Calculate the Relative Closeness to the Ideal Solution  $C_i^*$ .*

$$C_i^* = \frac{S_i^-}{S_i^* + S_i^-}, \quad i = 1, 2, \dots, m. \quad (13)$$

From the above formula, the system will select the options with  $C_i^*$  closest to 1.

$$\begin{aligned} C_1^* &= 0.775, \\ C_2^* &= 0.540, \\ C_3^* &= 0.270, \\ C_4^* &= 0.384. \end{aligned} \quad (14)$$

The results show that the ideal solution  $C_1^*$  of MET is 0.775, the ideal solution  $C_2^*$  of DPP-4 is 0.540, the ideal solution  $C_3^*$  of SU is 0.270, and the ideal solution  $C_4^*$  of the Insulin is 0.384. Because  $C_1^* > C_2^* > C_4^* > C_3^*$ , the recommendation priority of antidiabetic medications is MET > DPP-4 > Insulin > SU.

### 3. Results

According to the ADA and the EASD's statement, one of the major changes in treatment options is a new antidiabetic medication "SGLT2" which is added. So, we add "SGLT2" to this experiment. However, "Meglitinides" and " $\alpha$ -Glucosidase" have "Frequent dosing schedule" disadvantage, so we exclude those two antidiabetic medications. Finally, this system provides seven common antidiabetic medication choices which include "Biguanides," "SU," "TZDs," "DPP-4," "SGLT2," "GLP-1," and "Insulin."

Taichung Hospital is an accredited area hospital in central Taiwan. At the beginning of the system development, the endocrinologist who works in Taichung Hospital set up ten virtual patients' medical data to trial and adjusted clinical decision support system (CDSS). We used Mamdani-type fuzzy inference and mean of maximum (MOM) to perform defuzzification. Table 10 shows the virtual patient's medical data. In Table 10, the  $x_1, x_2, x_3, \dots, x_7$ , "History of Diseases" and "ADRs" are input variables for diabetes diplomat. The  $z$  is the fuzzy inference output of the  $HbA1c$  target which, respectively, are fuzzy safety rules (Method 1) and fuzzy positivity rules (Method 2). The "Recommend antidiabetic medications" shows the recommended medications for patients and the ranking value.

An attending physician, an endocrinologist, and a resident physician evaluated the decision support system for diabetes. They were direct using the system and evaluation it by their clinical experience. They evaluated the system using a 12-question, 5-point survey, regarding satisfaction degree, perceived usefulness, and behavioral intentions

TABLE 9: Risk of antidiabetic medications and cost for patient\_1.

Antidiabetic medications	Properties							Cost
	Hypo	Weight	Renal/GU	GI Sx	CHF	CVD	Bone	
MET	3	1	7	5	3	1	3	1
DPP-4	3	3	3	3	3	3	3	3
SU	7	7	7	3	3	5	3	1
Insulin	7	7	7	3	3	3	3	3

TABLE 10: Ten virtual patients' medical data.

ID	Age	Sex	$x_1, x_2, x_3, x_4, x_5, x_6, x_7$	has_History of Diseases	has_ADRs	$z$ (method 1)	$z$ (method 2)	Recommended antidiabetic medications
1	73	Female	3, 2, 3, NaN, NaN, NaN	increasing_LDL-C, Edema	GLP-1	8.6	8.9	(1) Biguanides (0.775) (2) DPP-4 (0.540) (3) Insulins (0.384) (4) Sulfonylureas (0.270)
2	75	Female	3, 2, 4, NaN, NaN, NaN	Heart_failure, increasing_LDL-C	NaN	8.6	9.0	(1) Biguanides (0.788) (2) DPP-4 (0.549) (3) GLP-1 (0.536) (4) Insulins (0.376) (5) Sulfonylureas (0.248)
3	64	Female	2, 1, 2, NaN, NaN, NaN	Bone_fractures, increasing_LDL-C	NaN	6.9	6.6	(1) Biguanides (0.788) (2) DPP-4 (0.549) (3) GLP-1 (0.536) (4) Insulins (0.376) (5) Sulfonylureas (0.248)
4	76	Female	4, 3, 3, 2, 1, NaN, NaN	increasing_LDL-C, Contraindications_CKD	DPP-4	8.8	7.8	(1) GLP-1 (0.631) (2) Insulins (0.445) (3) Sulfonylureas (0.369)
5	61	Female	4, 3, 2, 3, 2, NaN, NaN	Heart_failure, increasing_LDL-C, Contraindications_CKD, Weight_gain	NaN	8.6	7.8	(1) GLP-1 (0.534) (2) DPP-4 (0.466)
6	64	Female	2, 1, 1, NaN, NaN, 2, NaN	NaN	NaN	6.9	6.6	(1) Biguanides (0.731) (2) SGLT2 (0.648) (3) DPP-4 (0.619) (4) GLP-1 (0.586) (5) TZDs (0.549) (6) Sulfonylureas (0.377) (7) Insulins (0.365)
7	62	Male	2, 2, 3, NaN, NaN, 3, 1	Gastrointestinal_side_effects_abdominal_cramping, increasing_LDL-C	NaN	8.6	6.6	(1) DPP-4 (0.703) (2) GLP-1 (0.543) (3) Insulins (0.481) (4) Sulfonylureas (0.417)
8	81	Female	4, 3, 4, 4, 4, 4, 2	MI, increasing_LDL-C, Contraindications_CKD	DPP-4	8.6	9.0	(1) GLP-1 (0.631) (2) Insulins (0.445) (3) Sulfonylureas (0.369)
9	48	Female	1, 1, 2, 3, NaN, NaN, 1	Patient_reluctance_about_injection, increasing_LDL-C	NaN	7.9	6.6	(1) Biguanides (0.796) (2) GLP-1 (0.560) (3) DPP-4 (0.558) (4) Sulfonylureas (0.248)
10	56	Male	NaN, 2, 2, 2, 1, 1, NaN	Weight_gain, increasing_LDL-C, Gastrointestinal_side_effects_nausea	TZDs	7.9	7.8	(1) Biguanides (0.687) (2) DPP-4 (0.313)

(see Table 11). All the scores are expressed as percentage. The evaluation results are shown in Table 12. The clinical decision support system (CDSS) perceived 73% satisfactions.

The results of antidiabetic medication recommendation indicate that the system has 70% satisfaction and 71% has intentions to use it.

TABLE 11: Survey of “Patient-Centered Treatment Decision Support System for Diabetes Based on Fuzzy Logic and Domain Ontology”.

Question	Scoring				
	1	2	3	4	5
What do you think about “Patient ideal <i>HbA1c</i> target inference”?					
Are you satisfied with its accuracy?					
Is the “Method 1: Safety fuzzy rules” accurate?	<input type="checkbox"/>				
Are you satisfied with the results of the “Method 1: Safety fuzzy rules”?	<input type="checkbox"/>				
Is the “Method 2: Positivity fuzzy rules” accurate?	<input type="checkbox"/>				
Are you satisfied with the results of the “Method 2: Positivity fuzzy rules”?	<input type="checkbox"/>				
What do you think about “Antidiabetic medications reasoning and ranking”?					
Are you satisfied with its accuracy?					
Is the “Antidiabetic medications reasoning and ranking” accurate?	<input type="checkbox"/>				
Are you satisfied with the results of the “Antidiabetic medications reasoning and ranking”?	<input type="checkbox"/>				
Do you think the system can provide some benefits for you?					
Using the system improves my performance in my job.	<input type="checkbox"/>				
Using the system enhances my effectiveness in my job.	<input type="checkbox"/>				
I find the system to be useful in my job.	<input type="checkbox"/>				
If this system used in conjunction with the actual work, would you continue to use this system at work?					
I enjoy using this system at work.	<input type="checkbox"/>				
I will frequently use this system in the future.	<input type="checkbox"/>				
I will strongly recommend to others to use this system.	<input type="checkbox"/>				

Title: ○Endocrinologists ○Attending physicians ○Resident physicians

Gender: ○Male ○Female

E-mail:

TABLE 12: The evaluation result of the system.

Scores	Participants	
	Endocrinologist Attending physician Resident physician	Endocrinologist Attending physician
“ <i>HbA1c</i> target inference (Safety fuzzy rules: Method 1)” Satisfaction degree (%)	67%	80%
“ <i>HbA1c</i> target inference (Positivity fuzzy rules: Method 2)” Satisfaction degree (%)	67%	60%
“Antidiabetic medications reasoning and ranking” Satisfaction degree (%)	70%	85%
Perceived usefulness (%)	73%	87%
Intentions to use (%)	71%	77%

According to the feedback of resident physicians, because the inpatients may have too many complications, so the resident physicians mostly use insulin to control *HbA1c*. Therefore, the evaluation results of the resident physician will be relatively weak; this is because the recommendation system is only suitable for outpatients. So, if we exclude the results of the assessment of the resident physicians, the evaluation results of the system will be better. In this situation, the participating clinicians have 87% acceptance, and the likelihood of using the system at work and recommending it to others is 77%.

The fuzzy safety rules (Method 1) has 80% accuracy and satisfaction, but the fuzzy positivity rules (Method 2) is only 60%. So, the fuzzy safety rules (Method 1) is better than the fuzzy positivity rules (Method 2) for the patient ideal *HbA1c* target inference. The evaluation result of “Antidiabetic medications reasoning and ranking” shows that the system has

85% satisfaction which can assist clinicians to the management of T2DM while selecting antidiabetic medications.

The user interface for the antidiabetic medication recommendation system is shown in Figure 7. The website of the system is <http://120.109.46.42/T2DM/>. Doctors may refer to the system to make prescriptions. Of course, the system, which is likely to make misleading or inappropriate suggestions, cannot replace a doctor’s clinical experience and professional judgment. The doctor thus makes the final decision.

#### 4. Discussion

The number of patients with diabetes worldwide is significant and continually increasing. Diabetes imposes psychological, physical, and financial hardship on patients. Diabetes therapy, no doubt, is a complicated task. As regards the prescription strategy of clinical doctors, it is necessary that they consider

 A Patient-Centered Treatment Decision Support System for Diabetes Based on Fuzzy Logic and Domain Ontology

Please take a few minutes to [fill out the survey](#) concerning your vts/interaction with the system. Your responses will help us to improve the quality of our system.

**Patient Data**

Patient ID No.

1

**Modulation of the intensiveness of glucose lowering in T2DM (HbA1c target)**

Item	1	2	3	4	5	
Risks Of Hypoglycemia or Drug Effects	low	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	high
Disease Duration	newly diagnosed	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	long-standing
Life Expectancy	long	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	short
Important Comorbidities	absent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	severe
Established Vascular Complications	absent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	severe
Patient Attitude	highly motivated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	less motivated
Resources and Support System	readily available	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	limited

**Adverse Drug Reactions ADRs**

Biguanides

Sulfonylureas(SU)

TZDs

DPP-4

SGLT2

GLP-1

Insulins

**History**

GI Sx.	CVD	RENAL/GU	Others
<input type="checkbox"/> Gastrointestinal side effects:diarrhea	<input type="checkbox"/> Blunts myocardial ischemic preconditioning	<input type="checkbox"/> Increasing Creatinine(transient)	<input type="checkbox"/> Low durability
<input type="checkbox"/> Gastrointestinal side effects:abdominal cramping	<input type="checkbox"/> Increasing Heart rate	<input type="checkbox"/> Lactic acidosis risk(rare)	<input type="checkbox"/> Angioedema/urticaria and other immune-mediated dermatological effects
<input type="checkbox"/> Gastrointestinal side effects:nausea	<input checked="" type="checkbox"/> Increasing LDL-C	<input type="checkbox"/> Contraindications:CKD	<input type="checkbox"/> C-cell hyperplasia/medullary thyroid tumors in animals
<input type="checkbox"/> Gastrointestinal side effects:vomiting	<input type="checkbox"/> MI	<input type="checkbox"/> Contraindications:acidosis	<input type="checkbox"/> Ingestible
<input type="checkbox"/> Acute pancreatitis	<input type="checkbox"/> Volume depletion/hypotension/dizziness	<input type="checkbox"/> Genitourinary infections	<input type="checkbox"/> Training requirements
<input type="checkbox"/> Vitamin B12 deficiency	<input type="checkbox"/> Contraindications:hypoxia	<input type="checkbox"/> Polyuria	<input type="checkbox"/> Mitogenic effects
<b>CHF</b>	<input type="checkbox"/> Contraindications:dehydration	<b>Weight</b>	<input type="checkbox"/> Patient reluctance about injection
<input checked="" type="checkbox"/> Edema	<b>Hypo</b>	<input type="checkbox"/> Weight gain	
<input type="checkbox"/> Heart failure	<input type="checkbox"/> Hypoglycemia	<b>Bone</b>	
<input type="checkbox"/> Heart failure hospitalizations		<input type="checkbox"/> Bone fractures	

Reasoning    Reset

**Patient Ideal HbA1c target inference**

**Method 1: Safety fuzzy rules**

HbA1c Target: 8.6

**Method 2: Positivity fuzzy rules**

HbA1c Target: 8.9

**Antidiabetic medications reasoning and ranking**

Antidiabetic medications (The relative closeness to the ideal solution):

1. Biguanides (0.775)
2. DPP-4 (0.540)
3. Insulins (0.384)
4. Sulfonylureas (0.270)

FIGURE 7: User interface for the antidiabetic medication recommendation system.

many factors. However, the following two reasons will affect the doctor's decision to use the system:

- (1) If a doctor uses the new and expensive drugs as a treatment prescription, he will worry that the health insurance will not pay medical expenses. Then, he will only use the generic antidiabetic drugs.
- (2) The system only provides a single drug treatment prescription, for long-term diabetes patients may need a multidrug treatment prescription to reduce HbA1c effectively.

Even so, CDSS is used to assist humans in making decisions rather than replacing human decisions. The system shows the following clinical values:

- (a) Define appropriate therapeutic goals implementing patient-centered medical care and prescriptions:

The patient-centered management strategy, by contrast, holds that not all patients can benefit from active glucose management. It stresses individualized therapeutic goals. However, diabetes, multiple complications, and the complexity inherent in antidiabetic medication use often make it difficult for doctors, especially young doctors, to select the best therapeutic strategy. Despite their awareness of the concept of "patient-centered management strategy," it has shown the difficulty in practice. Given this, we systematized the constructs to help doctors develop their

therapeutic goals and selection of prescriptions to meet the patient's needs. In addition to encouraging patients to follow doctors' instructions, this method can also reduce the risks resulting from medical treatment. Therapeutic goals may thus be achieved.

- (b) Doctors can save time and make the best use of medical resources:

The increasing number of patients with type 2 diabetes has been exhausting medical resources. This system can enable doctors to spend less time on medical diagnosis and adjustment of patients' prescriptions. This will reduce the impact on health care resources.

- (c) Doctors can employ the system with ease, and their clinical inertia can be reduced:

This system is manipulation-friendly. By inputting a few needed parameters, doctors can obtain recommended antidiabetic medications in order of effectiveness and thus make their treatment judgment accordingly. The system enables physicians to save time in answering patients' questions and can reduce the risk of developing clinical inertia.

## 5. Conclusion and Future Work

The prescription strategy of clinical doctors must take many factors into account. To address this, we developed an

individualized antidiabetic medication recommendation system for patients with diabetes. This system, which can be manipulated with relative ease, tailors *HbA1c* levels to mitigate patients' differences. Currently, 12 kinds of antidiabetic medications, both oral and injected, are available. Manually considering all possible conditions is not only a waste of medical resources but also a burden on the system, not to mention that it is impractical. This study, which combines fuzzy logic and ontology reasoning, proposes an antidiabetic medication recommendation system for patients with diabetes. It promotes a new concept of "patient-centered diabetes therapy." Antidiabetic medications are recommended for the outpatients, and useful ranking of medications is conducted. In addition to aiding doctors' clinical diagnosis, the system can not only serve as a guide for doctors specializing in diabetes but also help family practitioners and interns in prescribing medications.

Based on results of the study using the feedback system of operations, the seven factors analyzed can provide dynamic correlations. We will improve our system interface and dynamic weighting calculations in future research. Besides, we propose an architecture based on rules to build an antidiabetic medication recommendation system. In the future, we will combine rule-based and case-based reasoning to solve the special case issues.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the Ministry of Science and Technology, Taiwan (Grant nos. 103-2632-E-324-001-MY3, MOST-104-2221-E-324-019-MY2, MOST-106-2221-E-324-025, and MOST-106-2218-E-324-002).

## References

- [1] Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century*, The National Academies Press, Washington, DC, USA, 2001.
- [2] C.-T. Bau, R.-C. Chen, and C.-Y. Huang, "Construction of a clinical decision support system for undergoing surgery based on domain ontology and rules reasoning," *Telemedicine and e-Health*, vol. 20, no. 5, pp. 460–472, 2014.
- [3] F. J. Ampudia-Blasco, P. Y. Benhamou, G. Charpentier et al., "A decision support tool for appropriate glucose-lowering therapy in patients with type 2 diabetes," *Diabetes Technology & Therapeutics*, vol. 17, no. 3, pp. 194–202, 2014.
- [4] C.-L. Hwang, Y.-J. Lai, and T.-Y. Liu, "A new approach for multiple objective decision making," *Computers & Operations Research*, vol. 20, no. 8, pp. 889–899, 1993.
- [5] R. F. Alharbi, J. Berri, and S. El-Masri, "Ontology based clinical decision support system for diabetes diagnostic," in *2015 Science and Information Conference (SAI)*, pp. 597–602, London, UK, 2015.
- [6] Stanford Center for Biomedical Informatics Research (BMIR), "Protégé," June 2017, <http://protege.stanford.edu>.
- [7] R.-C. Chen, Y.-H. Huang, C.-T. Bau, and S.-M. Chen, "A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection," *Expert Systems with Applications*, vol. 39, no. 4, pp. 3995–4006, 2012.
- [8] P. Sharma and P. D. Kaur, "Effectiveness of web-based social sensing in health information dissemination—a review," *Telematics and Informatics*, vol. 34, no. 1, pp. 194–219, 2017.
- [9] Y.-F. Zhang, Y. Tian, T.-S. Zhou, K. Araki, and J.-S. Li, "Integrating HL7 RIM and ontology for unified knowledge and data representation in clinical decision support systems," *Computer Methods and Programs in Biomedicine*, vol. 123, pp. 94–108, 2016.
- [10] D. E. Forbes, P. Wongthongtham, J. Singh, and S. C. Thompson, "Ontology supported assistive communications in healthcare," *Communications of the Association for Information Systems*, vol. 34, pp. 297–322, 2014.
- [11] P. C. Sherimon and R. Krishnan, "OntoDiabetic: an ontology-based clinical decision support system for diabetic patients," *Arabian Journal for Science and Engineering*, vol. 41, no. 3, pp. 1145–1160, 2016.
- [12] L. Marco-Ruiz, C. Pedrinaci, J. A. Maldonado, L. Panziera, R. Chen, and J. G. Bellika, "Publication, discovery and interoperability of clinical decision support systems: a linked data approach," *Journal of Biomedical Informatics*, vol. 62, pp. 243–264, 2016.
- [13] S. E. Inzucchi, R. M. Bergenstal, J. B. Buse et al., "Management of hyperglycemia in type 2 diabetes: a patient-centered approach: position statement of the American Diabetes Association (ADA) and the European Association for the Study of diabetes (EASD)," *Diabetes Care*, vol. 35, no. 6, pp. 1364–1379, 2012.
- [14] S. E. Inzucchi, R. M. Bergenstal, J. B. Buse et al., "Management of hyperglycemia in type 2 diabetes, 2015: a patient-centered approach: update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes," *Diabetes Care*, vol. 38, no. 1, pp. 140–149, 2015.
- [15] C.-H. Chang, Y.-D. Jiang, C.-H. Chung, L.-T. Ho, and L.-M. Chuang, "National trends in anti-diabetic treatment in Taiwan, 2000–2009," *Journal of the Formosan Medical Association*, vol. 111, no. 11, pp. 617–624, 2012.
- [16] H.-T. Ou, K.-C. Chang, Y.-M. Liu, and J.-S. Wu, "Recent trends in the use of antidiabetic medications from 2008 to 2013: a nation-wide population-based study from Taiwan," *Journal of Diabetes*, vol. 9, no. 3, pp. 256–266, 2016.
- [17] T. Kohro, T. Yamazaki, H. Sato et al., "Trends in antidiabetic prescription patterns in Japan from 2005 to 2011," *International Heart Journal*, vol. 54, no. 2, pp. 93–97, 2013.
- [18] C. Rafaniello, V. Arcoraci, C. Ferrajolo et al., "Trends in the prescription of antidiabetic medications from 2009 to 2012 in a general practice of Southern Italy: a population-based study," *Diabetes Research and Clinical Practice*, vol. 108, no. 1, pp. 157–163, 2015.
- [19] S.-H. Ko, D.-J. Kim, J.-H. Park et al., "Trends of antidiabetic drug use in adult type 2 diabetes in Korea in 2002–2013: nationwide population-based cohort study," *Medicine*, vol. 95, no. 27, article e4018, 2016.
- [20] A. S. Abdelmoneim, D. T. Eurich, J.-M. Gamble, and S. H. Simpson, "Use patterns of antidiabetic regimens by patients with type 2 diabetes," *Canadian Journal of Diabetes*, vol. 37, no. 6, pp. 394–400, 2013.

- [21] J. C. Hsu, D. Ross-Degnan, A. K. Wagner et al., "Utilization of oral antidiabetic medications in Taiwan following strategies to promote access to medicines for chronic diseases in community pharmacies," *Journal of Pharmaceutical Policy and Practice*, vol. 8, no. 1, p. 15, 2015.
- [22] R.-C. Chen, C.-Y. Huang, C.-T. Bau, and L.-S. Chen, "A decision support system for diabetes medicine selection using patient centered treatment based on fuzzy logic and domain ontology," *International Journal of Innovative Computing, Information and Control*, vol. 13, no. 5, pp. 1681–1692, 2017.
- [23] A. J. Garber, M. J. Abrahamson, J. I. Barzilay et al., "Consensus statement by the American Association of Clinical Endocrinologists and American College of endocrinology on the comprehensive type 2 diabetes management algorithm – 2016 executive summary," *Endocrine Practice*, vol. 22, no. 1, pp. 84–113, 2016.
- [24] Y. Handelsman, Z. T. Bloomgarden, G. Grunberger et al., "American Association of Clinical Endocrinologists and American College of Endocrinology - clinical practice guidelines for developing a diabetes mellitus comprehensive care plan - 2015," *Endocrine Practice*, vol. 21, Supplement 1, pp. 1–87, 2015.
- [25] J. Rada-Vilela, "FuzzyLite: a fuzzy logic control library," June 2017, <http://www.fuzzylite.com>.
- [26] Stanford Center for Biomedical Informatics Research (BMIR), "WebProtégé," June 2017, <http://protege.stanford.edu/products.php#web-protege>.
- [27] Apache Software Foundation, "Apache Jena," June 2017, <http://jena.apache.org>.
- [28] H. Rodbard, P. Jellinger, J. Davidson et al., "Statement by an American Association of Clinical Endocrinologists/American College of Endocrinology Consensus Panel on type 2 diabetes mellitus: an algorithm for glycemic control," *Endocrine Practice*, vol. 15, no. 6, pp. 540–559, 2009.
- [29] A. J. Garber, M. J. Abrahamson, J. I. Barzilay et al., "AAACE/ACE Comprehensive diabetes management algorithm 2015," *Endocrine Practice*, vol. 21, no. 4, pp. 438–447, 2015.
- [30] Y. Handelsman, Z. T. Bloomgarden, G. Grunberger et al., "American Association of Clinical Endocrinologists and American College of Endocrinology - clinical practice guidelines for developing a diabetes mellitus comprehensive care plan - 2015 - executive summary," *Endocrine Practice*, vol. 21, no. 4, pp. 413–437, 2015.
- [31] A. Garber, M. Abrahamson, J. Barzilay et al., "American Association of Clinical Endocrinologists' comprehensive diabetes management algorithm 2013 consensus statement," *Endocrine Practice*, vol. 19, Supplement 2, pp. 1–48, 2013.
- [32] C.-L. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications: A State-of-the-Art Survey. Lecture Notes in Economics and Mathematical Systems*, vol. 186, Springer-Verlag, Berlin, Heidelberg, 1981.
- [33] Wikipedia, "TOPSIS," June 2017, <https://en.wikipedia.org/wiki/TOPSIS>.
- [34] S.-S. Chang, *Fuzzy Multi-Criteria Decision Making for Evaluation Method*, Wu-Nan Book Inc., Taiwan, 2012.