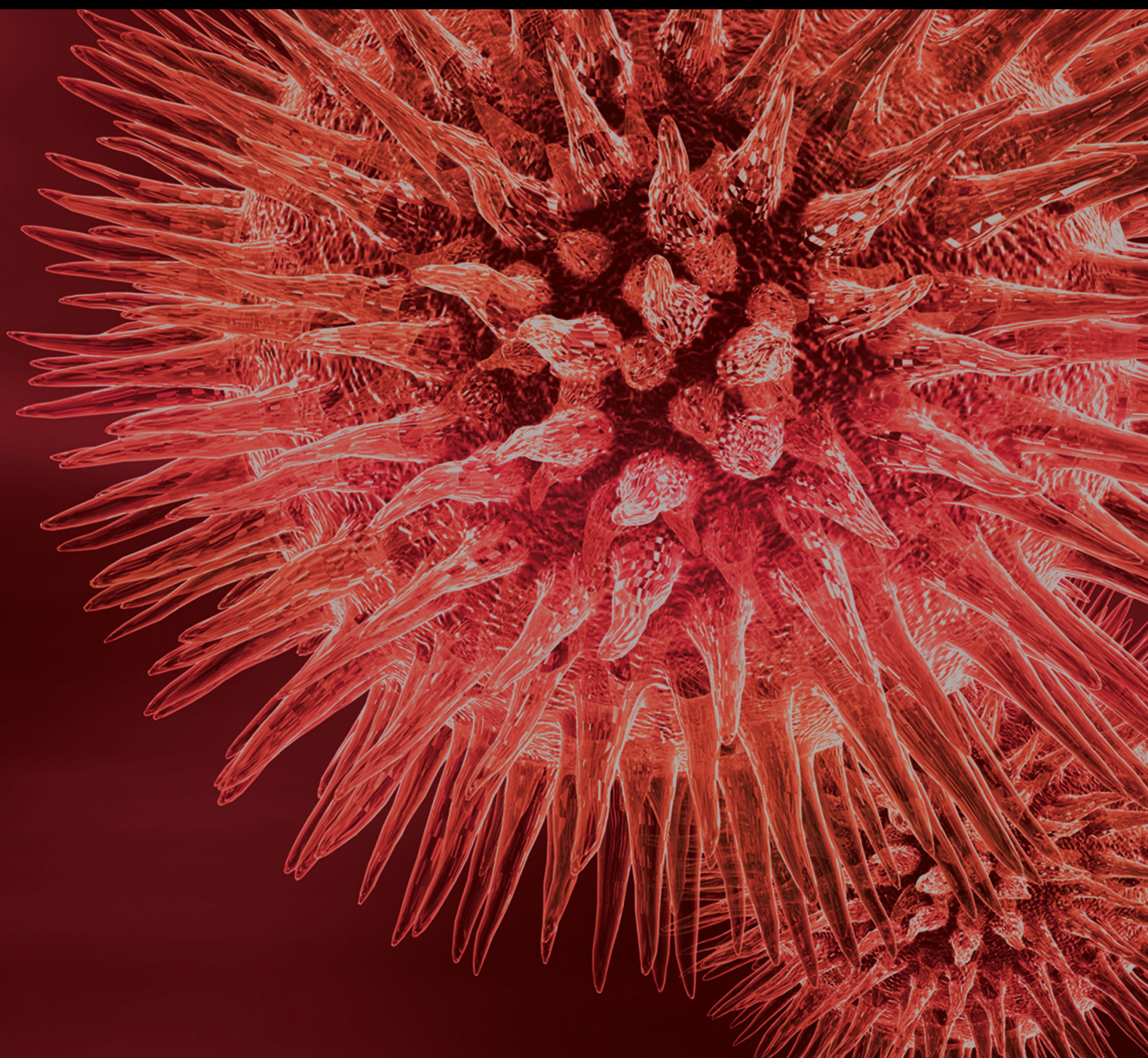


BioMed Research International

Pattern Recognition in Medical Decision Support

Lead Guest Editor: Shadnaz Asgari

Guest Editors: Fabien Scalzo and Magdalena Kasprowicz





Pattern Recognition in Medical Decision Support

BioMed Research International

Pattern Recognition in Medical Decision Support

Lead Guest Editor: Shadnaz Asgari

Guest Editors: Fabien Scalzo and Magdalena Kasprowicz




Copyright © 2019 Hindawi. All rights reserved.


This is a special issue published in “BioMed Research International.” 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.

Contents



Pattern Recognition in Medical Decision Support

Shadnaz Asgari , Fabien Scalzo, and Magdalena Kasprowicz
Editorial (2 pages), Article ID 6048748, Volume 2019 (2019)

Decision Support Systems in Prostate Cancer Treatment: An Overview

Y. van Wijk , I. Halilaj, E. van Limbergen, S. Walsh, L. Lutgens, P. Lambin, and B. G. L. Vanneste 
Review Article (10 pages), Article ID 4961768, Volume 2019 (2019)





Computational Analysis of Protein Structure Changes as a Result of Nondeletion Insertion Mutations in Human β -Globin Gene Suggests Possible Cause of β -Thalassemia

Talal Qadah  and Mohammad Sarwar Jamal 
Research Article (8 pages), Article ID 9210841, Volume 2019 (2019)



Automatic Detection of Hard Exudates in Color Retinal Images Using Dynamic Threshold and SVM Classification: Algorithm Development and Evaluation

Shengchun Long , Xiaoxiao Huang , Zhiqing Chen, Shahina Pardhan , and Dingchang Zheng 
Research Article (13 pages), Article ID 3926930, Volume 2019 (2019)

A Study of Machine-Learning Classifiers for Hypertension Based on Radial Pulse Wave

Zhi-yu Luo, Ji Cui, Xiao-juan Hu , Li-ping Tu, Hai-dan Liu, Wen Jiao, Ling-zhi Zeng, Cong-cong Jing, Li-jie Qiao, Xu-xiang Ma, Yu Wang, Jue Wang, Ching-Hsuan Pai, Zhen Qi , Zhi-feng Zhang , and Jia-tuo Xu 
Research Article (12 pages), Article ID 2964816, Volume 2018 (2019)




Computer Aided Diagnosis System for Detection of Cancer Cells on Cytological Pleural Effusion Images

Khin Yadanar Win , Somsak Choomchuay, Kazuhiko Hamamoto, Manasanan Raveesunthornkiat , Likit Rangsirattanakul, and Suriya Pongsawat
Research Article (21 pages), Article ID 6456724, Volume 2018 (2019)

A Deep Learning Approach to Vascular Structure Segmentation in Dermoscopy Colour Images

Joanna Jaworek-Korjakowska 
Research Article (8 pages), Article ID 5049390, Volume 2018 (2019)

PCG Classification Using Multidomain Features and SVM Classifier

Hong Tang , Ziyin Dai , Yuanlin Jiang , Ting Li, and Chengyu Liu 
Research Article (14 pages), Article ID 4205027, Volume 2018 (2019)

Editorial

Pattern Recognition in Medical Decision Support

Shadnaz Asgari ^{1,2} **Fabien Scalzo**^{3,4} and **Magdalena Kasprowicz**⁵

¹*Biomedical Engineering Department, California State University, Long Beach, USA*

²*Computer Engineering and Computer Science Department, California State University, Long Beach, USA*

³*Department of Neurology, University of California, Los Angeles, USA*

⁴*Department of Computer Science, University of California, Los Angeles, USA*

⁵*Department of Biomedical Engineering, Wrocław University of Science and Technology, Wrocław, Poland*

Correspondence should be addressed to Shadnaz Asgari; shadnaz.asgari@csulb.edu

Received 27 May 2019; Accepted 27 May 2019; Published 13 June 2019

Copyright © 2019 Shadnaz Asgari 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.

Recent advances in data acquisition and various monitoring modalities have resulted in generating and collecting a growing volume of biological and medical data at unprecedented speed and scale [1]. These accumulated data can be utilized for a more effective delivery of care and enhanced clinical decision-making [2]. However, analysis of these tremendous amounts of data—collected from electronic health records or monitoring devices—to extract useful information for a more broad-based health-care delivery is one of the main challenges of today's medicine [3].

Medical decision support systems help clinicians to best exploit these overwhelming amount of data by providing a computerized platform for integrating evidence-based knowledge and patient-specific information into an enhanced and cost-effective health care [4]. Over the last decade, various pattern recognition techniques have been applied to biomedical data (including signals and images) for automatic and machine-based clinical diagnostic and therapeutic support. The development of novel pattern recognition methods and algorithms with high performances, in terms of accuracy and/or time complexity, improves the health-care outcome by allowing clinicians to make a better-informed decision in a timelier manner. This is of vital importance especially when a rapid clinical decision needs to be made in a stressful environment such as intensive care units [5]. Development of predictive computational models and pattern recognition algorithms with performances and capabilities matching the complexity of rapidly evolving clinical measurement and monitoring systems is an ongoing research area and, thus, it requires continuous update on the current status of its advances [6].

With that scope in mind, this special issue focuses on the applications of pattern recognition for clinical decision support. For this purpose, we selected five research articles that discussed the design and clinical applications of feature extraction and/or classification of large-scale or high-dimensional biomedical data including biomedical signals and images.

S. Long et al. developed and evaluated an automatic retinal image processing algorithm to detect hard exudates (HE). Their algorithm consisted of four main stages: (i) image preprocessing; (ii) localization of optic disc; (iii) identification of potential HE region using dynamic threshold and fuzzy C-means clustering; and (iv) extraction of texture features from the HE region being fed into a support vector machine classifier. The proposed algorithm was trained and cross-validated on a publicly available e-ophtha EX database, achieving the overall average sensitivity of 76% and positive predictive value of 83%. Testing their algorithm on DIARETDB1 database resulted in better sensitivity (97%) and specificity (98%).

Z. Luo et al. developed and tested a machine learning approach to detect hypertension by automatic processing of radial artery blood pressure signals collected from healthy and hypertensive subjects by a noninvasive measurement device designed and built in house. They applied K-means clustering to exclude the noisy pulses from the data. Then they identified a subset of features that helped the most in detecting the hypertensive cases and applied various pattern classification algorithms such as AdaBoost and random forest. The authors were able to achieve the highest accuracy of 86% using AdaBoost.

K. Y. Win et al. proposed a novel methodology for the detection of cancer cells in cytological pleural effusion (CPE) images. Following some image intensity adjustment and application of median filtering to improve the image quality, cell nuclei were extracted by a hybrid segmentation method based on the fusion of simple linear iterative clustering and K-means clustering. Shape and contour concavity analyses were carried out to detect and split any overlapped nuclei into individual ones. Then several morphometric, colorimetric, and texture features were extracted. A novel hybrid feature selection method was developed using an artificial neural network to select the most discriminant and biologically interpretable features. Finally, an ensemble classifier of bagged decision trees was utilized to classify cells into being either benign or malignant. The proposed method achieved sensitivity of 88% and specificity of 99% on a dataset of 125 images containing more than 10,000 cells.

J. Jaworek-Korjakowska employed analysis of vascular structures in dermoscopy color images to distinguish benign and malignant pigmented skin lesions using a segmentation technique and convolutional neural network. The proposed method achieved sensitivity of 85% and specificity of 81%. The author concluded that small size and similarity of vascular structures to other local structures makes the segmentation and classification of dermoscopy color images challenging.

H. Tang et al. proposed employment of various features (from time/frequency domains) of phonocardiogram data and support vector machine for classifying normal and abnormal heart sound data. They applied correlation analysis to quantify discrimination level of the features and used support vector machine with radial basis kernel function for classification of phonocardiogram data from publicly available PhysioNet database. Their methodology achieved an average sensitivity of 88% and a specificity of 87%.

With upcoming technology upgrades in clinical measurement and monitoring systems, it is reasonable to assume that the cost of acquiring and storing biomedical data will decrease dramatically in the near future [1, 7]. This substantiates the need to further enhance the existing signal processing and pattern classification algorithms or develop novel ones to analyze the resulting complex data and help health-care providers in a variety of clinical decision-making processes [3]. The papers presented in this special issue outline some of the recently developed pattern recognition methodologies in clinical decision systems.

Conflicts of Interest

The authors declare no conflicts of interest.

Shadnaz Asgari
Fabien Scalzo
Magdalena Kasprowicz

References

- [1] J. Luo, M. Wu, D. Gopukumar, and Y. Zhao, "Big data application in biomedical research and health care: a literature review," *Biomedical Informatics Insights*, vol. 8, Article ID BII.S31559, 2016.
- [2] S. Rea, J. Pathak, G. Savova et al., "Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project," *Journal of Biomedical Informatics*, vol. 45, no. 4, pp. 763–771, 2012.
- [3] K. Najarian, K. R. Ward, and S. Shirani, "Biomedical signal and image processing for clinical decision support systems," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 974592, 2 pages, 2013.
- [4] L. Moja, K. H. Kwag, T. Lytras et al., "Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis," *American Journal of Public Health*, vol. 104, no. 12, pp. e12–e22, 2014.
- [5] G. K. Lighthall and C. Vazquez-Guillamet, "Understanding decision making in critical care," *Clinical Medicine & Research*, vol. 13, no. 3-4, pp. 156–168, 2015.
- [6] I. Contreras and J. Vehi, "Artificial intelligence for diabetes management and decision support: literature review," *Journal of Medical Internet Research*, vol. 20, no. 5, p. e10775, 2018.
- [7] H. Thimbleby, "Technology and the future of healthcare," *Journal of Public Health Research*, vol. 2, no. 3, p. 28, 2013.

Review Article

Decision Support Systems in Prostate Cancer Treatment: An Overview

Y. van Wijk ¹, I. Halilaj,¹ E. van Limbergen,² S. Walsh,¹
L. Lutgens,² P. Lambin,¹ and B. G. L. Vanneste ²

¹The D-Lab: Decision Support for Precision Medicine GROW - School for Oncology and Developmental Biology & MCCC, Maastricht University Medical Center, Maastricht, Netherlands

²Department of Radiation Oncology (MAASTRO), GROW - School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, Netherlands

Correspondence should be addressed to Y. van Wijk; y.vanwijk@maastrichtuniversity.nl

Received 26 October 2018; Revised 2 April 2019; Accepted 6 May 2019; Published 6 June 2019

Guest Editor: Shadnaz Asgari

Copyright © 2019 Y. van Wijk 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.

Background. A multifactorial decision support system (mDSS) is a tool designed to improve the clinical decision-making process, while using clinical inputs for an individual patient to generate case-specific advice. The study provides an overview of the literature to analyze current available mDSS focused on prostate cancer (PCa), in order to better understand the availability of decision support tools as well as where the current literature is lacking. **Methods.** We performed a MEDLINE literature search in July 2018. We divided the included studies into different sections: diagnostic, which aids in detection or staging of PCa; treatment, supporting the decision between treatment modalities; and patient, which focusses on informing the patient. We manually screened and excluded studies that did not contain an mDSS concerning prostate cancer and study proposals. **Results.** Our search resulted in twelve diagnostic mDSS; six treatment mDSS; two patient mDSS; and eight papers that could improve mDSS. **Conclusions.** Diagnosis mDSS is well represented in the literature as well as treatment mDSS considering external-beam radiotherapy; however, there is a lack of mDSS for other treatment modalities. The development of patient decision aids is a new field of research, and few successes have been made for PCa patients. These tools can improve personalized medicine but need to overcome a number of difficulties to be successful and require more research.

1. Introduction

Worldwide, prostate cancer (PCa) is the second most occurring type of cancer in men and the most commonly diagnosed cancer for men living in developed countries, making it a very relevant topic for cancer research [1].

A variety of treatment options is available to treat PCa, such as external-beam radiotherapy (EBRT) and radical prostatectomy [2–4], which have similar long-term survival outcomes. Other treatments such as brachytherapy [5–8] are gaining popularity, and active surveillance is an increasingly viable option as well [9, 10], due to the slow progression of some kinds of PCa. Retrospective studies comparing different treatment modalities tend to be conflicting and biased. Consensus on the best treatment choices for men with PCa remains absent because prospective trials for different

treatments report different toxicities [4, 11, 12]. Due to this, the treatment choice is largely dependent on both patient and physician subjective preferences, rather than knowledge-based decision-making [13]. Additionally, treatment outcome is dependent on a large number of features, including treatment, patient, tumor, clinical, and genetic features [14]. These factors further complicate the integration of evidence-based decision-making into clinical practice due to the limitations of human cognitive capacity, which can only take a relatively small number of factors into account on which to base a decision [15, 16]. In order to meet the aspiration of personalized medicine, the need for multifactorial decision support systems (mDSS) is growing [17–23]. An mDSS is a tool designed to improve the difficult medical decision-making process. It uses multifactorial inputs (treatment, patient, tumor, clinical, genetic, etc.) for a given patient to generate

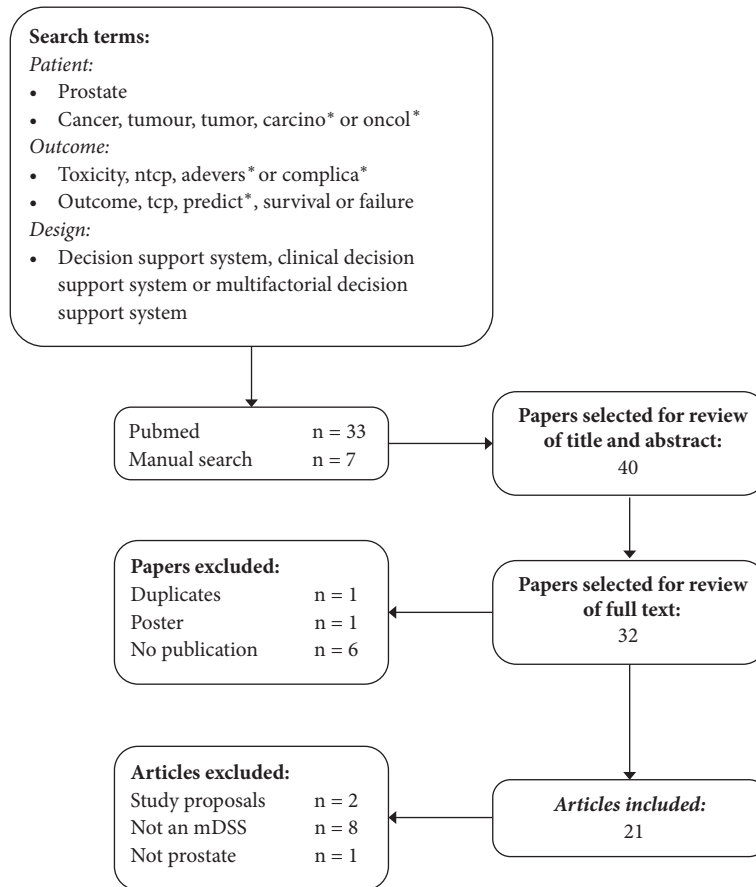


FIGURE 1: Flowchart of the results of the literature search in PubMed.

case-specific advice for patients, clinicians, or other medical professionals. Due to the variety of treatment options for PCa, all equally efficacious for outcome, but having different secondary effects, this disease is an interesting subject for the use of mDSS.

In addition to the need for mDSS for treatment selection, similar systems can be used for the diagnosis of PCa, improving early detection as well as reducing overdiagnosis and unnecessary testing. These mDSS can use imaging, clinical, biological, and other parameters to improve detection and risk classification of PCa in a minimally invasive method to maximize individual treatment.

This study provides an overview of the literature to analyze current available mDSS focused on PCa, in order to better understand the availability of mDSS as well as noting where the current literature is lacking. We aim to provide an update for clinicians about recent advances in mDSS for personalized PCa oncology, which may improve clinical decision-making.

2. Materials and Methods

2.1. Search Strategy. In order to identify all mDSS with relation to treatment for PCa, we performed a MEDLINE/PubMed literature search in July 2018, restricted to English. Details of the strategy we used are shown in Figure 1.

2.2. Selection Criteria. Prior to reviewing full texts, we manually checked the abstracts and titles to select papers for this study. Duplicates, posters, or abstracts that did not include a published work were excluded. Additionally, we excluded studies which clearly did not include an mDSS for PCa. After full text review, we excluded any papers that described study proposals that did not describe an mDSS in PCa. We selected appropriate studies by manually screening and considering the aforementioned criteria.

2.3. Study Characteristics. The studies we included were divided into sections according to the type of mDSS: diagnostic mDSS, which support the staging of PCa or support the decision for more invasive or expensive diagnostic tests; treatment mDSS, which support the decision between treatment modalities or treatment plans; and patient mDSS, which focus on informing the patient.

We described each study using the number of patients, the decision that the system supports, and the system inputs and outcomes. We also commented on the general applicability, the reported performance of the system, and the limitations. In order to assess the reporting quality of the studies, we tested each paper for its compliance to the TRIPOD (Transparent Reporting of studies on prediction models for Individual Prognosis Or Diagnosis) reporting guideline [24]. We reported this in a percentage calculated using the document:

TABLE 1: Overview of diagnosis support systems.

Study	N	Decision/Diagnose	Inputs	Outcomes	TRIPOD
(Roach et al. 1994)[25]	282	Low or high risk of LN involvement	PSA,GS, Clinical stage	P LN involvement	79%
(Diaz et al. 1994)[26]	217	Low or high risk of SV involvement	PSA, GS	P SV involvement	69%
(D'Amico et al. 1998)[27]	1872	Patient risk group	PSA, GS, Clinical stage	5-year PSA outcome	72%
(Chang et al. 1999)[28]	43	Localized vs Advanced PCa	PSA, GS, TRUS, DRE	P advanced PCa and P localized PCa	65%
(Roach et al. 2000)[29]	895	Extracapsular extension	PSA, GS	Extracapsular extension risk	92%**
(Lee et al. 2010)[30]	1077	Biopsy	Clinical*, Imaging*, PSA	P PCa	86%
(Kim et al. 2011)[31]	532	Advanced PCa	TRUS, Clinical, PSA	P advanced PCa	79%
(Shah et al. 2012)[32]	31	Location PCa	MRI image	Cancer probability map	83%
(Mukai et al. 2013)[33]	30773	PSA test	Clinical	Recommendation	-***
(Sadoughi et al. 2014)[34]	360	PCa	PSA, Age	P PCa	33%
(van Leeuwen et al. 2017)[35]	591	Significant PCa	Clinical, PSA, PIRADS, DRE	P significant PCa	97%
(Tosoian, et al. 2017)[36]	4459	Pathological Stage	PSA, GS, Clinical stage	% likelihood of given stage	83%

Abbreviations. N: number of patients; P: probability; PCa: prostate cancer; PSA: prostate specific antigen; PIRADS: Prostate Imaging Reporting and Data System; TRUS: transrectal ultrasound scan; LN: lymph node; GS: Gleason score; SV: seminal vesicles; TRIPOD: adherence to the TRIPOD statement; DRE: digital rectal examination.

*Clinical, imaging, and tumor parameters.

** Abstract only.

*** No development or validation of mDSS: no TRIPOD evaluation possible.

“Appendix I: Scoring adherence of prediction model study reports to the TRIPOD reporting guideline”, available on <http://www.triPod-statement.org/> (available here).

3. Results

3.1. Included Studies

3.1.1. mDSS for Diagnosis and Diagnostic Interventions. The studies that contained a diagnostic mDSS are listed in Table 1. Two of the studies we found had the goal of supporting the use of a diagnostic tool: Lee et al. (2010) and Mukai et al. (2013). The application tested by Mukai et al. (2013) was meant to support the use of prostate specific antigen (PSA) tests for patients. The mDSS was a web-based application that would aid in the decision to perform PSA tests for general practitioners (GPs) in Denmark. The study leads to the conclusion that it was possible to grant GPs in Denmark easy access to web-based mDSS by replacing certain words in their medical records by hyperlinks. However it also showed that this mDSS did not change PSA-testing behavior. Since this study neither developed nor validated this mDSS, the compliance to the TRIPOD guidelines could not be tested.

Lee et al. (2010) attempted to support the use of a biopsy by predicting initial biopsy outcomes through three different models: support vector machine (SVM), artificial

neural network (ANN), and multiple logistic regression. The study trained each of the models on 600 patients who had undergone transrectal ultrasonography (TRUS)-guided prostate biopsies, tested them on 477, and compared the model performances. The parameters of the models were TRUS findings and clinical parameters, including age and PSA. The area under the receiver operating characteristic (ROC) curve (AUC) for the use of multiple logistic regression analysis, ANN, and the SVM was 0.768, 0.778, and 0.847, respectively, and pairwise comparison of the ROC curves showed that the SVM model had superior performance.

Kim et al. (2011), Sadoughi et al. (2014), Shah et al. (2012), van Leeuwen et al. (2017), and Chang et al. (1999) all aimed to detect, diagnose, or classify PCa using a variety of methods. Kim et al. (2011) performed a study similar to Lee et al. (2010), but with the aim of improving pathological staging, rather than reducing the number of biopsies. Two models were developed, SVM and ANN, for the prediction of advanced PCa and compared based on performance. The models used TRUS-guided biopsy parameters and were tested on 532 patients divided into training and test groups. The SVM model performed significantly better ($p=0.02$) than the ANN model based on ROC curve, with an AUC of 0.805 while that of the ANN model was 0.719. This study showed that these models could improve objective pathological staging of biopsy-proven PCa patients and could be applied in

combination with TRUS-guided biopsies once externally validated.

Another neural network was trained on laboratory results by Sadoughi et al. (2014) who then performed particle swarm optimization. The specific goal of this study was to aid in distinguishing between localized PCa and benign hyperplasia of the prostate. The model was internally validated on 60 patients, and the authors found an accuracy of 98.33%. The description of the methodology was limited, and the reporting conformed only to 33% of the TRIPOD guidelines. The model could potentially improve detection of PCa and possibly reduce the number of biopsies, but external validation is necessary.

Notable were Shah et al. (2012), who also used SVM, but, in contrast to Lee et al. (2011), the model was not trained on biopsy results, but on pathological regions of a magnetic resonance imaging (MRI) scan of postprostatectomy prostates. The aim was not only to diagnose PCa, but to locate it specifically on MRI scans by modeling voxel specific risk analyses. The sensitivity and the specificity of the model with optimized SVM parameters were 90%, and the kappa coefficient was 80%, where the raters were the mDSS and the ground truth histology. The study only included 24 patients, but since the model was trained on specific regions, the training was done on 225 cancer and 264 noncancerous regions. This model could be applied in any hospital with a 3.0 T endorectal multiparametric magnetic resonance imaging (mpMRI) scanner, although it still requires validation.

Van Leeuwen et al. (2017) developed a nomogram, rather than a deep learning algorithm, that included a larger number of parameters to diagnose significant PCa. The nomogram included Prostate Imaging Reporting and Data System (PIRADS), age, PSA, digital rectal examination (DRE), prostate volume, and prior biopsy. The model performed with an AUC of 0.864 on an external validation set, and the paper proposed an optimal strategy to reduce the number of biopsies needed with minimal risk of underdiagnosis. This paper conformed to 97% of the TRIPOD guidelines.

Chang et al. (1999) evaluated the usefulness and the performance of an mDSS, the Prostate Cancer Expert System (PCES), which was validated on 43 patients with confirmed PCa. The PCES system, which utilized PSA, Gleason score, TRUS, and DRE, was used to categorize the patients into localized or advanced PCa, and the same test was applied to four attending physicians and four residents. The results showed that the PCES performed with a higher accuracy than all residents and physicians, though the difference was only higher for two physicians. It also showed that after consultation of the PCES, the staging accuracy of the residents improved to the level of the attending physicians.

A number of prediction tools are currently being applied in the clinic to aid in the further diagnosis of the PCa disease, providing predictions on lymph node (LN) involvement, organ confinement (OC), seminal vesicle (SV) involvement and extracapsular extension, and risk of failure after treatment. The Partin tables [37, 38] are a set of nomograms to predict OC of PCa, initially introduced in 1993 and most recently updated by Tosoian et al. (2017) [36]. Based on these nomograms, Roach derived a set of formulas for the

prediction of SV involvement in a paper published by Diaz et al. (1994) [26]. Roach et al. (1994) [25] also derived formulas that predicted LN involvement based on PSA and Gleason score, and Roach et al. (2000) [29] did the same for the risk of failure following radiotherapy (RT) and extracapsular extension in patients with localized PCa.

Tosoian et al. (2017) validated and updated the Partin tables on a cohort of 4459 patients with the goal of predicting the pathological outcome after radical prostatectomy. The performance of the model was tested for binary regression where the AUC was calculated when comparing organ confined (OC) PCa to other pathological outcomes. The model performed best when predicting OC versus LN involvement (AUC = 0.918) and versus seminal vesicle (SV) involvement (AUC = 0.856). The weakest performance was for OC versus extraprostatic extension (AUC = 0.673).

Diaz et al. (1994) split patients into high risk and low risk groups of SV involvement using PSA and Gleason score and tested this on 217 patients. The incidence rate of SV involvement in the low risk group was 7%, while the incidence rate in the high risk group was 37%, resulting in a chi-square of 23.17.

Roach et al. (1994) performed a similar study on 282 patients, but divided the patients into low and high risk groups for LN involvement. This resulted in 6% incidence rate in the low risk group and a 40% incidence rate in the high risk group, resulting in a significant split ($p < 0.001$).

Roach et al. (2000) split 895 patients into low, intermediate, and high risk groups for extracapsular extension. This resulted in an incident rate of 17.8%, 46.7%, and 66.7% in low, intermediate, and high risk groups, respectively, which was a significant split ($p < 0.01$).

D'Amico et al. (1998) suggested a widely accepted risk classification for prostate cancer to help predict biochemical outcome after five years after PCa treatment, stratifying them into low, intermediate, and high risk PCa [27]. This study included 1872 patients who underwent radical prostatectomy, EBRT, or interstitial RT. Cox regression was used to calculate the relative risk between different groups of patients, based on risk level and treatment type. No validation was performed in this paper.

Memorial Sloan Kettering Cancer Center (MSKCC) also has a publically available set of nomograms that is based on data from more than 10,000 patients. They have nomograms available to predict outcome after radical prostatectomy both before and after treatment as well as after PSA elevation (see <https://www.mskcc.org/nomograms>).

3.1.2. mDSS Supporting Treatment Decisions. The studies that contained treatment mDSS are listed in Table 2. The studies described by Walsh et al. (2018), Smith et al. (2016), and van Wijk et al. (2018) compare different RT treatment plans performed on the same patients and the mDSS selects the best plan with most favorable outcome. This type of personalized mDSS is very suitable for the comparison of RT modalities, as the treatment plans are a predictor for the delivered dose, with consequent treatment outcome. Walsh et al. (2018) used a combination of existing models for tumor control probability (TCP) and normal tissue complication probability

TABLE 2: Overview of treatment support systems.

Study	N	Decision	Inputs	Outcomes	TRIPOD
(Hodges et al. 2012)[43]	Model	SBRT, IMRT	Utility, transitions	QALY, Costs, ICER	80%
(Reed et al. 2014)[44]	Model	ART	Risk group ProsVue slope	QALY, Cost, ICER	80%
(Vanneste et al. 2015)[45]	Model	IMRT+IRS, IMRT	Utility, transitions	QALY, Cost, ICER	81%
(Smith et al. 2016)[46]	Model	RT plan	DVH, Clinical*	TCP, NTCP, QALY	87%
(van Wijk et al. 2018)[47]	23	IRS in RT	DVH, Clinical	TCP, NTCP	84%
(Salem et al. 2018)[48]	200	Follow-up	Symptoms, Blood tests	Follow-up suggestion	71%
(Walsh et al. 2018)[49]	25	IMRT, V-mat, PSPT, IMPT	DVH	TCP, NTCP, Robustness, stability	84%

Abbreviations. N: number of patients; IMRT: intensity modulated radiotherapy; V-mat: volumetric-modulated arc therapy; PSPT: passively scattered proton therapy; IMPT: image modulated proton therapy; TCP: tumor control probability; NTCP: normal tissue complication probability; DVH: dose-volume histogram; QALY: quality adjusted life year; IRS: implantable rectum spacer; ART: adjuvant radiotherapy; ICER: incremental cost-effectiveness ratio; TRIPOD: adherence to the TRIPOD statement.

*Clinical parameters.

(NTCP) to compare different treatment plans with photon and proton RT for 25 patients. The study included extensive corrections for displacements during treatment, which aided in the prediction of delivered dose. As this was a modeling study, no validation was done. However, the concept could be used as a basis for RT plan selection between different modalities and could aid in the optimization of TCP and NTCP. Smith et al. (2016) utilized an advanced Bayesian network to optimize intensity modulated radiotherapy (IMRT) treatment plans based on outcome in terms of progression free survival and toxicity. The models were validated against independent clinical trials for the metastasis free survival and overall survival and resulted in uncertainties of 2.5% and 2%, respectively. This method could potentially be implemented into any IMRT planning system and has the potential to improve the quality of treatment plans, resulting in optimized outcomes.

A device called the implantable rectum spacer (IRS) has been developed to spare the rectum during IMRT by increasing the distance between the anterior rectum wall and the prostate [39, 40]. Van Wijk et al. (2018) made use of image deformation based on a *virtual* IRS [41, 42]: she published models to predict the sparing effect of an IRS before implanting the IRS. The model was tested in a proof of concept study with 16 patients, comparing the mDSS outcome for the *virtual* IRS to the *real* IRS in the patients, and the median discrepancies in outcome were 1.8%. Once validation has taken place, this DSS could be applied to any RT planning system and has the potential to personalize treatment choice.

Two studies that supported the decisions involving follow-up treatment were found. Reed et al. (2014) analyzed the cost-effectiveness (CE) of the use of a nucleic acid detection immunoassay (NADiA) ProsVue™ to support the decision for adjuvant radiotherapy (ART). This model showed that primarily for the intermediate risk patients, NADiA ProsVue had an incremental cost-effectiveness ratio (ICER) lower than \$50,000 in 83.6% of the simulations. This analysis in its current state is not applicable outside the USA. Salem et al. (2018) validated a computer-led decision

support system aimed at giving additional advice in follow-up treatment strategy. This system uses the clinical profile of the patient in combination with blood test results to propose a follow-up strategy. Validation was done on 200 patients by comparing the computer generated strategy to the advice given by trained urologists. In all cases, the agreement between the experts and the system was better than moderate ($\kappa > 0.6$). The paper argued that such a system could significantly reduce costs in follow-up care in the United Kingdom and believed the system can be used by any healthcare worker, regardless of urology background.

Hodges et al. (2012) developed a CE model to analyze the benefit of IMRT with respect to stereotactic body radiotherapy (SBRT). Though this study compared these two treatments, this study carried out analysis on cohort level and not case-specific, as is expected of personalized medicine. The CE analysis was done solely based on probabilistic simulations, thus not taking into account variations in cost, transition rates, or utility values. A sensitivity analysis was performed, revealing that in 66% of the iterations SBRT was cost-effective over IMRT.

Vanneste et al. (2015) constructed a CE model to compare IMRT + IRS with IMRT without IRS. He developed a decision-analytic Markov model to examine the effect of late rectal toxicity and compare the costs and quality-adjusted life years (QALYs). An ICER of €55,880 per QALY was gained. For a ceiling ratio of €80,000, IMRT+S had a 77% probability of being cost-effective.

3.1.3. mDSS Focusing on Providing Patient Information. This section describes the studies that attempted to improve shared decision-making. This type of mDSS fosters patient involvement in therapeutic decisions and emphasizes the provision of information needed to make such decisions (Table 3). Berry et al. (2013) described the testing of a personal patient-profile (P3P) intervention for PCa patients. P3P is a web-based program to help men prepare for shared decision-making about the management of early stage prostate cancer.

TABLE 3: Summary of patient support systems.

Study	N	Decision	Inputs	Outcomes	TRIPOD
(Nguyen et al. 2009)[50]	Literature	Treatment	Various	Various	86%
(Berry et al. 2013)[51]	494	Treatment	P3P	Decision satisfaction	82%

Abbreviations. N: number of patients; P3P: personal patient-profile for prostate cancer.

TABLE 4: Overview of excluded studies that described tools to improve mDSS.

Study	N	Tool	Inputs	Outcomes
(Daemen et al. 2009)[61]	55	Genetic integration	DNA, CNV	Cancer outcome
(Kuru et al. 2013)[59]	50	Diagnostics	mpMRI	PIRADS
(Zumsteg et al. 2013)[55]	1024	Risk stratification	Risk factors, Gleason score, biopsy	Risk group
(Beyan et al. 2014) [62]	Model	Genetic integration	SNPs	Various
(Kent and Vickers 2015)[60]	Model	Diagnostics	Clinical and tumor features	Life expectancy
(Gnanapragasam et al. 2016)[56]	10139	Risk stratification	PSA, stage, Gleason score	Risk group
(Epstein et al. 2016)[57]	26325	Risk stratification	PSA, stage	Gleason grade
(Gries et al. 2017)[58]	120	Utility values	Utility's 18 health states	Utility's 243 health states

Abbreviations. N: Number of patients; CNV: copy number variation; SNP: single nucleotide polymorphism; PSA: prostate specific antigen; mpMRI: multiparametric magnetic resonance imaging; PIRADS: Prostate Imaging Reporting and Data system.

The study hypothesized that when P3P was used in combination with a consultation with different clinicians to prepare for decision-making, patients were more satisfied with their treatment choice and experienced less regret, but testing did not result in significant improvement. However, this type of system has the capability to take into account patient preferences and priorities, while when only using consultations, these preferences can be misinterpreted by clinicians. The paper suggested similar studies to be performed in the future on larger cohorts.

Nguyen et al. (2009) performed an extensive literature research for predictive outcome models for 15 different treatment options for PCa, including active surveillance, combinations of radical prostatectomy, RT, hormonal therapy, and high intensity focused ultrasound. They attempted to create a comprehensive overview of the different outcome combinations, such as survival, metastasis, and various toxicities. This overview was designed to be comprehensive for patients so that they could use their own priorities and preferences to make an informed decision. Though they concluded that there is a great need for additional models, this paper provided a framework and is a step towards evidence-based personalized medicine. Additionally, this framework could be a useful tool for clinical decision-making by medical personnel when adapted for these users.

3.2. Excluded Studies. Two studies, Stacey et al. (2016) [52] and Jayadevappa et al. (2015) [53], were excluded because they contained study proposals. Both these studies will attempt to test patient decision aids. Another study, McRae et al. (2016) [54], was excluded because it was not about prostate cancer.

Eight studies in total were excluded as they did not describe an mDSS, but instead provided tools that could be used in the development of an mDSS. Though these studies can be used to further personalize PCa treatment, these tools cannot be applied in the clinic in their current form. A short overview of these papers is given in Table 4.

Two studies, Daemen et al. (2009) and Beyan et al. (2014), attempted to use genetic information to improve cancer prediction outcomes. Beyan specifically worked to incorporate single nucleotide polymorphisms (SNPs) into the national health information system of Turkey.

Zumsteg et al. (2013) [55], Gnanapragasam et al. (2016) [56], and Epstein et al. (2016) [57] provided new definitions for risk group stratification, increasing upon the current low, intermediate, and high risk groups. They found clear separation in progression free survival in their new risk groups, making these a potential tool for diagnostic mDSS. Gries et al. (2017) [58] provided a method to interpolate the utility values for all combinations of 18 different health states, which could be a valuable tool for mDSS to quantify QALYs. Kuru et al. (2013) [59] showed that a PIRADS score on mpMRI is prognostic for PCa.

Finally Kent et al. (2015) [60] performed a literature search and attempted to create a diagnostic tool which would predict the life expectancy in PCa patients. They concluded that no existing model was suitable for incorporation into an mDSS.

4. Discussion

4.1. Primary Findings. In response to the increasing number of PCa treatment options, the development of mDSS has

become a growing topic of interest to provide an aid in difficult medical decisions. Diagnosis and staging mDSS have also been a growing topic of interest, as well as a number of tools to improve cancer detection, predictions of treatment outcome, and outcome stratification. However, there is still a need for new mDSS in treatment decision aids and validation of diagnostic mDSS. Additionally, the field of patient informed decision-making is still in its infancy, but essential for the growth towards individualized medicine.

For the diagnostic tools, there are a number of viable tools available for the diagnosis of local PCa that have been validated on large cohorts [30, 31, 33, 35, 63]. Additionally, a study has been performed for detection of PCa on the voxel level for MRI images [32]. This type of diagnostic mDSS could additionally assist in treatment planning or treatment selection. One study explores diagnosis of PCa with LN metastasis [64], but this field remains largely uncharted, similar to the computer automated detection of biochemical failure or treatment failure, also explored in only one study [48].

For treatment mDSS, the proportion between the number of treatment options and the decision support tools remains somewhat skewed. Nguyen et al. (2009)[50] suggested 15 different combinations of treatment options, and this did not include the use of rectal displacement devices [65, 66] or proton therapy. The treatment mDSS that were found by the current study primarily involved RT, including treatment plan selection [46], proton compared to photon therapy [49], SBRT compared to IMRT [43], and the use of an implantable rectum spacer in EBRT [47]. Nguyen et al. (2009) attempted to create a general overview of prediction tools to create a clear overview for patients, and they found most prediction tools to be focused on RT and radical prostatectomy. Less prediction tools are available for brachytherapy, which is a very viable treatment option for PCa. One study, Alitto et al. (2017) [67], describes an Umbrella Protocol for the standardized development of new mDSS. This protocol could help in improving the application of new mDSS in clinics.

Development of patient decision aids is challenging, as cultural and language barriers are much more present in this field of research. Nguyen et al. (2009) developed a comprehensive treatment overview for patients, but found that the predictive tools available were limited, leaving an incomplete overview. Berry et al. (2013) [51] hypothesized that patients satisfaction was increased after treatment when they were actively involved in the decision-making process and comorbidities like anxiety, depression, and fatigue were reduced, and they proceeded to test this using the P3P intervention. They have found, however, that this method has not increased self-reported preparation for the intervention.

4.2. mDSS Design. The last decade deep learning algorithms have gained popularity in the development of mDSS for the classification of cancers, and the same is true within the field of PCa. Kim et al. (2011) and Lee et al. (2010) tackled a similar problem with both an ANN and an SVM, but the dataset of Kim et al. (2011) was approximately double

in size. This increased patient cohort resulted in a higher AUC performance for both models, which confirms that deep learning algorithms rely heavily on large datasets. Also, no external validation was done, so currently these models are not generally applicable. An approach to make these models both more accurate and more applicable for clinical use would be to use distributed learning, where the models are trained on centers all over the world, without the data having to leave the clinics. Shah et al. (2012) used an SVM for a more complex problem than just classification by doing voxel based analyses for the localization of PCa. The drawback of this model is that for training, a large amount of imaging data must be available, as well as 3D pathology, but the advantage is that the usage of subregions in the prostate allows for data augmentation.

Notable is that a number of nomograms initially developed more than two decades ago are currently still in use, though updated using newer datasets. The continued application of simple, easy to interpret models is something to keep in mind in the development of new mDSS. Though artificial intelligence has the potential to improve diagnosis of PCa, transparency plays a large role in clinical application. It also shows that predictive parameters for PCa are very consistent, with persistent usage of Gleason score, PSA, clinical stage, and age. Any new mDSS being developed should be tested against the performance of these parameters to avoid tackling a simple problem using computationally heavy machine learning.

When looking at the available treatment mDSS, we see that most of these are focused on RT. This is likely due to the patient specific treatment planning done in RT, which results in highly detailed dose maps before the start of treatment. This is ideal for mDSS, as different treatment plans can be directly compared, and the different outcomes can be predicted using dose response curves. Other types of treatment, such as prostatectomy or watchful waiting, rely much more heavily on clinical parameters and tumor parameters for predictions of their outcome or on subjective physician decisions. This makes it harder to compare outcomes of different treatments for the same patient. The development of new mDSS comparing completely different treatment types with each other for the same patient would be very beneficial for filling this gap in the current literature.

4.3. Study Limitations. This study focused on mDSS in PCa, and the scope did not include any other cancer types. It is possible that general mDSS, applicable for more than one cancer type, were therefore overlooked. The search was also focused on mDSS and thus did not include any predictive models that could aid in decision-making. The terms describing mDSS may not have been used in interesting studies that could have been included in this overview. 'Patient decision aids' was also not a search term, which may be the cause of the limited number of mDSS for patients included in this overview. The search performed was only MEDLINE/PubMed linked, so studies not available on these media were overlooked.

Inherent to literature overviews is that negative findings are not always reported, so failed attempts at creating mDSS

are often uncommon in overviews and literature reviews, and this may cause a biased view on the subject.

We assessed the reporting quality of each included study using the TRIPOD statement for quantification and comparison. Although the TRIPOD statement has been endorsed by a large number of medical journals and editorial organizations, it is not a universal gold standard. The checklist used for the assessment has limitations, such as the severe point punishment for the lack of specific keywords in title or abstract. This could reduce the score of a well-written paper. Additionally, the adherence to the TRIPOD statement was compressed into a single number, and the specific reporting issues were not named, such as improper analyses, lack of validation, or reporting of study specifics. However, it is a useful tool to show the strength of a report, as when adherence to the TRIPOD statement was high, the paper was clearly written, and all proper steps were taken for the development of quality mDSS tools.

5. Conclusion

A number of mDSS for the primary diagnosis and staging of localized PCa are available. Treatment mDSS were mostly focused on EBRT, for which several tools are available. However, a lack of mDSS for other treatment modalities suggests that the development of new tools is necessary to objectively compare different treatment modalities. The development of patient decision aids is a new field of research, and few successes have been made for PCa patients. Though the idea of informed decision-making by patients is in line with the goal of personalized medicine, the development of these tools needs to overcome a number of barriers to be successful, like comprehensiveness, language barriers, patient cooperation, and physician cooperation. More research needs to be performed to better empower clinical decisions in the diagnosis and treatment process.

Conflicts of Interest

Dr. Philippe Lambin reports grants/sponsored research agreements from Varian Medical, Oncoradiomics, ptTheragnostic, and, outside the submitted work, Health Innovation Ventures and DualTpharma. He received an advisor (SAB)/presenter fee and/or reimbursement of travel costs/external grant writing fee and/or in-kind manpower contribution from Oncoradiomics and, outside the submitted work, BHV and Convert Pharmaceuticals. Dr. Lambin has shares in the company Oncoradiomics SA and, outside the submitted work, Convert Pharmaceuticals SA and is co-inventor of two patents on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Oncoradiomics; one patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito; three non-patentable inventions (software) licensed to ptTheragnostic/DNAmito, Oncoradiomics, and Health Innovation Ventures. Dr. Seán Walsh is a paid employee at Oncoradiomics. Yvonka van Wijk is an employee at ptTheragnostic. Iva Halilaj is a paid employee at Health Innovations.

Supplementary Materials

The supplementary material we provided contains a checklist of TRIPOD points and the calculation method we used to calculate the adherence to the TRIPOD statement. The version put in the supplementary materials is the one available on the TRIPOD site. It first contains a description, then part A with general description of the study, and then part B where each point is checked. (*Supplementary Materials*)

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] F. C. Hamdy, J. L. Donovan, J. A. Lane et al., "10-Year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer," *The New England Journal of Medicine*, vol. 375, no. 15, pp. 1415–1424, 2016.
- [3] B. Lennernäs, K. Majumder, J.-E. Damber et al., "Radical prostatectomy versus high-dose irradiation in localized/locally advanced prostate cancer: a swedish multicenter randomized trial with patient-reported outcomes," *Acta Oncologica*, vol. 54, no. 6, pp. 875–881, 2015.
- [4] J. L. Donovan, F. C. Hamdy, J. A. Lane et al., "Patient-reported outcomes after monitoring, surgery, or radiotherapy for prostate cancer," *The New England Journal of Medicine*, vol. 375, no. 15, pp. 1425–1437, 2016.
- [5] N. N. Stone, L. Potters, B. J. Davis et al., "Multicenter analysis of effect of high biologic effective dose on biochemical failure and survival outcomes in patients with gleason score 7-10 prostate cancer treated with permanent prostate brachytherapy," *International Journal of Radiation Oncology • Biology • Physics*, vol. 73, no. 2, pp. 341–346, 2009.
- [6] J. E. Sylvester, P. D. Grimm, J. C. Blasko et al., "15-Year biochemical relapse free survival in clinical Stage T1-T3 prostate cancer following combined external beam radiotherapy and brachytherapy; Seattle experience," *International Journal of Radiation Oncology • Biology • Physics*, vol. 67, no. 1, pp. 57–64, 2007.
- [7] H. Ragde, G. L. Grado, and B. S. Nadir, "Brachytherapy for clinically localized prostate cancer: Thirteen-year disease-free survival of 769 consecutive prostate cancer patients treated with permanent implants alone," *Archivos Españoles de Urología*, vol. 54, no. 7, pp. 739–747, 2001.
- [8] P. Grimm, I. D. Billiet, Bostwick. et al., "Comparative analysis of prostate-specific antigen free survival outcomes for patients with low, intermediate and high risk prostate cancer treatment by radical therapy. Results from the Prostate Cancer Results Study Group," *BJU Int*, vol. 109, supplement 1, pp. 22–29, 2012.
- [9] H. B. Musunuru, T. Yamamoto, L. Klotz et al., "Active surveillance for intermediate risk prostate cancer: survival outcomes in the sunnybrook experience," *The Journal of Urology*, vol. 196, no. 6, pp. 1651–1658, 2016.
- [10] G. Agarwal, D. Buethe, C. Russell, A. Luchey, and J. M. Pow-Sang, "Long term survival and predictors of disease reclassification in patients on an active surveillance protocol for prostate cancer," *Canadian Journal of Urology*, vol. 23, no. 2, pp. 8215–8219, 2016.

- [11] M. Soloway and M. Roach III, "Prostate cancer progression after therapy of primary curative intent: A review of data from the prostate-specific antigen era," *Cancer*, vol. 104, no. 11, pp. 2310–2322, 2005.
- [12] M. Roach III, "Radical prostatectomy or external beam radiotherapy: One step forward or two steps back?" *Cancer*, vol. 95, no. 2, pp. 215–218, 2002.
- [13] B. G. L. Vanneste, E. J. van Limbergen, E. N. van Lin et al., "Prostate cancer radiation therapy: what do clinicians have to know?" *BioMed Research International*, vol. 2016, Article ID 6829875, 14 pages, 2016.
- [14] V. Valentini, N. Dinapoli, and A. Damiani, "The future of predictive models in radiation oncology: from extensive data mining to reliable modeling of the results," *Future Oncology*, vol. 9, no. 3, pp. 311–313, 2013.
- [15] A. Dekker, S. L. Vinod, Holloway. et al., "Rapid learning in practice: a lung cancer survival decision support system in routine patient care data," *Radiotherapy and Oncology*, vol. 113, no. 1, pp. 47–53, 2014.
- [16] A. P. Abernethy, L. M. Etheredge, P. A. Ganz et al., "Rapid-learning system for cancer care," *Journal of Clinical Oncology*, vol. 28, no. 27, pp. 4268–4274, 2010.
- [17] P. Lambin, R. G. P. M. Van Stiphout, M. H. W. Starmans et al., "Predicting outcomes in radiation oncology-multifactorial decision support systems," *Nature Reviews Clinical Oncology*, vol. 10, no. 1, pp. 27–40, 2013.
- [18] P. Lambin, J. Zindler, B. Vanneste et al., "Modern clinical research: how rapid learning health care and cohort multiple randomised clinical trials complement traditional evidence based medicine," *Acta Oncologica*, vol. 54, no. 9, pp. 1289–1300, 2015.
- [19] P. Lambin, E. Roelofs, B. Reymen et al., "'Rapid Learning health care in oncology' - an approach towards decision support systems enabling customised radiotherapy," *Radiotherapy & Oncology*, vol. 109, no. 1, pp. 159–164, 2013.
- [20] P. Lambin, S. F. Petit, H. J. W. L. Aerts et al., "The ESTRO Breur lecture 2009. from population to voxel-based radiotherapy: exploiting intra-tumour and intra-organ heterogeneity for advanced treatment of non-small cell lung cancer," *Radiotherapy & Oncology*, vol. 96, no. 2, pp. 145–152, 2010.
- [21] P. Lambin, J. B. G. Zindler, Vanneste. et al., "Decision support systems for personalized and participative radiation oncology," *Advanced Drug Delivery Reviews*, vol. 109, pp. 131–153, 2017.
- [22] P. A. Pawloski, G. A. Brooks, M. E. Nielsen, and B. A. Olson-Bullis, "A systematic review of clinical decision support systems for clinical oncology practice," *Journal of the National Comprehensive Cancer Network*, vol. 17, no. 4, pp. 331–338, 2019.
- [23] T. R. McNutt, S. H. Benedict, D. A. Low et al., "Using big data analytics to advance precision radiation oncology," *International Journal of Radiation Oncology • Biology • Physics*, vol. 101, no. 2, pp. 285–291, 2018.
- [24] K. G. M. Moons, D. G. Altman, J. B. Reitsma, and G. S. Collins, "New guideline for the reporting of studies developing, validating, or updating a multivariable clinical prediction model: The TRIPOD statement," *Advances in Anatomic Pathology*, vol. 22, no. 5, pp. 303–305, 2015.
- [25] M. Roach III, C. Marquez, H.-S. Yuo et al., "Predicting the risk of lymph node involvement using the pre-treatment prostate specific antigen and gleason score in men with clinically localized prostate cancer," *International Journal of Radiation Oncology • Biology • Physics*, vol. 28, no. 1, pp. 33–37, 1994.
- [26] A. Diaz, M. Roach III, C. Marquez et al., "Indications for and the significance of seminal vesicle irradiation during 3D conformal radiotherapy for localized prostate cancer," *International Journal of Radiation Oncology • Biology • Physics*, vol. 30, no. 2, pp. 323–329, 1994.
- [27] A. V. D'Amico, R. Whittington, S. Bruce Malkowicz et al., "Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer," *Journal of the American Medical Association*, vol. 280, no. 11, pp. 969–974, 1998.
- [28] P. L. Chang, Y.-C. Li, T. M. Wang, S. T. Huang, M. L. Hsieh, and K. H. Tsui, "Evaluation of a decision-support system for preoperative staging of prostate cancer," *Medical Decision Making*, vol. 19, no. 4, pp. 419–427, 1999.
- [29] M. Roach 3rd, A. Chen, J. Song et al., "Pretreatment prostate-specific antigen and Gleason score predict the risk of extracapsular extension and the risk of failure following radiotherapy in patients with clinically localized prostate cancer," *Seminars in Urologic Oncology*, vol. 18, no. 2, pp. 108–114, 2000.
- [30] H. J. Lee, S. I. Hwang, S.-M. Han et al., "Image-based clinical decision support for transrectal ultrasound in the diagnosis of prostate cancer: Comparison of multiple logistic regression, artificial neural network, and support vector machine," *European Radiology*, vol. 20, no. 6, pp. 1476–1484, 2010.
- [31] S. Y. Kim, S. K. Moon, D. C. Jung et al., "Pre-operative prediction of advanced prostatic cancer using clinical decision support systems: accuracy comparison between support vector machine and artificial neural network," *Korean Journal of Radiology*, vol. 12, no. 5, pp. 588–594, 2011.
- [32] V. Shah, B. Turkbey, H. Mani et al., "Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging," *Medical Physics*, vol. 39, no. 7, pp. 4093–4103, 2012.
- [33] T. O. Mukai, F. Bro, F. Olesen, and P. Vedsted, "To test or not: A registry-based observational study of an online decision support for prostate-specific antigen tests," *International Journal of Medical Informatics*, vol. 82, no. 10, pp. 973–979, 2013.
- [34] F. Sadoughi and M. Ghaderzadeh, "A hybrid particle swarm and neural network approach for detection of prostate cancer from benign hyperplasia of prostate," *Studies in Health Technology and Informatics*, vol. 205, pp. 481–485, 2014.
- [35] P. J. van Leeuwen, A. Hayen, J. E. Thompson et al., "A multiparametric magnetic resonance imaging-based risk model to determine the risk of significant prostate cancer prior to biopsy," *BJU International*, vol. 120, no. 6, pp. 774–781, 2017.
- [36] J. J. Tosoian, M. Chappidi, Z. Feng et al., "Prediction of pathological stage based on clinical stage, serum prostate-specific antigen, and biopsy Gleason score: Partin Tables in the contemporary era," *BJU International*, vol. 119, no. 5, pp. 676–683, 2017.
- [37] A. W. Partin, J. Yoo, H. B. Carter et al., "The use of prostate specific antigen, clinical stage and Gleason score to predict pathological stage in men with localized prostate cancer," *The Journal of Urology*, vol. 150, no. 1, pp. 110–114, 1993.
- [38] A. W. Partin, L. A. Mangold, D. M. Lamm, P. C. Walsh, J. I. Epstein, and J. D. Pearson, "Contemporary update of prostate cancer staging nomograms (Partin Tables) for the new millennium," *Urology*, vol. 58, no. 6, pp. 843–848, 2001.
- [39] B. G. L. Vanneste, K. van De Beek, L. Lutgens, and P. Lambin, "Implantation of a biodegradable rectum balloon implant: tips, tricks and pitfalls," *International Brazilian Journal of Urology*, vol. 43, no. 6, pp. 1033–1042, 2017.

- [40] B. G. Vanneste, F. Buettner, M. Pinkawa, P. Lambin, and A. L. Hoffmann, "Ano-rectal wall dose-surface maps localize the dosimetric benefit of hydrogel rectum spacers in prostate cancer radiotherapy," *Clinical and Translational Radiation Oncology*, vol. 14, pp. 17–24, 2019.
- [41] Y. van Wijk, B. G. L. Vanneste, S. Walsh et al., "Development of a virtual spacer to support the decision for the placement of an implantable rectum spacer for prostate cancer radiotherapy: Comparison of dose, toxicity and cost-effectiveness," *Radiotherapy & Oncology*, vol. 125, no. 1, pp. 107–112, 2017.
- [42] B. G. L. Vanneste, Y. van Wijk, L. C. Lutgens et al., "Dynamics of rectal balloon implant shrinkage in prostate VMAT: Influence on anorectal dose and late rectal complication risk," *Strahlentherapie und Onkologie*, vol. 194, no. 1, pp. 31–40, 2018.
- [43] J. C. Hodges, Y. Lotan, T. P. Boike, R. Benton, A. Barrier, and R. D. Timmerman, "Cost-effectiveness analysis of SBRT versus IMRT: an emerging initial radiation treatment option for organ-confined prostate cancer," *American Journal of Managed Care*, vol. 18, no. 5, pp. e186–e193, 2012.
- [44] S. D. Reed, S. B. Stewart, C. D. Scales Jr., and J. W. Moul, "A framework to evaluate the cost-effectiveness of the NADiA prosvue slope to guide adjuvant radiotherapy among men with high-risk characteristics following prostatectomy for prostate cancer," *Value in Health*, vol. 17, no. 5, pp. 545–554, 2014.
- [45] B. G. L. Vanneste, M. Pijls-Johannesma, L. Van De Voorde et al., "Spacers in radiotherapy treatment of prostate cancer: is reduction of toxicity cost-effective?" *Radiotherapy & Oncology*, vol. 114, no. 2, pp. 276–281, 2015.
- [46] W. P. Smith, M. Kim, C. Holdsworth, J. Liao, and M. H. Phillips, "Personalized treatment planning with a model of radiation therapy outcomes for use in multiobjective optimization of IMRT plans for prostate cancer," *Journal of Radiation Oncology*, vol. 11, no. 1, article no. 38, 2016.
- [47] Y. van Wijk, B. G. Vanneste, A. Jochems et al., "Development of an isotoxic decision support system integrating genetic markers of toxicity for the implantation of a rectum spacer," *Acta Oncologica*, vol. 57, no. 11, pp. 1499–1505, 2018.
- [48] H. A. Salem, G. Caddeo, J. McFarlane et al., "A multicentre integration of a computer-led follow-up of prostate cancer is valid and safe," *BJU International*, vol. 122, no. 3, pp. 418–426, 2018.
- [49] S. Walsh, E. Roelofs, P. Kuess et al., "Towards a clinical decision support system for external beam radiation oncology prostate cancer patients: Proton vs. photon radiotherapy? a radiobiological study of robustness and stability," *Cancers*, vol. 10, no. 2, 2018.
- [50] C. T. Nguyen and M. W. Kattan, "Development of a prostate cancer metagram: A solution to the dilemma of which prediction tool to use in patient counseling," *Cancer*, vol. 115, no. 13, pp. 3039–3045, 2009.
- [51] D. L. Berry, Q. Wang, B. Halpenny, and F. Hong, "Decision preparation, satisfaction and regret in a multi-center sample of men with newly diagnosed localized prostate cancer," *Patient Education and Counseling*, vol. 88, no. 2, pp. 262–267, 2012.
- [52] D. Stacey, M. Taljaard, J. Smylie et al., "Implementation of a patient decision aid for men with localized prostate cancer: evaluation of patient outcomes and practice variation," *Implementation Science*, vol. 11, no. 1, p. 87, 2015.
- [53] R. Jayadevappa, S. Chhatre, J. J. Gallo et al., "Treatment preference and patient centered prostate cancer care: design and rationale," *Contemporary Clinical Trials*, vol. 45, pp. 296–301, 2015.
- [54] M. P. McRae, G. Simmons, J. Wong, and J. T. McDevitt, "Programmable bio-nanochip platform: a point-of-care biosensor system with the capacity to learn," *Accounts of Chemical Research*, vol. 49, no. 7, pp. 1359–1368, 2016.
- [55] Z. S. Zumsteg, D. E. Spratt, I. Pei et al., "A new risk classification system for therapeutic decision making with intermediate-risk prostate cancer patients undergoing dose-escalated external-beam radiation therapy," *European Urology*, vol. 64, no. 6, pp. 895–902, 2013.
- [56] V. J. Gnanapragasam, A. Lophatananon, K. A. Wright et al., "Improving clinical risk stratification at diagnosis in primary prostate cancer: a prognostic modelling study," *PLoS Medicine*, vol. 13, no. 8, p. e1002063, 2016.
- [57] J. I. Epstein, M. J. Zelefsky, D. D. Sjoberg et al., "A contemporary prostate cancer grading system: a validated alternative to the gleason score," *European Urology*, vol. 69, pp. 428–435, 2016.
- [58] K. S. Gries, D. A. Regier, S. D. Ramsey, and D. L. Patrick, "Utility estimates of disease-specific health states in prostate cancer from three different perspectives," *Applied Health Economics and Health Policy*, vol. 15, no. 3, pp. 375–384, 2017.
- [59] T. H. Kuru, M. C. Roethke, P. Rieker et al., "Histology core-specific evaluation of the European Society of Urogenital Radiology (ESUR) standardised scoring system of multiparametric magnetic resonance imaging (mpMRI) of the prostate," *BJU International*, vol. 112, no. 8, pp. 1080–1087, 2013.
- [60] M. Kent and A. J. Vickers, "A systematic literature review of life expectancy prediction tools for patients with localized prostate cancer," *The Journal of Urology*, vol. 193, no. 6, pp. 1938–1942, 2015.
- [61] A. Daemen, O. Gevaert, F. Ojeda et al., "A kernel-based integration of genome-wide data for clinical decision support," *Genome Medicine*, vol. 1, no. 4, p. 39, 2009.
- [62] T. Beyan and Y. Aydın Son, "Incorporation of personal single nucleotide polymorphism (SNP) data into a national level electronic health record for disease risk assessment, part 1: an overview of requirements," *JMIR Medical Informatics*, vol. 2, no. 2, p. e15, 2014.
- [63] M. Ghaderzadeh, "Clinical decision support system for early detection of prostate cancer from benign hyperplasia of prostate," *Studies in Health Technology and Informatics*, vol. 192, p. 928, 2013.
- [64] P. L. Chang, T. M. Wang, S. T. Huang, M. L. Hsieh, K. H. Tsui, and R. H. Lai, "Use of a medical decision support system to improve the preoperative diagnosis of prostate cancer with pelvic lymph node metastases," *Changcheng Yi Xue Za Zhi*, vol. 22, no. 4, pp. 556–564, 1999.
- [65] L. Wilton, M. Richardson, S. Keats et al., "Rectal protection in prostate stereotactic radiotherapy: a retrospective exploratory analysis of two rectal displacement devices," *Journal of Medical Radiation Sciences*, vol. 64, no. 4, pp. 266–273, 2017.
- [66] G. Mok, E. Benz, J.-P. Vallee, R. Miralbell, and T. Zilli, "Optimization of radiation therapy techniques for prostate cancer with prostate-rectum spacers: a systematic review," *International Journal of Radiation Oncology • Biology • Physics*, vol. 90, no. 2, pp. 278–288, 2014.
- [67] A. Alitto, R. Gatta, B. Vanneste et al., "PRODIGE: PRediction models in prOstate cancer for personalized meDICine challenge," *Future Oncology*, vol. 13, no. 24, pp. 2171–2181, 2017.

Research Article

Computational Analysis of Protein Structure Changes as a Result of Nondeletion Insertion Mutations in Human β -Globin Gene Suggests Possible Cause of β -Thalassemia

Talal Qadah ¹ and Mohammad Sarwar Jamal ^{1,2}

¹Department of Medical Laboratory Technology, Faculty of Applied Medical Science, King Abdulaziz University, Jeddah, Saudi Arabia

²Hematology Research Group, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Correspondence should be addressed to Talal Qadah; thqadah@kau.edu.sa

Received 30 June 2018; Revised 28 February 2019; Accepted 11 April 2019; Published 29 May 2019

Academic Editor: Magdalena Kasprowicz

Copyright © 2019 Talal Qadah and Mohammad Sarwar Jamal. 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.

Beta-thalassemia is described as a group of hereditary blood disorders characterized by abnormalities in the synthesis of beta chains of hemoglobin. These anomalies result in different phenotypes ranging from moderate to severe clinical symptoms to no symptoms at all. Most of the defects in hemoglobin arise directly from the mutations in the structural β -globin gene (*HBB*). Recent advances in computational tools have allowed the study of the relationship between the genotype and phenotype in many diseases including β -thalassemia. Due to high prevalence of β -thalassemia, these analyses have helped to understand the molecular basis of the disease in a better way. In this direction, a relational database, named HbVar, was developed in 2001 by a collective academic effort to provide quality and up-to-date information on the genomic variations leading to hemoglobinopathies and thalassemia. The database recorded details about each variant including the altered sequence, hematological defects, its pathology, and its occurrence along with references. In the present study, an attempt was made to investigate nondeletion mutations in the *HBB* picked up from HbVar and their effects using the *in silico* approach. Our study investigated 12 nucleotides insertion mutations in six different altered sequences. These 12 extra nucleotides led to the formation of a loop in the protein structure and did not alter its function. It appears that these mutations act as 'silent' mutations. However, further *in vitro* studies are required to reach definitive conclusions.

1. Introduction

Hemoglobinopathies are genetic disorders caused by single-gene variations in the α -like and β -like human globin gene clusters. These are the most common inherited disorders in humans with nearly 7% of the human population acting as carriers of the mutations in globin genes. Single nucleotide substitutions in the coding or regulatory regions of these genes can lead to varying degrees of defects in their expression [1]. The *HBB* gene belongs to the β -globin gene cluster that encodes β -globin polypeptide. It is located in the short arm of chromosome 11 and contains two introns and three exons. Molecular defects in human *HBB* may result in structural defects causing abnormality in hemoglobins, such as HbS, HbC, and HbD, or may result in absence or

reduced synthesis of β -globin chains causing β -thalassemia [2]. Mutations in the *HBB* may involve substitution, deletion, or insertion of one or multiple nucleotides within the gene or its flanking regions resulting in anemia and low RBC production [3]. β -Thalassemia is inherited as an autosomal recessive trait and its clinical manifestation can be divided as thalassemia major, intermedia, and thalassemia minor (trait) [4]. Some mutations in the *HBB* lead to complete inactivation of the gene resulting in the absence of β -globin chains (β^0), in turn, leading to the most severe form of thalassemia. Other mutations allow production of β -globin chains in varying proportions leading to β^+ thalassemia. This case is most commonly found in the Middle East, Central Asia, Mediterranean countries, India, and southern China and in some parts of Africa and South America [5]. It is one of the

most common genetic disorders caused by point mutations in Saudi Arabia causing variable phenotypic effects. These phenotypic severities may arise from defects in transcription, RNA processing, or translation of the HBB gene [6]. The most common mutations in most Arab-populated countries include IVSI-110(G>A), IVSI-1(G>A), IVSI-6(T>C), IVSII-1(G>A), IVSI-5(G>C), codon 5(-CT), and codon 39 (C>T) [7].

Due to the high prevalence of variable phenotypes of thalassemia and striking heterogeneity of its molecular defects, various strategies were employed to investigate the molecular mechanisms of this disease. Due to recent advances in computational tools, *in silico* analysis has become one of the chosen methods to investigate links between genomic and resulting phenotypic characteristics in thalassemia. HbVar (<http://globin.cse.psu.edu/globin/hbvar/>) is the oldest and most appreciated database of hemoglobin variants and thalassemia mutations established in 2001 [8]. It is a locus specific database, which was developed as a combined academic effort to keep a record of hemoglobin variants, new data entries, updates, and corrections. It provides high quality and up-to-date information on the genomic variations, associated phenotypic and hematological effects, pathology, frequency of different mutations, ethnic prevalence, and references [9]. HbVar has become a primary resource for the research community working on globin proteins and for physicians dealing with patients with hemoglobinopathies, to help them with making proper diagnoses.

The objective of this study is to investigate the effects of insertion mutations in the *HBB* exons using the *in silico* approach. We aimed to search the HbVar database to select sequences with uncharacterized insertion mutations and study their effects on the structure and function of β -globin protein.

2. Methodology

To perform this study, data from the HbVar database (<http://globin.bx.psu.edu/hbvar>) were used. Using this database, we identified uncharacterized *HBB* sequences containing nondeletion mutations and picked them up for *in silico* investigation. The sequence of a wild *HBB* (gene ID 3043) was taken as reference (<https://www.ncbi.nlm.nih.gov/gene/?term=3043>). The potential 5' and 3' sites in the gene sequences were identified using the Human Splicing Finder software [10]. These predictions were compared and multiple sequence alignment was performed using the available web server. An online translation tool (<https://web.expasy.org/translate/>) was used to translate the nucleotide sequences into amino acids, and the types of mutations as well as their respective positions were noted. Mutations located in the defined donor and acceptor splice sites were included here. The relative strength of the sites obtained from the bioinformatics tool was given values between 0 to 100. Splice sites with high value were considered functional. Homology modeling of the wild type and mutant sequences was performed to compare the 3D structures of the proteins using the SWISS-MODELER [11].

TABLE 1: Variation in the protein sequences.

Amino Acids	Sequence_ID
R31	Seq1, Seq3, Seq 4, Seq5, Seq6
S31	Seq2
L32	Seq1, Seq2, Seq3, Seq6
M32	Seq5
V32	Seq4
V33	Seq1, Seq2, Seq3, Seq 4, Seq5
I33	Seq6

3. Molecular Dynamic Simulation

The molecular dynamic simulation of the wild type and mutant proteins was performed using the GROMACS software [12]. The force field used for simulation was GRO-MOS96 53A6. The model structures (wild type and mutants) were solvated with water molecules in an octahedral box. Sodium ions (Na^+) were added for neutralization. The solvated systems were then subjected to 5000 steps of energy minimization using the steepest descent method to remove the steric clashes. Convergence was achieved in the energy minimization when the maximum force was smaller than $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. The NPT ensemble was performed for 1000 ps at 300 K. The production simulation was executed at 300 K for 30 ns for the wild type and mutant proteins. Protein visualization and superimposition were performed using PyMOL software (<https://pymol.org/2/>). Root mean square deviation was analyzed using the PyMOL align module.

4. Results

Multiple sequence alignment between reference and mutated gene sequences showed that at least 12 nucleotides have been inserted in the sequences 1, 2, 3, 4, 5, and 6 from position 93 to 104 as compared to the wild type (Figure 1). The inserted nucleotides showed maximum variation in the positions 93 and 94. In sequence 1 and sequence 2, adenine and thymine were inserted at position 93, respectively, whereas guanine appeared at position 93 in sequences 3, 4, 5, and 6. At position 94, thymine was inserted in sequences 1, 2, and 6, cytosine in sequence 3, guanine in sequence 4, and adenine in sequence 5. From position 95 to 104, the six sequences presented no variation among the inserted nucleotides, except for sequence 6 which had adenine inserted at position 97, instead of guanine like the remaining sequences.

Variations among these inserted nucleotides gave rise to changes in amino acids in protein sequences. Twelve nucleotides in total gave rise to four amino acids where different variants have been identified (Figure 2 and Table 1).

The construction of the 3D structure of the HBB protein from the given mutated sequences showed that the inserted amino acids formed a loop structure (secondary structure) in the protein. The homology modeling of the wild type and mutant type HBB protein showed that the mutated segment did not form any well-defined secondary structure.

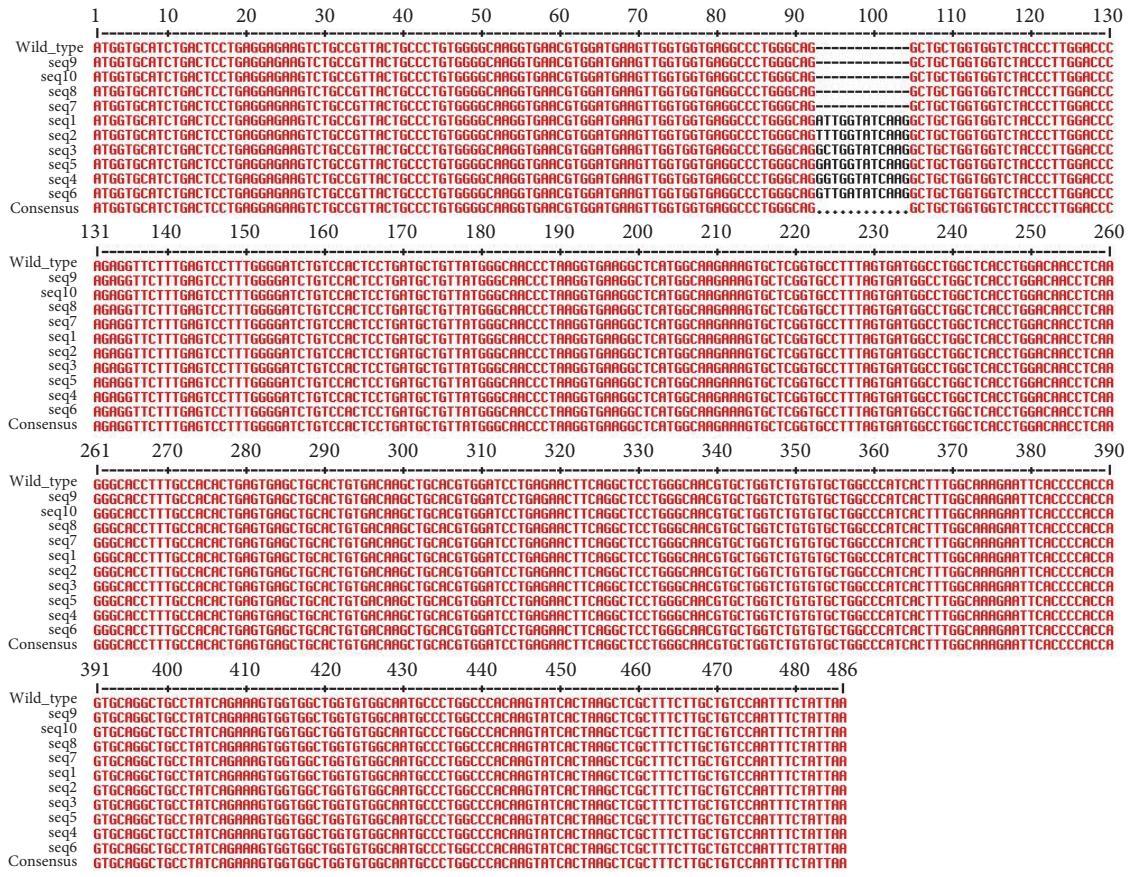


FIGURE 1: Multiple sequence alignment of the reference sequence and the multiple sequences, ATC, ATT, ACT, picked up from the HbVar database (after removing the introns). Twelve nucleotides have been inserted from position 93 to 104 in sequence numbers 1, 2, 3, 4, 5, and 6.

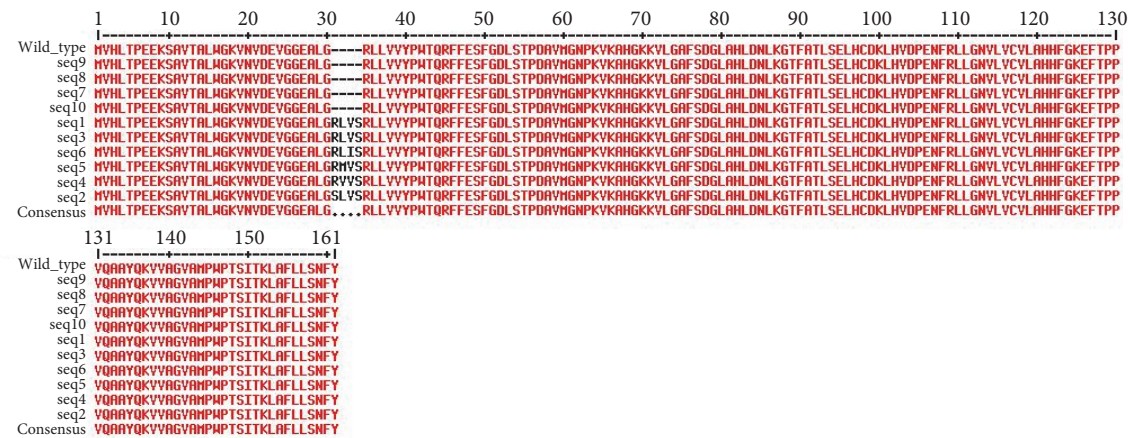


FIGURE 2: Protein sequence alignment of wild type and multiple sequences. Insertion of four amino acids has taken place at positions 31 to 34.

The mutated segment formed a loop and connected two α -helical chains (Figure 3).

The stability and properties of the wild type and its mutant structures were studied by explicit solvent MD stimulation. Root mean square deviations have been calculated between different structures to form a complete picture of deviation

in the structures of mutated proteins from the wild type (Table 2). The root mean square deviation (RMSD) analysis not only reflects the change of protein backbone versus simulation time, but also indicates the divergence of two structures. The RMSD of the homolog became stable at 30ns. The RMSD value of the wild type was 0.27 nm (Figure 4(a)).

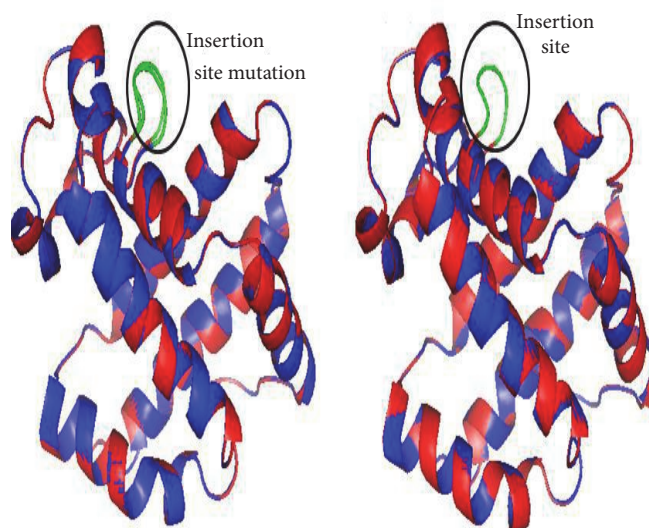


FIGURE 3: 3D structures of wild type (red color) and mutant type (blue color) of the HBB protein. Insertion sequence has been differentiated in green in the provided figures.

TABLE 2: Root mean square deviation (RMSD) between the model structures of the sequences.

Model	Model 2	RMS	Sec. str.	Mutation
WILD	Seq1	1.115	Loop	Insertion
	Seq2	0.690	Loop	Insertion
	Seq3	1.115	Loop	Insertion
	Seq4	0.533	Loop	Insertion
	Seq5	0.534	Loop	Insertion
	Seq6	1.107	Loop	Insertion
Seq2	Seq1	0.246	Loop	
	Seq3	0.246	Loop	
	Seq4	0.247	Loop	
	Seq5	0.256	Loop	
	Seq6	0.135	Loop	
Seq5	Seq4	0.013	Loop	

This result indicated that an accepted structure was obtained by the simulation that was reliable for further analyses. The root mean square fluctuation (RMSF) reflects the mobility of a certain residue around its mean position, which is another tool for studying the dynamics stability of the system. Although there were some deviations among the trajectories (especially in the loop region), the present data suggested less fluctuations, which further highlighted the reliability of the modeled structure (Figure 4(b)). The RMSF analysis can be used as a tool to describe local flexibility differences among residues throughout the MD simulation [62]. According to Figure 4(b), the wild type protein and Seq6 showed an overall higher degree of flexibility when compared to the mutants. A difference in RMSF value was seen on residues 76-91. The wild type and Seq6 proteins showed a fluctuation of 0.32nm, while the fluctuation at the same position on Seq1 was 0.1nm, thus indicating a flexibility loss. Whereas residues 54-80 showed

fluctuation values ranging from 0.15nm to 2.0nm in the wild type, while in the Seq1 and Seq6, these fluctuation values ranged from 0.09nm to 0.10nm. These results suggested that Seq1 affected the flexibility of the protein. This flexibility loss may affect protein function. Surprisingly, these fluctuations were found in the region away from the insertion or mutation site. However, Seq6 was found to have some fluctuation at this site as compared to the wild type and Seq1 (Figure 4(b)). We also analyzed the radius of gyration (Rg) values in the simulation. Rg is an indicator of structure compactness and overall dimension of the protein. It explains how regular secondary structures are compactly packed into the 3D structure of a protein. If a protein is folded well, it will likely maintain a relatively steady value of Rg, whereas it will change over time for unfolded proteins [13]. We found a stable Rg for Seq1 as compared to the wild type and Seq6 (Figure 4(c)). Low value of Rg for Seq1, as compared to the wild type and Seq6,

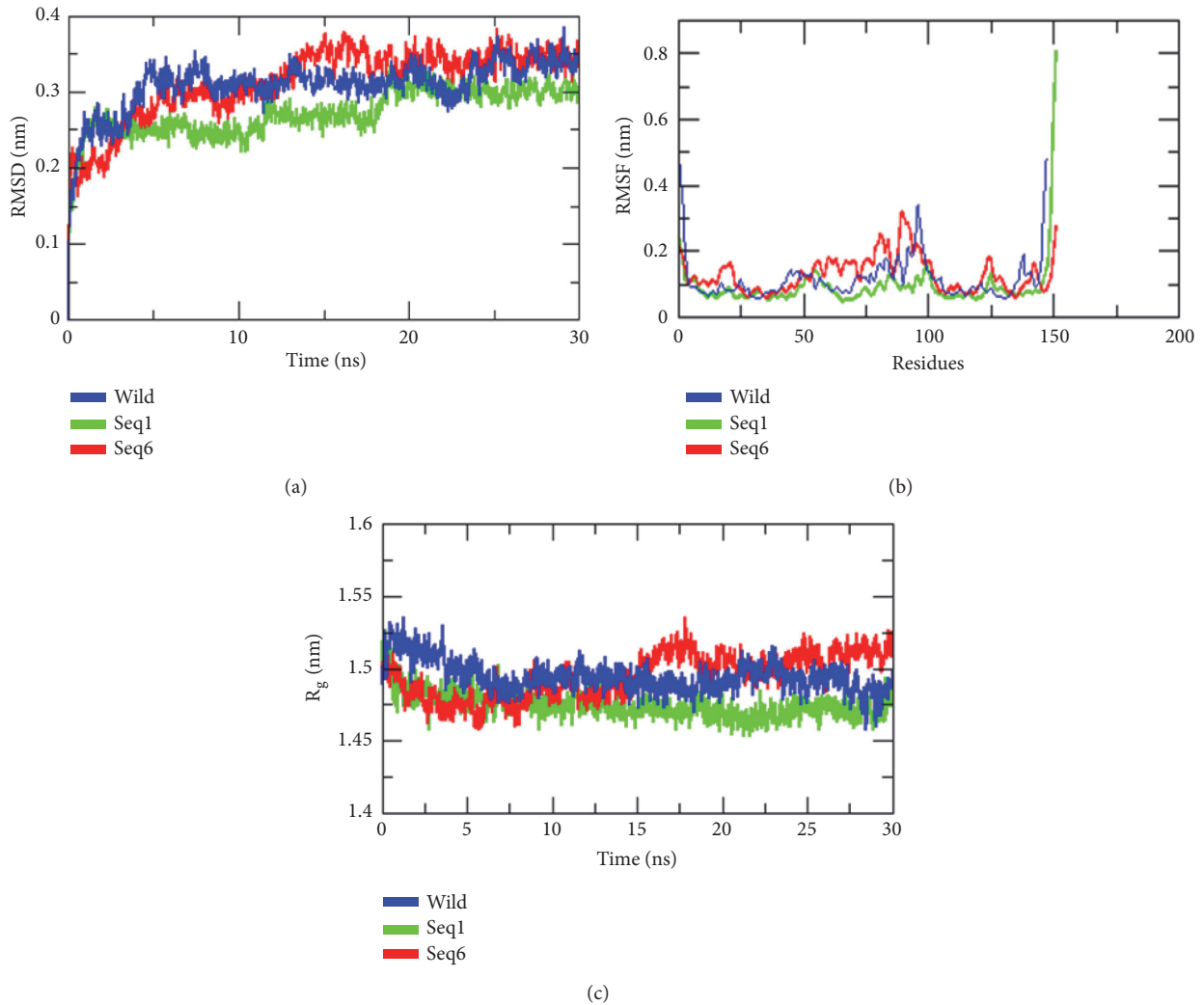


FIGURE 4: Molecular dynamic simulation: (a) root mean square deviation; (b) root mean square fluctuations; (c) radius of gyration of C α atoms of wild type and mutant proteins; time at 300 K.

suggested tight packing of these structures, making them relatively stable. The conformation of the modeled structure of Seq1 with different times, from 10ns, 20ns, and 30ns, was found to be similar. However, a slight difference has been observed at the insertion site (Figure 5).

5. Discussion

Constructing a relationship between the genotype and phenotype experimentally is an important aspect of research [14], but it can prove to be highly difficult, in particular, when studying a large number of subjects. The in silico analysis provides a solution here. It helps researchers analyze enormous amounts of data in biology to narrow down the positive leads that can be further analyzed experimentally for validation. This saves an extensive amount of labor, time, and costs. In silico analysis of large number of mutations is also easier and faster to accomplish, as this type of investigation is performed by comparing and studying alterations in the

nucleotide and/or amino acid sequences with the wild/native type and then correlating these alterations with the changed phenotypes [15, 16, 16].

Generally, β -thalassemia affects the people of the Gulf, Middle Eastern, and Mediterranean regions [17]. Especially in Saudi Arabia, β -thalassemia is prevalent, though there are variations in the frequency of the gene and in the type of mutations [18]. Many previous studies have screened and reported different mutations in the *HBB* and their frequencies in the Saudi population [19–21].

Most of the mutations affecting the expression of the *HBB* are linked to the gene physically and form different alleles of the gene, but some mutations that affect the gene expression, while also segregating the *HBB* cluster, have also been identified [22, 23]. Although the *HBB* is well characterized, some mutations in this gene recorded in the HbVar database are poorly understood and have not been properly studied before. In this study, we picked up such mutations and performed in silico analysis to understand their effect on the protein structure and function.

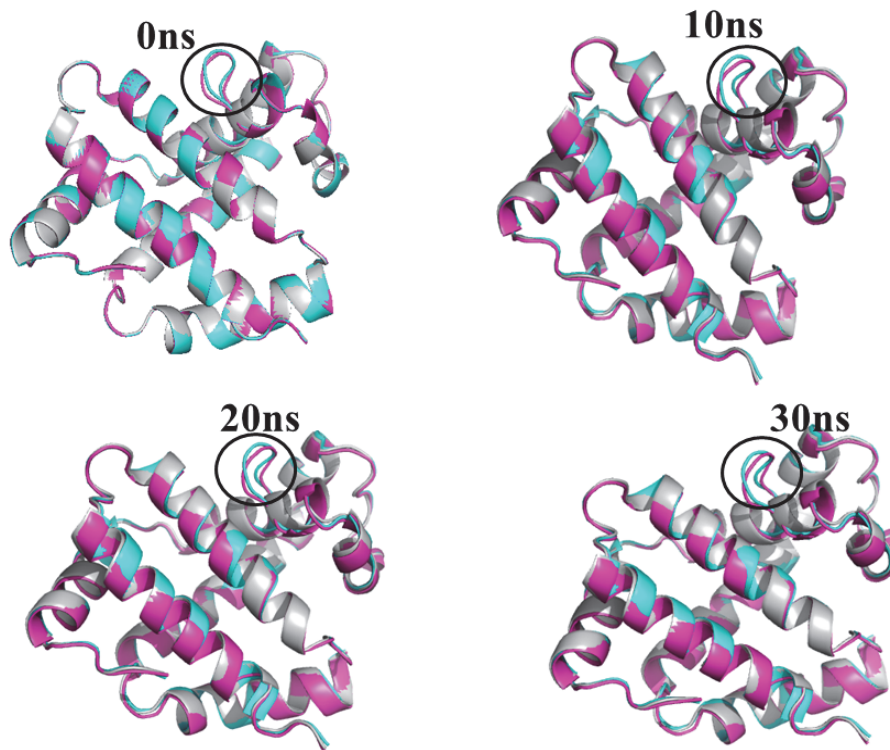


FIGURE 5: Conformation of model structures at 0ns, 10ns, 20ns, and 30ns. Cyan (Seq1), Magenta (Seq2), and Black brown (wild).

The nondeletion forms of defects in the *HBB* account for a large number of the β -thalassemia alleles [24]. These include small insertions and single nucleotide substitutions or deletions within the gene or its nearby sequences [23]. Some of the alleles of β -thalassemia are very mild, in that carriers (heterozygotes) of such alleles are almost normal with no apparent signs of the disease, except for imbalanced synthesis of globin chains [25, 26]. One of the fairly common 'silent' mutations in the Mediterranean population is 101 C \rightarrow T. It interacts with other more severe mutations of β -thalassemia to cause milder forms [24]. Other 'silent' mutations have been reported in the 5' UTR region of the *HBB* [25, 26].

The present study aimed to understand the nondeletion mutations using *in silico* tools. We identified 6 different sequences carrying insertion mutation of 12 nucleotides from position 93 to 104 in the gene sequence. Variations also occurred among the inserted nucleotides among these 6 different sequences. *In silico* analysis showed that these inserted nucleotides translated into 4 additional amino acids. These additional amino acids acquired a loop formation in the 3D structure of the protein (Figure 3). The difference in amino acids did not show any variation in the secondary loop structure acquired by these amino acids, but the molecular dynamics simulations presented evidence of effects caused by these mutations on the overall protein flexibility. The RMSF analysis showed a high degree of flexibility in the wild type protein and Seq6 in comparison to other mutant forms (Figure 4(b)). Additionally, significant flexibility loss

was seen in the Seq1 mutant form, especially in residues from positions 54-80 and 76-91. Though this effect was seen far from the site of mutation in Seq1, it is quite possible that insertion of new amino acids may likely disturb the internal environment of the protein, resulting in a whole new set of interactions between amino acids, which in turn, might have affected protein flexibility. This loss in flexibility may result in the loss of function of the protein. Studies suggest that a change in protein structure and, consequently, in function could be because of genetic variation in distal effect, because of the temporal effect due to folding of the protein sequence culminating into the final protein structure. This particularly happens because of the change in the properties related to the physicochemical, such as hydrophobicity, charge, and geometry due to the side chain of the amino acid residues. If such changes occur at critical sites, such as catalytic positions or important interacting sites called interfaces, then it is likely the reason for disease causing variations, which further tend to destabilize various hydrogen bonds and the salt bridge [25–28]. To further prove the effects of these insertion mutations on protein structure and function, more in depth analysis is required.

Furthermore, the studies on radius of gyration of C α atoms of the wild type and mutant proteins depicted in the 3D structure of the Seq1 mutant protein, are seen to be more compact and stable, as compared to the wild type and Seq6 proteins. This data suggests that the insertion mutations in *HBB* protein might be affecting its overall structure and function as shown in Seq1 and Seq6, but more

intensive studies are required to fully understand the scope of these effects. We are yet to determine how, if at all, these mutations affect the flexibility of HBB protein and whether this loss affects protein function and to what extent. In vitro studies will further assess the functional behavior of mutated proteins.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors have no conflicts of interest.

Acknowledgments

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. J-726-142-38. The authors, therefore, acknowledge with thanks DSR for technical and financial support. The authors also would like to thank Prof. Abdulwahab Noorwali for providing the authors with his technical support.

References

- [1] D. J. Weatherall, "Hemoglobinopathies worldwide: present and future," *Current Molecular Medicine*, vol. 8, no. 7, pp. 592–599, 2008.
- [2] D. J. Weatherall and J. B. Clegg, *The Thalassaemia Syndromes*, Blackwell Science, Oxford, UK, 2001.
- [3] R. Galanello and R. Origa, "Beta-thalassemia," *Orphanet Journal of Rare Diseases*, vol. 5, article 11, 2010.
- [4] A. S. Warsy, M. A. ElHazmi, A. K. Al Momin, A. AlHazmi, and A. Aleem, "Extensive polymorphisms in saudi beta thalassaemia patients," *Biosciences, Biotechnology Research Asia*, vol. 10, no. 1, pp. 127–132, 2013.
- [5] L. Zahed, "The Spectrum of beta-Thalassemia Mutations in the Arab Populations," *Journal of Biomedicine and Biotechnology*, vol. 1, no. 3, pp. 129–132, 2001.
- [6] M. H. Qari et al., "Regional consensus opinion for the management of Beta thalassaemia major in the Arabian Gulf area," *Orphanet Journal of Rare Diseases*, vol. 8, p. 143, 2013.
- [7] B. Giardine, S. van Baal, P. Kaimakis et al., "HbVar database of human hemoglobin variants and thalassaemia mutations: 2007 update," *Human Mutation*, vol. 28, no. 2, p. 206, 2007.
- [8] F. O. Desmet, D. Hamroun, M. Lalande, G. Collod-B eroud, M. Claustres, and C. B eroud, "Human splicing finder: an online bioinformatics tool to predict splicing signals," *Nucleic Acids Research*, vol. 37, no. 9, article e67, 2009.
- [9] T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch, "SWISS-MODEL: an automated protein homology-modeling server," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3381–3385, 2003.
- [10] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GRGMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008.
- [11] P. Rogerson and G. A. Artica, "Molecular size scaling in families of protein native folds," *Journal of Mathematical Chemistry*, vol. 49, no. 8, pp. 1493–1506, 2011.
- [12] T. Carllice-dos-Reis, J. Viana, F. C. Moreira et al., "Investigation of mutations in the HBB gene using the 1,000 genomes database," *PLoS ONE*, vol. 12, no. 4, 2017.
- [13] R. Dibyajyoti, "In silico analysis of β -thalassaemia mutations in india and its neighbouring south east asian countries," *Journal of Pharmaceutical, Chemical and Biological Sciences*, vol. 6, no. 3, pp. 188–198, 2018.
- [14] K. M. K. De Vooght, R. V. Wijk, and W. W. Van Solinge, "Management of gene promoter mutations in molecular diagnostics," *Clinical Chemistry*, vol. 55, no. 4, pp. 698–708, 2009.
- [15] S. AbdulAzeez and J. F. Borgio, "In-silico computing of the most deleterious nsSNPs in HBA1 gene," *PLoS ONE*, vol. 11, no. 1, p. e0147702, 2016.
- [16] T. Jalil, "IN-silico assessment of common β -globin gene mutations found in Khyber Pakhtunkhwa, Pakistan," *Pakistan Journal of Physiology*, vol. 14, no. 3, pp. 68–73, 2018.
- [17] A. Al-Sultan, S. Phanasgaonkar, A. Suliman, M. Al-Baqushi, Z. Nasrullah, and A. Al-Ali, "Spectrum of β -thalassaemia mutations in the eastern province of Saudi Arabia," *Hemoglobin*, vol. 35, no. 2, pp. 125–134, 2011.
- [18] F. H. Hasounah, S. A. Sejeny, J. A. Omer, J. M. Old, and R. W. A. Oliver, "Spectrum of β -thalassaemia mutations in the population of saudi arabia," *Human Heredity*, vol. 45, no. 4, pp. 231–234, 1995.
- [19] M. S. Akhtar, F. Qaw, J. Francis Borgio et al., "Spectrum of α -thalassaemia mutations in transfusion-dependent β -thalassaemia patients from the eastern province of Saudi Arabia," *Hemoglobin*, vol. 37, no. 1, pp. 65–73, 2013.
- [20] A. M. Abuzenadah, I. M. R. Hussein, G. A. Damanhour et al., "Molecular basis of β -thalassaemia in the western province of Saudi Arabia: Identification of rare β -thalassaemia mutations," *Hemoglobin*, vol. 35, no. 4, pp. 346–357, 2011.
- [21] A. K. Al-Ali, S. Al-Ateeq, and B. W. Imamwerdi, "Molecular bases of β -thalassaemia in the Eastern province of Saudi Arabia," *Journal of Biomedicine and Biotechnology*, vol. 2005, no. 4, pp. 322–325, 2005.
- [22] S. L. Thein and W. G. Wood, "The molecular basis of β thalassaemia, $\delta\beta$ thalassaemia, and hereditary persistence of fetal hemoglobin," in *Disorders of Hemoglobin: Genetics, Pathophysiology*, M. H. Steinberg, B. G. Forget, D. R. Higgs, and R. L. Nagel, Eds., pp. 323–356, Cambridge University Press, Cambridge, UK, 2009.
- [23] S. L. Thein, "The molecular basis of β -thalassaemia," *Cold Spring Harbor Perspectives in Medicine*, vol. 3, no. 5, Article ID a011700, 2013.
- [24] J. M. Gonzalez-Redondo, T. A. Stoming, A. Kutlar et al., "A C \rightarrow T substitution at nt -101 in a conserved DNA sequence of the promoter region of the β -globin gene is associated with 'silent' β -thalassaemia," *Blood*, vol. 73, no. 6, pp. 1705–1711, 1989.
- [25] E. Maragoudaki, E. Kanavakis, J. Traeger-Synodinos et al., "Molecular, haematological and clinical studies of the -101 C \rightarrow T substitution of the β -globin gene promoter in 25 β -thalassaemia intermedia patients and 45 heterozygotes," *British Journal of Haematology*, vol. 107, no. 4, pp. 699–706, 1999.
- [26] A. David and M. J. E. Sternberg, "The contribution of missense mutations in core and rim residues of protein-protein interfaces to human disease," *Journal of Molecular Biology*, vol. 427, no. 17, pp. 2886–2898, 2015.

- [27] M. Petukh, T. G. Kucukkal, and E. Alexov, "On human disease-causing amino acid variants: Statistical study of sequence and structural patterns," *Human Mutation*, vol. 36, no. 5, pp. 524–534, 2015.
- [28] Y. Yang, B. Chen, G. Tan, M. Vihinen, and B. Shen, "Structure-based prediction of the effects of a missense variant on protein stability," *Amino Acids*, vol. 44, no. 3, pp. 847–855, 2013.

Research Article

Automatic Detection of Hard Exudates in Color Retinal Images Using Dynamic Threshold and SVM Classification: Algorithm Development and Evaluation

Shengchun Long ¹, Xiaoxiao Huang ¹, Zhiqing Chen,²
Shahina Pardhan ³ and Dingchang Zheng ⁴

¹College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

²Eye Center, the Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou 310000, China

³Vision and Eye Research Unit (VERU), School of Medicine, Anglia Ruskin University, Chelmsford, UK

⁴Department of Medical Science and Public Health, Faculty of Medical Science, Anglia Ruskin University, Chelmsford, UK

Correspondence should be addressed to Shengchun Long; longsc@zjut.edu.cn and Xiaoxiao Huang; huangxiaoxiao@zjut.edu.cn

Received 12 March 2018; Revised 1 December 2018; Accepted 6 January 2019; Published 23 January 2019

Guest Editor: Fabien Scalzo

Copyright © 2019 Shengchun Long 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.

Diabetic retinopathy (DR) is one of the most common causes of visual impairment. Automatic detection of hard exudates (HE) from retinal photographs is an important step for detection of DR. However, most of existing algorithms for HE detection are complex and inefficient. We have developed and evaluated an automatic retinal image processing algorithm for HE detection using dynamic threshold and fuzzy C-means clustering (FCM) followed by support vector machine (SVM) for classification. The proposed algorithm consisted of four main stages: (i) imaging preprocessing; (ii) localization of optic disc (OD); (iii) determination of candidate HE using dynamic threshold in combination with global threshold based on FCM; and (iv) extraction of eight texture features from the candidate HE region, which were then fed into an SVM classifier for automatic HE classification. The proposed algorithm was trained and cross-validated (10 fold) on a publicly available e-optha EX database (47 images) on pixel-level, achieving the overall average sensitivity, PPV, and F-score of 76.5%, 82.7%, and 76.7%. It was tested on another independent DIARETDB1 database (89 images) with the overall average sensitivity, specificity, and accuracy of 97.5%, 97.8%, and 97.7%, respectively. In summary, the satisfactory evaluation results on both retinal imaging databases demonstrated the effectiveness of our proposed algorithm for automatic HE detection, by using dynamic threshold and FCM followed by an SVM for classification.

1. Introduction

Diabetic retinopathy (DR) is one of the major complications of diabetes that can lead to vision loss. The prevalence of DR is expected to grow exponentially, and the global population of DR patients is expected to increase to 191.0 million by 2030 [1]. The severity of DR is categorized according to the number of microaneurysms, hemorrhages, exudates, and neovascularization. The progress of DR is normally classified into normal retina, background DR, nonproliferative DR (NPDR), proliferative DR (PDR), and/or macular edema (ME) [2]. Regular screening to detect retinopathy can potentially reduce the risk of blindness of patients.

It is known that the occurrence of hard exudates (HE) is one of the main threats to vision loss especially when they occur near or on fovea [3]. Figure 1 shows an example of color retinal fundus image with HE. HE appears at late background and NPDR stages on the surface of retina as bright yellowish or white at different locations [4] and with variable shapes and sizes ranging from a few pixels to thousands of pixels in the retinal images. It is well accepted that the detection of HE in color retinal images plays a vital role in DR diagnosis and monitoring the progress of treatment. HE detection is therefore the main emphasis of this study.

HE is usually visually graded which is time-consuming and susceptible to observer errors [5]. The computer-aided

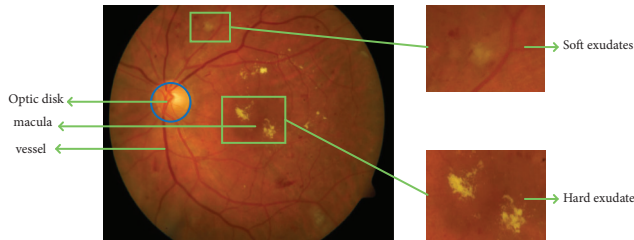


FIGURE 1: Example of retinal fundus image with exudates regions (zoom into the soft exudates region and hard exudates region).

detection of HE would potentially assist in achieving fast and accurate diagnosis. Many published algorithms have been developed for automatic HE detection in retinal images using four main strategies: thresholding, edge detection, region growing, and classification. Using the global threshold method and edge detection to achieve exudates detection automatically and accurately is very challenging due to the uneven intensity of the exudates, and the low contrast between exudates and retinal background [6]. Liu et al. [7] proposed a semiautomatic approach to detect low intensity exudates using local thresholds, which required the operator to select the local threshold manually based on the histogram of subimages. Region growing has also been implemented [8] to detect the exudates, which suffers from the difficulties of selecting the seed point and stopping criteria in region growing due to the wide variety of color distribution and nonhomogeneous illumination. Recently, different classification methods for exudates detection have been proposed to achieve fully automatic detection. An SVM classifier in combination with a Gaussian scale space approach has been used to differentiate between soft exudates, HE and outliers [9]. Other classification methods, including the bootstrapped decision trees [10], a Naïve-Bayes classifier optimized further by an adaptive boosting technique [11], and random forest method [12], have also been used. Unfortunately, the classification results from applying the above methods for HE detection are not clinically satisfactory enough due to various qualities of retinal images. This requires a more effective image segmentation method before classification.

Due to the large variety of the exudates in size, intensity, shape, and contrast, and the noise or artifacts during the image acquisition process, segmenting the small proportion of exudates pixels from the whole retinal images is challenging, leading to unsatisfactory detection accuracy for clinical applications. For general color image segmentation, fuzzy C-means (FCM), an unsupervised fuzzy clustering, has been widely used [13], where the global threshold is commonly used. However, using a global threshold may ignore the local details of the image. Dynamic threshold has been used, but this is more prone to shadow and man-made boundaries. It has been approved by Moghaddam and Cheriet [14] that using dynamic threshold in combination with the global threshold can significantly improve the effectiveness of segmentation of areas of interest in other fields, such as melasma image segmentation and cell cluster segmentation for in situ microscopy [15]. To date, the application of employing

dynamic threshold in combination with global threshold based on FCM has not been attempted in retinal image segmentation. In this study, we use this combined approach for determining candidates of HE from retinal images. After image segmentation, the segmented regions are normally classified into two disjoint classes using a neural network or support vector machines (SVM). Literature suggests that SVM is more practical than neural networks for small size of training data [16]. In machine learning, SVM is a supervised learning model with associated learning algorithms that analyze data for classification and regression analysis. The SVM is characterized by the ability to simultaneously minimize empirical errors and maximize the geometric edge region [17]. SVM is therefore implemented in this study.

The aim of this study was therefore to develop and evaluate a HE detection algorithm using dynamic threshold and FCM in combination with SVM. Specifically, after OD localization based on the image gray-scale value and retinal blood vessels distribution, the dynamic threshold matrix will be obtained using FCM in each subimage, which will be combined with global threshold matrix to obtain the exudates candidate regions, and followed with SVM classification to achieve automatic HE detection.

2. Methods

2.1. Retinal Image Databases. Our proposed algorithm was developed and tested on two publicly available databases of retinal images (the DIARETDB1 [18] and the e-optha EX [12]). DIARETDB1 database contains 89 color fundus images with 50° Field of View (FOV) and the size of 1500×1152 pixels, of which only 5 are normal; others contain different lesions. In this database, the different regions with HE have been manually labelled by four specialists from each image to determine whether a retinal image contains exudates. Human graders marked 571 regions as exudates DIARETDB1. The e-optha EX dataset contains 47 images with exudates regions and 35 exudates-free images, where only the 47 images with exudates regions were used in this study. Since the ground truth in DIARETDB1 is based on image-level and the e-optha EX is the only publicly available database which has provided pixel-level annotation for exudates segmentation, the e-optha EX was selected to train and cross-validate our algorithm on pixel-level, and the DIARETDB1 was used for additional independent test to discriminate whether a retinal image contains exudates.

2.2. Algorithm Development for Automatic HE Detection. As shown in Figure 2, our proposed algorithm for HE detection was composed of four main stages: (i) image preprocessing, (ii) OD localization, (iii) exudates candidate regions determination, and (iv) HE features extraction and classification. Matlab (2016a) was used in the environment of 64 bit Windows 10 operating system with 2.9 GHz Intel Core i5 CPU and 16GB memory.

2.2.1. Retinal Imaging Preprocessing. The preprocessing stage is crucial due to the intrinsic characteristics of retinal images. Retinal images often have poor and varying contrasts due

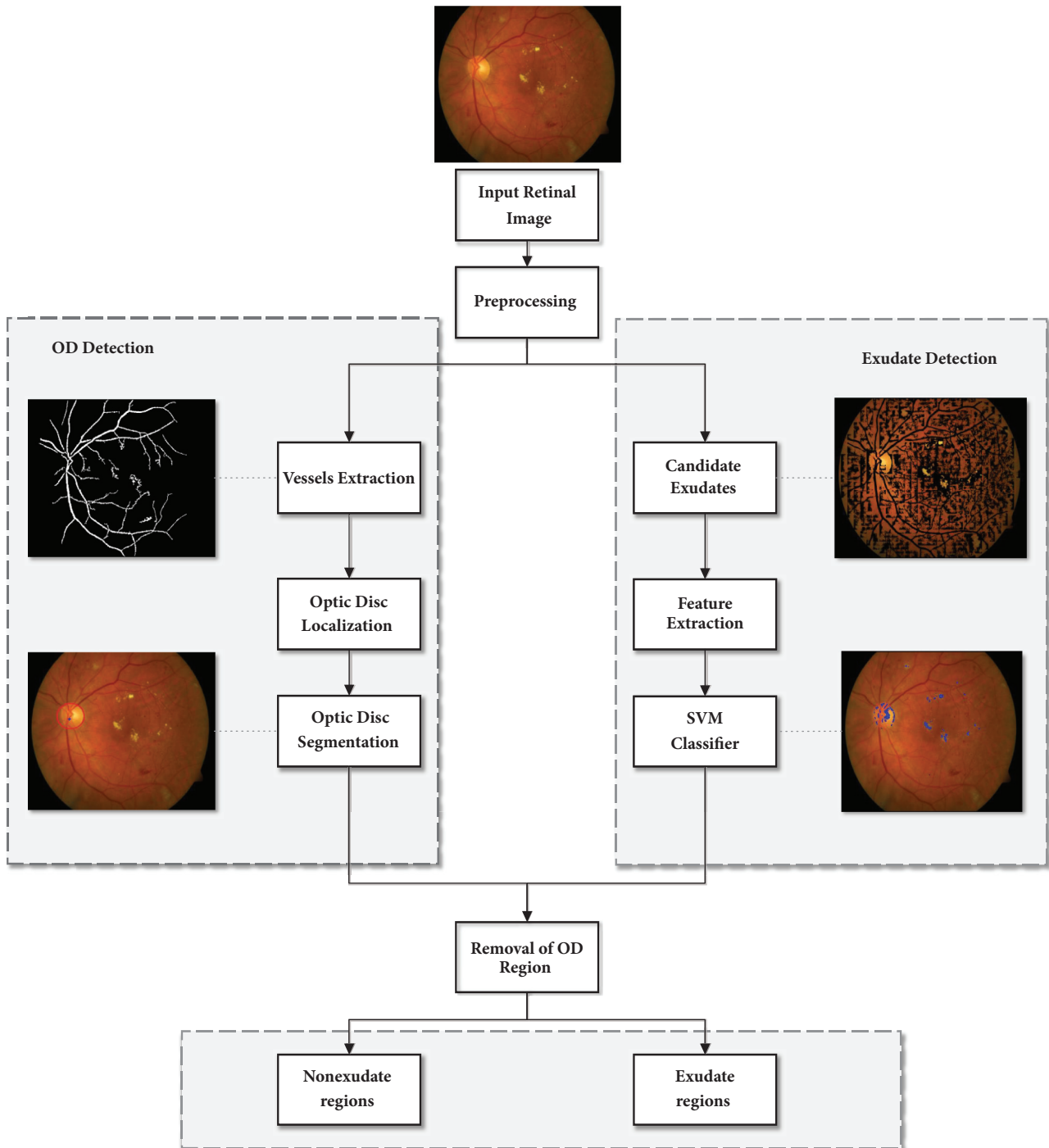


FIGURE 2: Flow chart of our proposed algorithm for automatic detection of HE.

to many factors including the noises introduced during the imaging acquisition process and the improper reflection of camera flash and retinal pigmentation. Additionally, the uneven illumination increases the intensity level near OD and decreases in regions away from OD. All these factors have significant impact on HE detection.

In our algorithm, color intensity normalization and contrast enhancement of the fundus photographs were operated with the size of retinal image rescaled to 512×512 pixels.

As proposed by Clara et al. [19], color normalization was performed by enhancing luminance plane of YIQ color model instead of enhancing each color plane of RGB. The modified process is as follows:

$$Y_{mod} = aY - bI - cQ \tag{1}$$

The modified color model YIQ was then converted back to RGB color model, as shown in the first three images in Figure 3. The empirical values of 1.8, 0.9, and 0.9 were used for

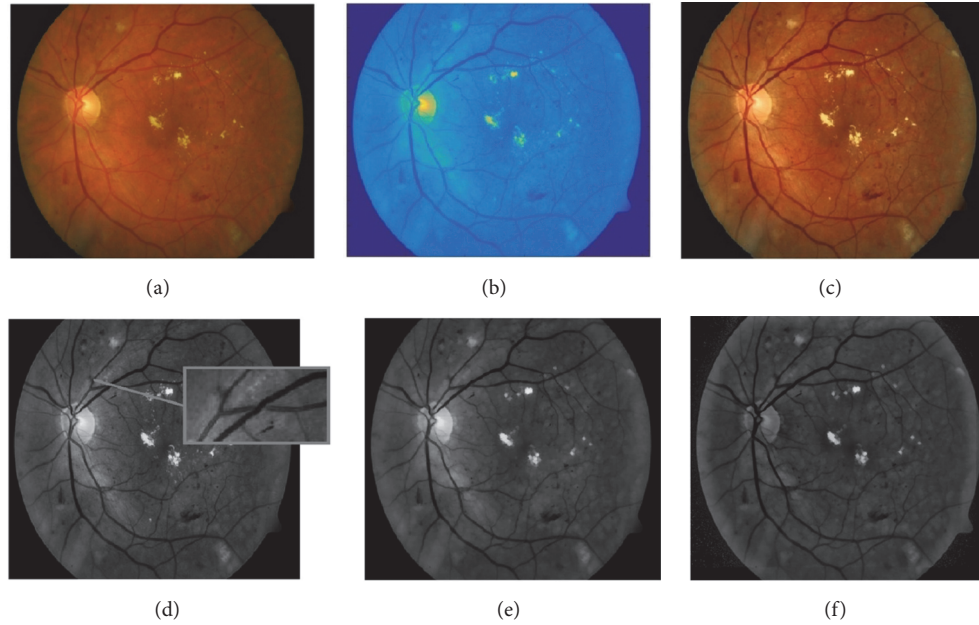


FIGURE 3: Example of retinal fundus image preprocessing. (a) Original retinal image. (b) Color normalized YIQ plane image. (c) Enhanced RGB plane image. (d) Green channel image after CLAHE (zoom into the blood vessels with brighter strip). (e) Green channel image after morphological opening. (f) Mean filtering of (e).

parameters a , b , and c , respectively, with which satisfactory results were achieved when the images were converted back to RGB color model, producing greater contrast between the HE and the background for the next step of HE detection.

It has been observed that the OD appears most contrasted in the green channel when compared to red and blue channels in the RGB retinal images [20]. Additionally, as the red channel is too saturated and the blue channel is the darkest color channel that does not contain much information, the green channel image was only used for the HE detection. Furthermore, in order to remove some bright strips down the central length of the blood vessels, the green plane of the image after contrast limited adaptive histogram equalization (CLAHE) was filtered by applying a morphological opening using a three-pixel diameter disc [21]. Next, the illumination equalization method in [22] was used to correct shade as follows:

$$I_{ie} = I - I_{bg} + u \quad (2)$$

where a mean filter of size 51×51 was applied to the green channel image I to generate a background image I_{bg} , which was then subtracted from the I to correct for shade variations. Finally, the average intensity u of green channel image I was added to keep the gray range same as in the I . The example images during the process are shown in Figure 3.

2.2.2. Optic Disc Detection and Masking. OD localization is an essential stage in our proposed algorithm because OD has similar properties as exudates in terms of color and brightness. The OD is a bright yellow disc in the retina where retinal blood vessels emerge. Therefore, the disc should be masked from the fundus image before further HE detection.

OD localization is relatively simple and fast in normal retinal images because it is where the largest cluster of brightest pixels is; however, this becomes more challenging in the images where the area of bright lesions is also large or OD is obscured by retinal blood vessels, for example, when there is a large hemorrhage on the disc [6]. In our proposed algorithm, the information of image brightness and retinal vasculature features were used for OD localization [23], which involved three steps: retinal blood vessels extraction, the center of OD localization, and OD segmentation.

Retinal Blood Vessels Extraction. In general, retinal blood vessels in the green channel fundus images do not have enough contrast in comparison with the surrounding background. An enhancement method of CLAHE [24] was applied to solve this problem. Next, a mean filtering with a 9×9 pixel-kernel was used to blur the image to reduce the noises. The retinal blood vessels image $I_{bv'}$ was obtained by subtracting the blurred image from the enhanced image by CLAHE, and the retinal blood vessels image I_{BV} was obtained by thresholding operator [25] applied to $I_{bv'}$. This process is shown in Figure 4, where two example images with different illumination conditions are given.

The Center of Optic Disc Localization. Retinal blood vessels originate from OD and spread outwards to the retina and the macular region. The vessels are generally aligned vertically in the vicinity of OD [26]. In order to obtain retinal blood vessels position information, a mean filter of size 61×61 was applied to the green channel image I to generate an average intensity image $I_{G'}$, and the $I_{BV'}$ (local average intensity of I_{BV}) was computed from the average intensity of the pixels

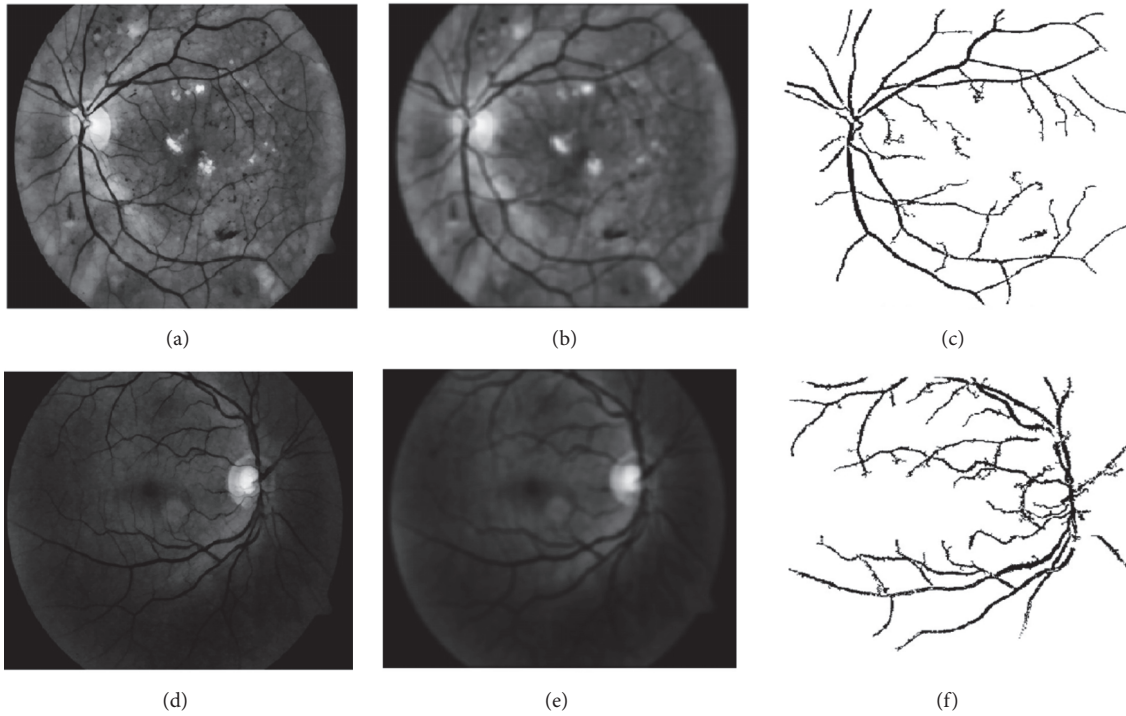


FIGURE 4: Examples of retinal blood vessels extraction on two retinal images with different illumination conditions. (a)+(d) Green channel images after CLAHE. (b)+(e) Mean filtering of (a) and (d) respectively. (c)+(f) Extracted retinal blood vessels.

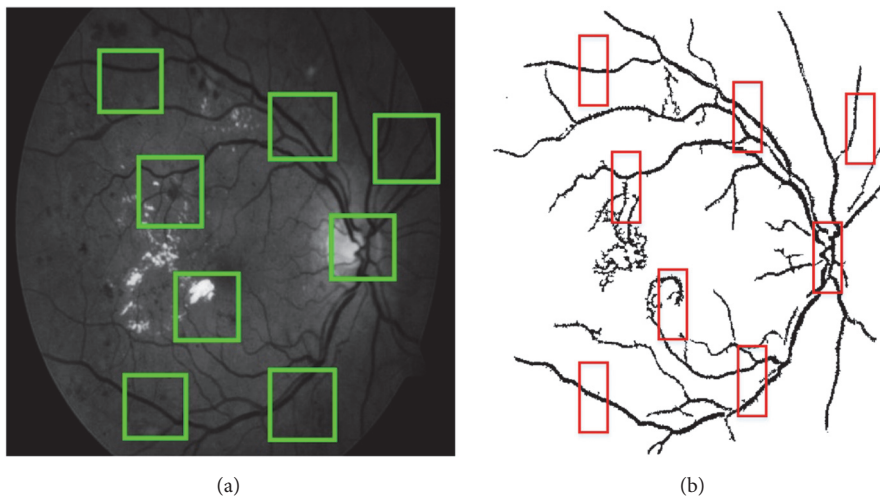


FIGURE 5: Illustration of the process for localizing the center of the optic disc. (a) Green channel image. The green boxes indicate the size of the mean filter. (b) The red boxes show the size of the window used to calculate the local average intensity of I_{BV} .

within an $N \times M$ window as illustrated in Figure 5. In this study, the window size N was between 50 and 60 pixels, and M was between 20 and 25 pixels. Next, in order to combine the brightness features and blood vessels position information from the green channel image, each pixel $I_{OD}(r, c)$ in the image was adjusted as follows:

$$I_{OD}(r, c) = I_{BV'}(r, c) - 1.2 * I_G'(r, c) \quad (3)$$

The image I_{OD} was then traversed with the minimum point identified as the center of OD, as shown in Figure 6(a).

Optic Disc Segmentation. To detect the OD boundary, the size $m \times n$ of region of interest (ROI) was defined based on the localization result of OD center, where m and n were one-ninth of the respective dimensions of the image multiplied. Since the OD in the retinal images has circular boundary shape [27], a circular Hough transform was applied to segment the OD boundary [23, 28, 29]. The Hough transform is a widely considered technique in Computer Vision and Pattern Recognition to detect geometrical features that can be defined through parametric equations like straight

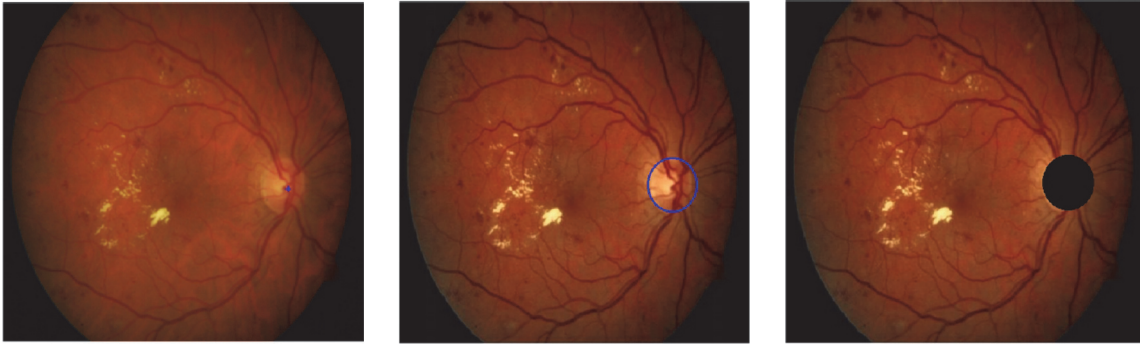


FIGURE 6: Demonstration of optic disc (OD) localization, segmentation, and masking. (a) Localization of OD center. (b) Segmentation of OD. (c) Masking OD.

lines and circles. The OD segmentation by applying Hough transform is shown in Figure 6(b). Lastly, the segmented OD was masked to avoid the interference to the following HE detection, as shown in Figure 6(c).

2.2.3. Detection of Hard Exudates. There were two main procedures. FCM clustering was firstly used to get the local dynamic threshold of each subimage, which was then combined with global threshold matrix to segment color retinal images. Next, an SVM classification was applied to distinguish exudates and nonexudates regions.

Retinal Image Segmentation Using FCM. The following describes the image segmentation process using the dynamic threshold in combination with global threshold based on FCM clustering:

(1) The retinal image was divided into a series of subimages (K subimages), and FCM algorithm was used to assign pixels in each subimage to different categories by using fuzzy memberships. FCM is an iterative optimization that minimized the cost function defined as follows:

$$J(U, V) = \sum_{i=1}^n \sum_{k=1}^c (u_{ki})^m \|x_i - v_k\|^2 \quad (4)$$

where u_{ki} represents the membership of pixel x_i in the k th cluster and v_k represents the clustering center of the k th cluster. Considering that the gray-scale value was used as the only feature for clustering, the midpoint of the clustering center line was used as the threshold in the segmentation sense, where the mean of the two clustering centers was obtained as the threshold of the subimage;

(2) The entire original retinal image pixels were classified in a similar way as above to obtain the global threshold and construct the global matrix S with the same size as the original image.

(3) After the interpolation of the thresholds of the respective subimages into a dynamic threshold matrix D of the same size as the entire original image, a mean filter of size 10×10 was applied to the matrix D .

(4) The final threshold matrix T was constructed as

$$T = kS + (1 - k)D \quad (5)$$

where the value of k was set to 0.1.

(5) The segmentation result was obtained by comparing the threshold matrix T with the retinal image.

The size of the subimage affects the retinal imaging segmentation results. Figure 7 shows the FCM clustering results for different subimage sizes. Taking both the running time and accuracy of local threshold into consideration, the size of 30×40 pixels was selected as the most suitable subimages size.

Feature Extraction for Hard Exudates Detection. In order to further segment the exudates regions from the exudates candidates, some significant features that were commonly used by eye care practitioners to visually distinguish HE from other types of lesions were extracted from each region and used as inputs of SVM. The key features included the following:

- (i) Mean green channel intensity (f1): a mean filter of size 3×3 was applied to the green channel image. This feature indicates the gray-scale intensity for all pixels. Again, only the features from the green channel were extracted.
- (ii) Gray intensity (f2): it was the gray-scale value of each pixel.
- (iii) Mean hue (f3), mean saturation (f4), and mean value (f5) of retinal image in HSV color model: a mean filter of size 3×3 was, respectively, applied to the three channel image I_h, I_s, I_v . Because exudates are the bright lesions on the surface of retina, the information about saturation and brightness (f4 and f5) of retinal image is also important.
- (iv) Energy (f6): energy was the sum of intensity squares of all pixel values in eight-convexity.
- (v) Standard deviation (SD) of the green channel image (f7): the morphological opening operation was applied to the green channel image to preserve foreground regions that have a similar shape to the structuring element or that completely contain the structuring element, while eliminating all the other regions of foreground pixels.
- (vi) Mean gradient magnitude (f8): it was the magnitude of the directional change in intensity of edge pixels.

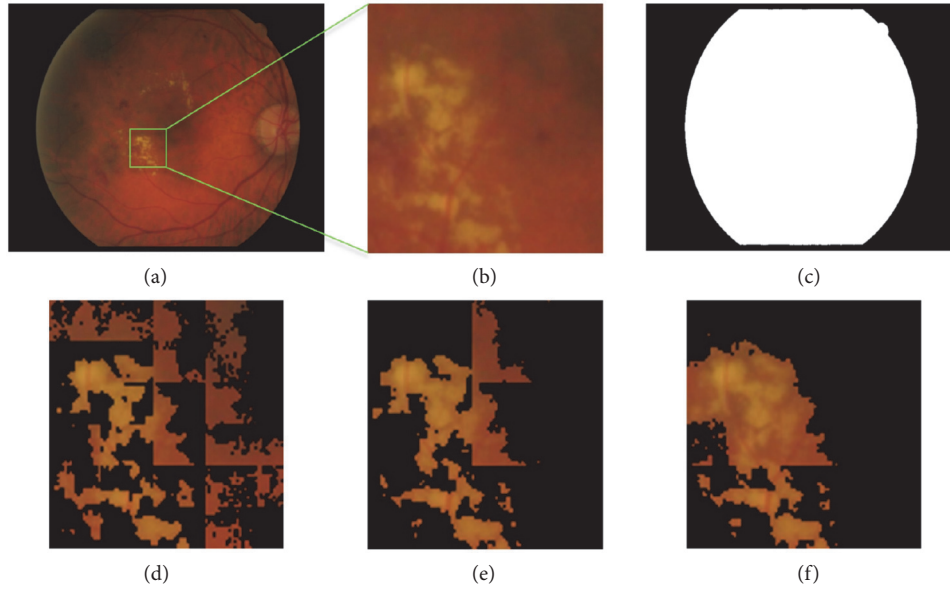


FIGURE 7: FCM clustering segmentation with different subimage sizes. (a) Original retinal image. (b) Zooming into the exudates region. (c) Segmentation result of original image using FCM. (d) Segmentation result with the subimage size of 15×20 pixels. (e) With the subimage size of 30×40 pixels. (f) With the subimage size of 60×80 pixels. Note: the background is filled with black.

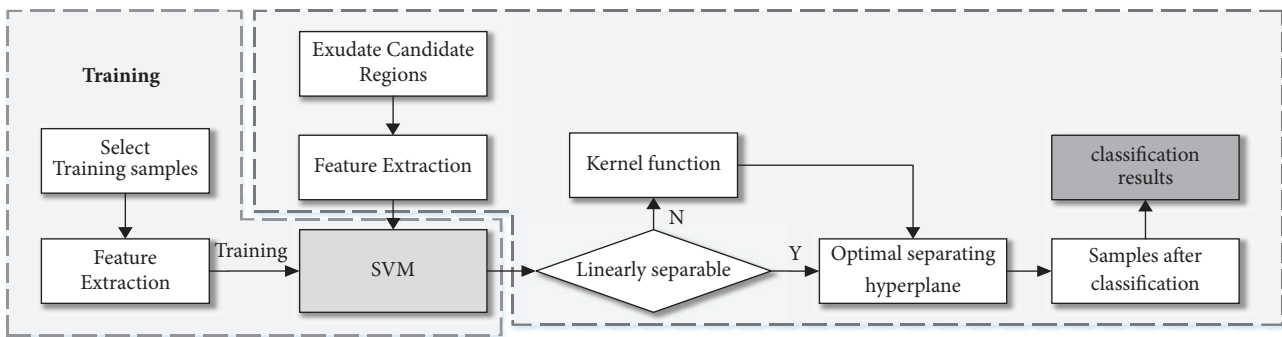


FIGURE 8: Flow chart of the SVM classification process.

It helps in distinguishing strong and blurry edges to differentiate between exudates and other bright lesions [3].

In comparison with other published algorithms where dozens of features were used [3, 9, 11], only eight key features were extracted in this study to reduce processing time while maintaining the accuracy of HE extraction.

SVM Classification. The flow chart of the SVM classification algorithm is shown in Figure 8. Briefly, the features extracted from the test images were fed into the trained SVM classifier to output a binary matrix representing the classification results. In this study, SVM was applied along with kernel function based on radial basis function (RBF). RBF kernel function has been widely used with two parameters (C and γ) obtained from the grid search method.

For training and cross-validation purposes, a few small regions (each image is about 1-10 regions, size between 50

and 250 pixels) of each of the 47 ground truth images were manually selected from the e-ophtha EX dataset as training samples. These selected regions have been divided into exudates regions and nonexudates regions. Using the e-ophtha EX dataset, a 10-fold cross-validation was applied to evaluate the ability of SVM classifier on pixel-level. The database was randomly split into 10 mutually exclusive subsets (the folds) $D_1, D_2, D_3, \dots, D_{10}$, approximately of equal size. The classifier was trained on 42 selected training images and tested the remaining 5 images to output a binary matrix representing the classification result. This procedure was repeated 10 times.

For each training image, a certain number of pixels (ranges from 50 to 250) were manually selected to construct training vector set. Each pixel constituted a feature vector from the eight key features. x_i represents the input sample feature vector set as follows:

$$x_i = (f_1, f_2, f_3, \dots, f_8) \quad (6)$$

The acquired training sample set (x_j, y_j) was input to train the SVM. y_j is the category flag:

$$y_j = \begin{cases} -1, & x_j \in A \\ 1, & x_j \in B \end{cases} \quad (7)$$

$j \in \{1, 2, \dots, W\}$, W is the dimension of the set of sample feature vectors. A and B , respectively, represent the HE and non-HE regions. In this study, around 7200 training vectors (or pixels) from the 42 training images were manually selected by an operator ($W=7200$).

The 10-fold cross-validation procedure was repeated five times by five different operators to manually select a region from each training image and then run the above procedure to evaluate the algorithm reliability.

2.3. Ensemble Evaluation Criteria. The evaluation criteria for HE identification were presented at two levels: pixel-level and image-level depending on which database was used. The pixel-level determination was based on whether each pixel of the classification result from the e-ophtha EX dataset has exudates in comparison with precisely labelled ground truth. The image-level HE detection was based on the presence or absence of HE in the classification result to determine whether a retinal image in the DIARETDB1 contains exudates.

2.3.1. Pixel-Level Evaluation on e-Ophtha EX Database. The evaluation can be classically performed by counting the number of pixels which were correctly classified. However, this approach was inappropriate for exudates segmentation evaluation because the contours of exudates do not match perfectly between the determinations from different observers, resulting in weak agreement on exudates determination. In this study, a hybrid validation method was used, where a minimal overlap ratio between ground truth and candidates was required.

Given the segmented exudates connected component set $\{D_1, D_2, \dots, D_N\}$ and the ground truth exudates component set $\{G_1, G_2, \dots, G_M\}$, we have the following.

A pixel was considered as a true positive (TP) if it belongs to

$$\begin{aligned} & \{D \cap G\} \cup \left\{ D_i \mid \frac{|D_i \cap G|}{|D_i|} > \sigma \right\} \\ & \cup \left\{ G_j \mid \frac{|G_j \cap D|}{|G_j|} > \sigma \right\} \end{aligned} \quad (8)$$

where $|\cdot|$ is the cardinal of a set and σ is a parameter ranging from 0 to 1. σ was set to 0.2 as used by Zhang et al. [12].

A pixel was considered as a false positive (FP) if it belongs to

$$\{D_i \mid D_i \cap G = \emptyset\} \cup \left\{ D_i \cap \bar{G} \mid \frac{|D_i \cap \bar{G}|}{|D_i|} \leq \sigma \right\} \quad (9)$$

or as a false negative (FN) pixel if it belongs to

$$\{G_j \mid G_j \cap D = \emptyset\} \cup \left\{ G_j \cap \bar{D} \mid \frac{|G_j \cap \bar{D}|}{|G_j|} \leq \sigma \right\} \quad (10)$$

The remaining pixels were considered as true negative (TN) pixels.

In this study, the four classes were clearly unbalanced as TP, FN, and FP were negligible in practice with respect to TN, computing the specificity, i.e., $TN/(FP+TN)$, and a receiver operating characteristic (ROC) curve, which is not appropriate. Sensitivity ($S = TP/(TP + FN)$), positive prediction value ($PPV = TP/(TP + FP)$), and F-score ($(2 \times S \times PPV)/(S + PPV)$) were therefore used as the performance of HE detection. The PPV combined both TP and FP, indicating the ratio of detected exudates pixels annotated as exudates pixels by specialists.

2.3.2. Image-Level Evaluation on DIARETDB1 Database. From clinical point of view, it would also be useful to evaluate the presence of exudates at the image-level, especially for DR screening applications. In order to evaluate the robustness of our algorithm, our algorithm was independently tested to determine whether the testing image contains exudates using the 89 images in the DIARETDB1 database, which has been labelled with ground truth at the image-level. As shown in Figure 9, each image was labelled by four specialists, if the ground truth confidence level is greater than or equal to 75%, the image was diagnosed with HE. At the image-level, if the image according to our algorithm and the ground truth both contain exudates region, the classification result for this retinal image was concluded as a TP. Matlab functionality for computing performance measures is publicly available at the DIARETDB1 web page [18]. For example, the processed image as Figure 9(d) was fed as an input into the evaluation protocol to obtain the evaluation outcomes (TP, TN, FP, FN). Three different evaluation parameters, including the sensitivity, specificity, and accuracy, were then used to determine the overall performance of HE detection. Their calculation formulas are shown as follows:

$$\begin{aligned} Accuracy &= \frac{TN + TP}{TP + FP + TN + FN} \\ Specificity &= \frac{TN}{TN + FP} \\ Sensitivity &= \frac{TP}{TP + FN} \end{aligned} \quad (11)$$

2.4. Data Statistical Analysis. For the 10-fold cross-validation using the e-ophtha EX database, the sensitivity, PPV, and F-score were calculated for each image, with their mean and standard deviation (SD) across all the images calculated. Their SD between the five repeats performed by the five different operators were also calculated to demonstrate the reliability of our algorithm. ANOVA analysis was then performed to check the repeatability between the five repeats. For the independent test on the DIARETDB1 database, the overall mean sensitivity, specificity, and accuracy were calculated

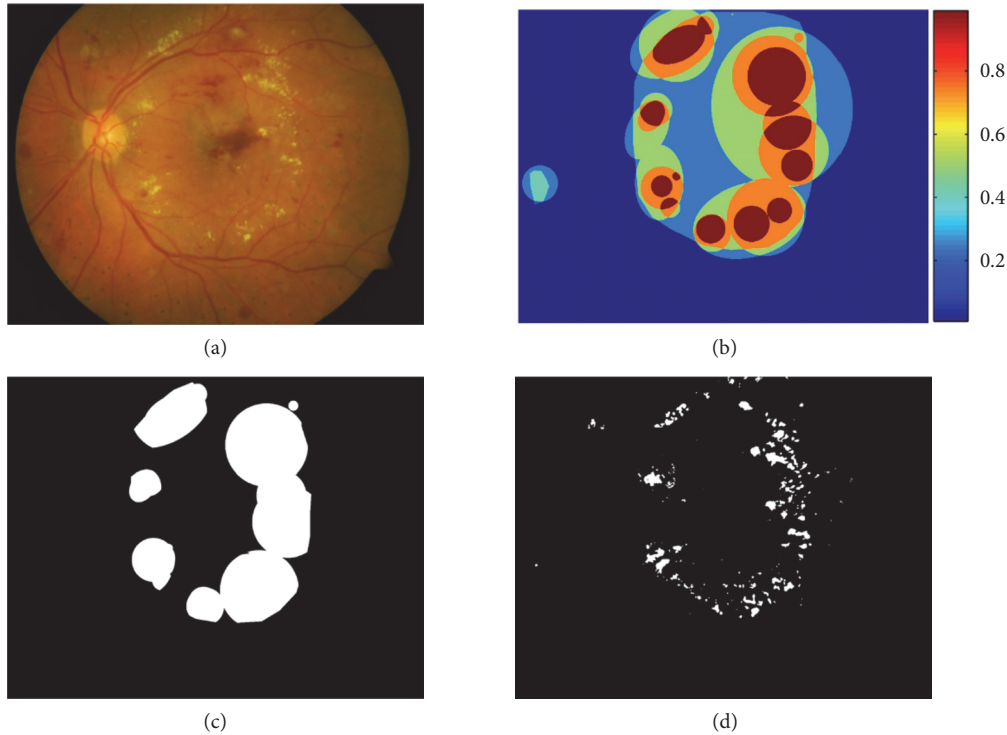


FIGURE 9: Example of one retinal image in DIARETDB1 database. (a) Original image. (b) Exudates regions labelled by four specialists (color decodes the ground truth confidence). (c) The exudates regions where the ground truth confidence level $\geq 75\%$. (d) Segmentation result of our algorithm.

from all the 89 images, which were simply compared with other published results using the same database.

3. Results

3.1. 10-Fold Cross-Validation Results on the e-Ophtha EX Database. Statistical analysis showed that there was no significant difference between the five repeat measurements for the evaluation parameters (all $p > 0.8$). As shown in Figure 10(a), the overall mean and SD of sensitivity, PPV, and F-score across all the images e-ophtha EX database were $76.5\% \pm 15.1\%$, $82.6\% \pm 16.7\%$, and $76.7\% \pm 12.7\%$. The measurement repeatability (SD of the five measurements) of sensitivity, PPV, and F-score for each individual image is shown in Figure 10(b). It ranged from $0.3\% \sim 16\%$, indicating that our algorithm proposed in this study for HE detection is sufficiently stable.

Table 1 also shows our algorithm achieved a higher score of PPV values in comparison with other published results also using pixel-level evaluation on the same database, indicating that our method could distinguish HE from other bright lesions more effectively. To visualize the HE detection from different retinal images, three example images are provided in Figure 11. Only the exudates regions (the left three subfigures) were cropped from the original retinal images. Figure 11(a4, b4, c4) shows the results of validation results at the pixel-level with $\sigma = 0.2$, where the green, red, blue, and black pixels are the TP, FN, FP, and TN pixels, respectively. It can be seen that

TABLE 1: Overall performance comparison of our proposed algorithm with published studies for HE detection on e-ophtha EX dataset.

Methods	Sensitivity	PPV	F-score
Zhang et al. (2014) [12]	74%	72%	73%
Welfer et al. (2010) [38]	79%	55%	69%
Imani et al. (2016) [39]	80.32%	77.28%	-
Liu et al. (2017) [30]	76%	75%	76%
Kusakunniran et al. (2018) [40]	56.4%	-	-
Our proposed algorithm	76.5%	82.7%	76.7%

most of the large exudates could be identified successfully. Some FPs with wrongly detected HEs could be caused by the presence of other bright lesions, such as cotton wool spots and drusens. Some small HE pixels were missed by our proposed algorithms because of their low contrasts.

3.2. Validation Results on DIARETDB1 Database. Table 2 lists the overall evaluation performance of our proposed algorithm using image-level evaluation in the DIARETDB1 database. The overall mean sensitivity, specificity, and accuracy were 97.5% , 97.8% , and 97.7% , respectively, which compared well with other published results. Some example images from DIARETDB1 database are shown in Figure 12 to demonstrate whether an image has been correctly or wrongly detected with exudates.

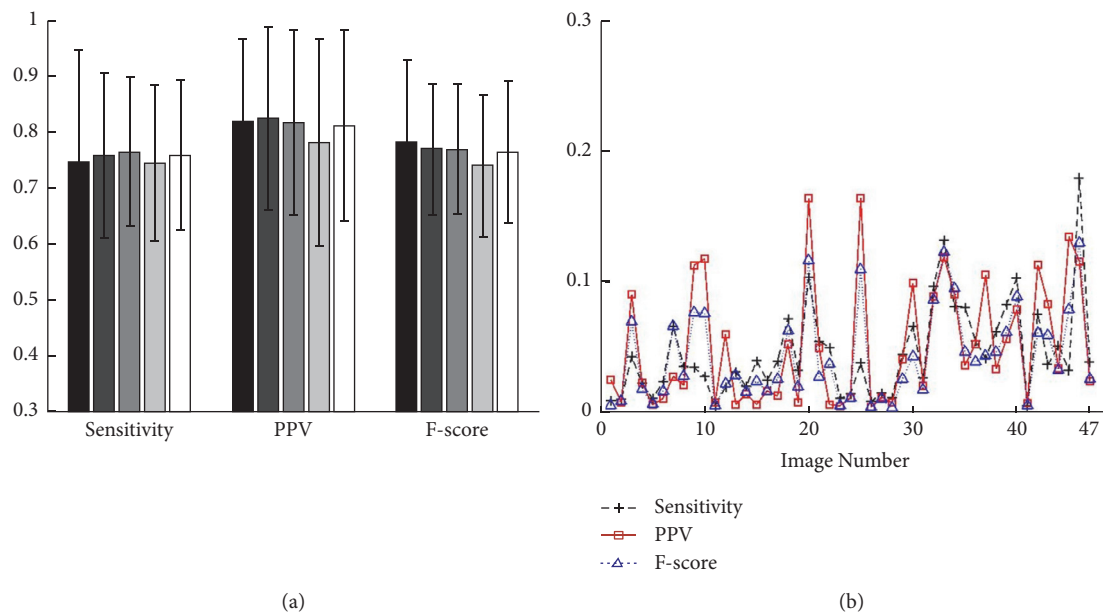


FIGURE 10: Data statistical analysis. (a) The overall mean and standard deviation of sensitivity, PPV, and F-score for all the images in the e-optha EX database. They are given separately for the five repeat measurements by five different operators. (b) The repeatability (standard deviation of 5 repeat measurements by 5 operators on each image) of sensitivity, PPV, and F-score of our algorithm on the e-optha EX database.

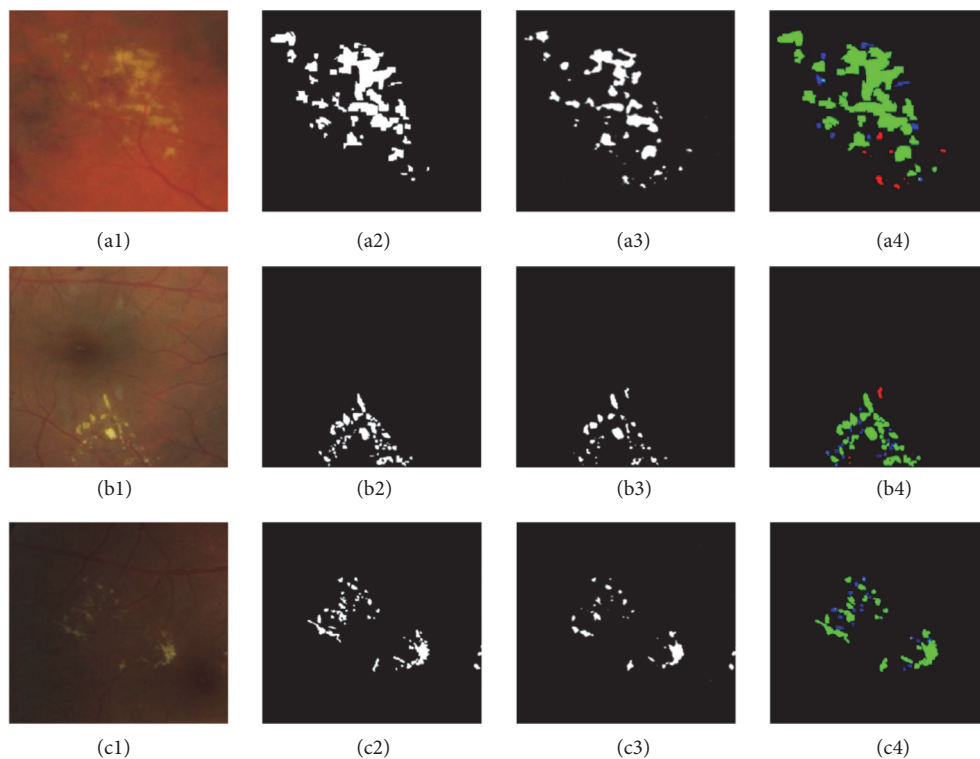


FIGURE 11: Example of pixel-level validation from three example images. (a1, b1, c1) Exudates regions cropped from the original retinal fundus image. (a2, b2, c2) The ground truth images in e-optha EX dataset. (a3, b3, c3) The segmented results with our algorithm. (a4, b4, c4) The results of pixel-level validation. The green, blue, red, and black pixels are the TP, FN, and FP, and TN pixels, respectively.

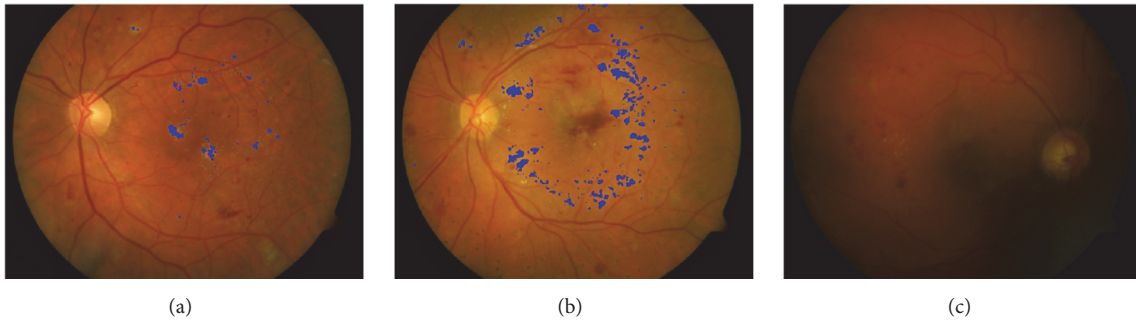


FIGURE 12: Example images showing the results of HE detection in the DIARETDB1 database. (a) HE was correctly detected. (b) Some small HE was missed in this image. (c) Failed to detect HE correctly.

TABLE 2: Overall performance comparison of our proposed algorithm with published algorithms for HE detection on DIARETDB1 database.

Methods	Sensitivity	Specificity	Accuracy
Harangi et al. (2014) [11]	92%	68%	82%
Haloi et al. (2015) [9]	96.54%	93.15%	-
Imani et al. (2016) [39]	89.01%	99.93%	-
Liu et al. (2017) [30]	83%	75%	79%
Rekhi et al. (2017) [31]	91.67%	92.68%	92.13%
Fraz et al. (2017) [10]	92.42%	81.25%	87.72%
Kusakunniran et al. (2018) [40]	89.1%	99.7%	96.2%
Our proposed algorithm	97.5%	97.8%	97.7%

4. Discussion and Conclusion

We have developed and evaluated an automatic retinal image processing algorithm to detect HEs using dynamic threshold, FCM and SVM. The color retinal images were segmented using dynamic threshold in combination with the global threshold, and the segmented regions were classified into two disjoint classes (exudates and nonexudates pixels) using SVM. The algorithm was tested on two publicly available databases (DIARETDB1 and e-optha EX database), and the evaluation results quantitatively demonstrated that our proposed algorithm is reliable in terms of repeatability and also achieved high accuracy for HE detection.

It is known that OD has similar properties with exudates in terms of color and brightness, masking or removing OD from the fundus image before further processing for HE detection is therefore important, which would improve the HE detection accuracy [10, 30, 31]. This study has presented a method for OD localization by combining the information of brightness and retinal vasculature features. Our method is inspired by Medhi et al. [23] who used a vertical Sobel mask and considered OD as the region with maximum value of edge pixels. Unlike other methods with more complicated process [29, 32], we only need to traverse the entire image twice to find the pixel with the largest gray-scale value and the most densely distributed of blood vessels, achieving fast localization of OD. Rahebi et al.'s [32] study applied the firefly algorithm and reported a success rate of 94.38% for OD localization in the DIARETDB1 database. Using the same

database in this study, an accuracy of 89.9% was achieved. Although our OD detection was slightly less accurate than theirs, our method was much simpler and faster. More importantly, our method is very suitable for the application of HE detection as an intermediate step, and the relatively high accuracy was comparable with many other complex algorithms with specific aim for OD detection.

FCM has been implemented in exudates segmentation algorithms [13, 33]. Sopharak et al. [34] proposed an FCM based method to determine whether a pixel has exudates or not, but they only achieved moderately acceptable segmentation result with the sensitivity of 80% on DIARETDB1 database. Global threshold is commonly used for image segmentation. However, using the global information only may ignore the details from those small HEs. If the gray-scale value of background is constant, using global threshold for segmentation would achieve satisfactory results. However, in many cases, because the contrast between the object and background changes in different regions, the gray-scale value of background varies, resulting in a poor segmentation outcome. In other fields, it has been shown that using dynamic threshold in combination with the global threshold can significantly improve the segmentation results. For instance, the combined thresholds have been applied successfully to distinguish the human skin in color image and melasma image segmentation, where good segmentation results were achieved [35, 36]. The key advantage of combining the image's global information with the local details could overcome the problems associated with using local threshold alone. After employing this combined approach, the satisfactory evaluation results (97.5% of sensitivity on DIARETDB1 database, 76.5% of sensitivity on e-optha EX database) were achieved in this study. It is noted that only one feature (the gray-scale value of retinal images) was input into the FCM. More input features and the FCM clustering combined with the morphological technique could be also considered in future to achieve higher accuracy.

SVM classifier was selected in this study to distinguish true exudates regions from nonexudates regions. One of the key reasons is that the sample size of retinal image database used in this paper is not large enough. Using SVM was expected to have better classification result because SVM can apply the nonlinear relationship between data

and features better than other classifiers [16]. Secondly, SVM can have rapid training phase [17]. Akram et al. [3] proposed a hybrid classifier as a GMM and SVM for exudates detection; however, training GMM model and finding the optimized parameters for GMM were complicated. In this study, the combined approach using FCM and SVM required less computational expenses. Only eight key features were used when compared with other algorithms with dozens of features [9, 11]. The distinguishing features of HE, in comparison with other lesions as having sharper margins and bright yellow color, enabled the most representative of eight features to be used to achieve more efficient process while maintaining the accuracy of HE extraction. Jaya et al. [37] proposed an expert decision-making system designed using a fuzzy support vector machine (FSVM) classifier to detect hard exudates. Color and texture features are extracted from the images as input to the FSVM classifier. However, using one classifier to detect HE and candidate regions of HE not extracted in advance, the computational complexity of the classifier will increase greatly, resulting in low final detection efficiency.

One limitation of our algorithm is that its performance depends on the OD detection and retinal blood vessels removal. Since the applied OD detection was quite simple in this study, the performance of our method could be further improved by improving the robustness of OD localization and blood vessel detection. Secondly, while the retinal image quality was very poor, such as the whole image is very dark with large artificial shadow (e.g., image029, image047 in DIARETDB1 database), and the contrast between HE and the background is not strong enough (e.g., image044, image052 in DIARETDB1 database), the HE detection result was poor. In addition, some big and bright cotton wool spots have been wrongly detected as HE and some small HE were ignored. In future studies, we will improve algorithms to achieve more effective detection. Furthermore, we suggest more evaluations to be carried out with the proposed algorithms on other clinically available data. Such tests could contribute to further improvements on the algorithms, resulting in more robust and more accurate detection. In summary, the satisfactory evaluation results on both retinal imaging databases demonstrated the effectiveness of employing dynamic threshold, fuzzy C-means and SVM in our proposed automatic HE detection methods, providing scientific evidence that it has potential for clinical DR diagnosis.

Data Availability

The DIARETDB1 and the e-ophtha EX databases used to support this study are from freely available databases of retinal images at <http://www.it.lut.fi/project/imageret/diaret/db1/> and <http://www.eophtha.com>, which have been cited. The processed data during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Shengchun Long and Xiaoxiao Huang conceived and designed the experiments. Xiaoxiao Huang performed the experiments. Zhiqing Chen and Xiaoxiao Huang analyzed the results. All authors reviewed the manuscript. Shahina Pardhan and Dingchang Zheng approved the final version. Zhiqing Chen, Shahina Pardhan, and Dingchang Zheng contributed equally to this work.

Acknowledgments

Authors thank e-ophtha EX dataset (<http://www.eophtha.com>) and DIARETDB1 database (<http://www.it.lut.fi/project/imageret/diaret/db1/>) for providing the fundus images for this work.

References

- [1] Y. Zheng, M. He, and N. Congdon, "The worldwide epidemic of diabetic retinopathy," *Indian Journal of Ophthalmology*, vol. 60, pp. 428–431, 2012.
- [2] U. R. Acharya, E. Y. Ng, J. H. Tan, S. V. Sree, and K. H. Ng, "An integrated index for the identification of diabetic retinopathy stages using texture parameters," *Journal of Medical Systems*, vol. 36, pp. 2011–2020, 2012.
- [3] M. U. Akram, A. Tariq, S. A. Khan, and M. Y. Javed, "Automated detection of exudates and macula for grading of diabetic macular edema," *Computer Methods and Programs in Biomedicine*, vol. 114, no. 2, pp. 141–152, 2014.
- [4] E. Group, "Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified airline house classification," *Ophthalmology*, vol. 98, no. 5, pp. 786–806, 1991.
- [5] A. Fagotcampagna, I. Romon, N. Poutignat, and J. Bloch, "Non-insulin treated diabetes: relationship between disease management and quality of care," *La Revue Du Praticien*, vol. 57, pp. 2209–2216, 2007.
- [6] H. Li and O. Chutatape, "Fundus image features extraction," in *Proceedings of the International Conference of the IEEE In Engineering in Medicine and Biology Society*, vol. 4, pp. 3071–3073, 2000.
- [7] Z. Liu, C. Opas, and S. M. Krishnan, "Automatic image analysis of fundus photograph," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 524–525, November 1997.
- [8] H. Li, "Model-based approach for automated feature extraction in color fundus images," in *Proceedings of the 9th International Conference on Computer Vision*, vol. 1, Nice, France, 2003.
- [9] M. Haloi, S. Dandapat, and R. Sinha, "A gaussian scale space approach for exudates detection, classification and severity prediction," *Computer Science*, vol. 56, pp. 3–6, 2015.
- [10] M. M. Fraz, W. Jahangir, S. Zahid, M. M. Hamayun, and S. A. Barman, "Multiscale segmentation of exudates in retinal images using contextual cues and ensemble classification," *Biomedical Signal Processing and Control*, vol. 35, pp. 50–62, 2017.
- [11] B. Harangi and A. Hajdu, "Automatic exudate detection by fusing multiple active contours and regionwise classification," *Computers in Biology and Medicine*, vol. 54, pp. 156–171, 2014.
- [12] X. Zhang, G. Thibault, E. Decenci ere et al., "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Medical Image Analysis*, vol. 18, no. 7, pp. 1026–1043, 2014.

- [13] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham, "Automatic recognition of exudative maculopathy using fuzzy c - means clustering and neural networks," in *Proceedings of the Medical Image Understanding Analysis Conference*, vol. 3, pp. 49–52, 2001.
- [14] R. F. Moghaddam and M. Cheriet, "A multi-scale framework for adaptive binarization of degraded document images," *Pattern Recognition*, vol. 43, no. 6, pp. 2186–2198, 2010.
- [15] E. Espinoza, G. Martinez, J.-G. Frerichs, and T. Scheper, "Cell cluster segmentation based on global and local thresholding for in-situ microscopy," in *Proceedings of the 2006 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 542–545, April 2006.
- [16] A. Osareh, M. Mirmehdi, B. T. Thomas, and R. Markham, "Comparative Exudate Classification Using Support Vector Machines and Neural Networks," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI*, Springer Berlin Heidelberg, Berlin Heidelberg, 2002.
- [17] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
- [18] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen et al., "The DIARETDB1 diabetic retinopathy database and evaluation protocol," in *Proceedings of the 18th British Machine Vision Conference (BMVC '07)*, pp. 1–10, September 2007.
- [19] C. I. e. a. Sánchez, "A novel automatic image processing algorithm for detection of hard exudates based on retinal image analysis," *Medical Engineering & Physics*, vol. 30, p. 350, 2008.
- [20] R. J. Winder, P. J. Morrow, I. N. McRitchie, J. R. Bailie, and P. M. Hart, "Algorithms for digital image processing in diabetic retinopathy," *Computerized Medical Imaging and Graphics the Official Journal of the Computerized Medical Imaging Society*, vol. 33, no. 8, p. 608, 2009.
- [21] D. Marin, A. Aquino, M. E. Gegundezarias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Transactions on Medical Imaging*, vol. 30, Article ID 146158, pp. 146–158, 2011.
- [22] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 951–958, 2003.
- [23] J. P. Medhi and S. Dandapat, "An effective fovea detection and automatic assessment of diabetic maculopathy in color fundus images," *Computers in Biology and Medicine*, vol. 74, pp. 30–44, 2016.
- [24] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems (IV)*, P. Heckbert, Ed., Boston, MASS, USA, 1994.
- [25] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems Man & Cybernetics*, vol. 9, pp. 62–66, 2007.
- [26] J. M. Provis, P. L. Penfold, E. E. Cornish, T. M. Sandercoe, and M. C. Madigan, "Anatomy and development of the macula: Specialisation and the vulnerability to macular degeneration," *Clinical and Experimental Optometry*, vol. 88, no. 5, pp. 269–281, 2005.
- [27] C. Kimme, D. Ballard, and J. Sklansky, "Finding Circles by an Array of Accumulators," *Communications of the ACM*, vol. 18, no. 2, pp. 120–122, 1975.
- [28] T. Chen, Y. Luo, F. Xiao, D. Shi, and S. Zhang, "Uneven clustering algorithm based on clustering optimization for wireless sensor networks," *Computer Science*, vol. 41, no. 6A, pp. 289–292, 2014 (Chinese).
- [29] H. K. Hsiao, C. C. Liu, C. Y. Yu, S. W. Kuo, and S. S. Yu, "A novel optic disc detection scheme on retinal images," *Expert Systems with Applications*, vol. 39, pp. 10600–10606, 2012.
- [30] Q. Liu, B. Zou, J. Chen et al., "A location-to-segmentation strategy for automatic exudate segmentation in colour retinal fundus images," *Computerized Medical Imaging and Graphics*, vol. 55, pp. 78–86, 2017.
- [31] R. S. Rekhi, A. Issac, M. K. Dutta, and C. M. Travieso, "Automated classification of exudates from digital fundus images," in *Proceedings of the In International Conference and Workshop on Bioinspired Intelligence*, vol. 16, pp. 1–6, 2017.
- [32] J. Rahebi and F. Hardala, "A new approach to optic disc detection in human retinal images using the firefly algorithm," *Medical Biological Engineering & Computing*, Article ID 453461, pp. 453–461, 2016.
- [33] X. Y. Wang and J. Bu, "A fast and robust image segmentation using FCM with spatial information," *Digital Signal Processing*, vol. 20, 2010.
- [34] S Akara, U. Bunyarit, and B. Sarah, "Automatic exudate detection from non-dilated diabetic retinopathy retinal images using fuzzy c-means clustering," *Sensors*, vol. 9, no. 3, pp. 2148–2161, 2009.
- [35] P. Yogarajah, J. Condell, K. Curran, A. Cheddad, and P. McKevitt, "A dynamic threshold approach for skin segmentation in color images," *International Journal of Biometrics*, vol. 4, pp. 38–55, 2010.
- [36] Y. e. a. Liang, "Hybrid threshold optimization between global image and local regions in image segmentation for melasma severity assessment," *Multidimensional Systems & Signal Processing*, vol. 7, pp. 1–8, 2015.
- [37] T. Jaya, J. Dheeba, and N. A. Singh, "Detection of hard exudates in colour fundus images using fuzzy support vector machine-based expert system," *Journal of Digital Imaging*, vol. 28, no. 6, pp. 761–768, 2015.
- [38] D. Welfer, J. Scharcanski, and D. R. Marinho, "A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images," *Computerized Medical Imaging and Graphics*, vol. 34, no. 3, pp. 228–235, 2010.
- [39] E. Imani and H. R. Pourreza, *A Novel Method for Retinal Exudate Segmentation Using Signal Separation Algorithm*, Elsevier North-Holland, Inc, 2016.
- [40] W. Kusakunniran, Q. Wu, P. Ritthipravit, and J. Zhang, "Hard exudates segmentation based on learned initial seeds and iterative graph cut," *Computer Methods and Programs in Biomedicine*, vol. 158, pp. 173–183, 2018.

Research Article

A Study of Machine-Learning Classifiers for Hypertension Based on Radial Pulse Wave

Zhi-yu Luo,¹ Ji Cui,¹ Xiao-juan Hu ,² Li-ping Tu,¹ Hai-dan Liu,¹ Wen Jiao,¹ Ling-zhi Zeng,¹ Cong-cong Jing,¹ Li-jie Qiao,¹ Xu-xiang Ma,¹ Yu Wang,¹ Jue Wang,¹ Ching-Hsuan Pai,¹ Zhen Qi ,¹ Zhi-feng Zhang ,¹ and Jia-tuo Xu ¹

¹Department of Basic Medical College, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Pudong New Area, Shanghai 201203, China

²Shanghai Collaborative Innovation Center of Health Service in Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Shanghai 201203, China

Correspondence should be addressed to Zhi-feng Zhang; rchbt@163.com and Jia-tuo Xu; xjt@fudan.edu.cn

Received 1 July 2018; Revised 5 October 2018; Accepted 28 October 2018; Published 11 November 2018

Guest Editor: Shadnaz Asgari

Copyright © 2018 Zhi-yu Luo 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.

Objective. In this study, machine learning was utilized to classify and predict pulse wave of hypertensive group and healthy group and assess the risk of hypertension by observing the dynamic change of the pulse wave and provide an objective reference for clinical application of pulse diagnosis in traditional Chinese medicine (TCM). **Method.** The basic information from 450 hypertensive cases and 479 healthy cases was collected by self-developed H20 questionnaires and pulse wave information was acquired by self-developed pulse diagnostic instrument (PDA-1). H20 questionnaires and pulse wave information were used as input variables to obtain different machine learning classification models of hypertension. This method was aimed at analyzing the influence of pulse wave on the accuracy and stability of machine learning model, as well as the feature contribution of hypertension model after removing noise by K-means. **Result.** Compared with the classification results before removing noise, the accuracy and the area under the curve (AUC) had been improved. The accuracy rates of AdaBoost, Gradient Boosting, and Random Forest (RF) were 86.41%, 86.41%, and 85.33%, respectively. AUC were 0.86, 0.86, and 0.85, respectively. The maximum accuracy of SVM increased from 79.57% to 83.15%, and the AUC stability increased from 0.79 to 0.83. In addition, the features of importance on traditional statistics and machine learning were consistent. After removing noise, the features with large changes were h1/t1, w1/t, t, w2, h2, t1, and t5 in AdaBoost and Gradient Boosting (top10). The common variables for machine learning and traditional statistics were h1/t1, h5, t, Ad, BMI, and t2. **Conclusion.** Pulse wave-based diagnostic method of hypertension has significant value in reference. In view of the feasibility of digital-pulse-wave diagnosis and dynamically evaluating hypertension, it provides the research direction and foundation for Chinese medicine in the dynamic evaluation of modern disease diagnosis and curative effect.

1. Introduction

Hypertension is a clinical syndrome whose principal characteristic is an increase in systemic arterial pressure and it is the one of the most common cardiovascular diseases in the world [1–3]. According to 2017 Chinese guidelines for the management of hypertension [4], the prevalence of hypertension has been increasing in China for decades, reaching 23.2%, which has greatly affected the health status of people. In addition, hypertension and other cardiovascular diseases, whose prevention cannot be ignored, possess the

characteristics of high incidence, high mortality, and heavy medical burden.

At present, the data on the classification and prediction of hypertension mainly come from inpatient electronic medical records, environmental and genetic factors, and gene expression data [5]. Zhiyong Pei [6] used support vector machine (SVM) to classify and predict by inputting environmental factors, genetic factors variables, and environmental and genetic factors for 559 hypertensive patients and 641 healthy people. It was found that the combination of environmental variables and genetic factors would improve the prediction

accuracy, reaching 80.1%. Sherif Sakr [7] used cardiorespiratory fitness data and combined the different machine learning methods with the SMOTE algorithm to achieve the higher prediction accuracy of hypertension at 81.69% with AUC of 0.93. In short, there are some achievements on studies in the classification and prediction of hypertension, but these researches have some limitations to some extent. On one hand, they focus on the prediction of a certain stage of data and the preconditions are relatively numerous. On the other hand, the subjectivity of the data is relatively strong. Most of all, data features rarely involve the pulse wave-based hypertension prediction studies.

As one of the methods of traditional Chinese medicine (TCM) diagnosis, pulse diagnosis has always played an important role in clinical diagnosis and treatment and sub-health recuperation [8]. Since the 1950s, the integration of biomedicine, mathematics, physics, biomechanics, bioengineering, computer science, and traditional Chinese medicine has made great progress in the objective study of pulse diagnosis. The pulse wave reflects the shape of the pulse beats and pulse graph is used as the objective image for recording the pulse wave. By combining with the medical history, lab-test diagnosis, and four-combined Chinese medicine diagnosis, and analyzing the correlation of pulse graph with the disease and syndrome, the study gradually establishes the clinical standard diagnosis of pulse graph, which further provides the objective diagnostic method for clinical practice in Chinese medicine. Pearson's correlation coefficient and t-test are often applied for analysis of time domain or frequency domain features, either of which is the classical analysis method in pulse wave [9]. In addition, Hu [10] used KNN to undertake the classification prediction of the pulse wave of the elderly people with different blood pressure segments, but in this article she did not conduct the modelling prediction for multiple age groups. The age of the samples is mostly elderly, and the model has certain limitations. Zhang [11] used genetic algorithm to screen features, and then SVM predicted the time domain information of pulse wave, which obtained accurate rate and ROC of 76.0% and 0.83, respectively. Many studies have suggested that pulse waves have strong correlation with hypertension [12–15].

The research of pulse wave digital processing also has obtained better results and has been applied to the objectification of pulse diagnosis in TCM [16–18]. Based on this, some machine learning methods such as support vector machine, neural network, and random forest [19] have been used in pulse wave recognition or classification. Although some achievements have been made in pulse wave recognition and classification, there are still some problems, such as poor prediction accuracy and model stability, and fuzzy characteristic contribution in the hypertension model. The reason for this may be some deficiencies in model building and the impact of data noise.

Clustering analysis [20] is a classical method of unsupervised learning, and its most representative algorithm is K-means. We attempt to classify different samples into different groups based on K-means clustering analysis method and sort out the contribution of clustering according to the

sample characteristics, and then make further analysis after removing the low-quality samples. Additionally, cardiovascular monitoring commonly relies on sphygmomanometers and less on the risk prediction of chronic cardiovascular disease. The assessment of cardiovascular dynamic risk is more comprehensive by combining pulse wave and H20 scale [8, 19, 21] with symptom assessment.

The main purposes of this paper are as follows: (1) to remove the noise of pulse wave by K-means, so as to further improve the accuracy and stability of the model; (2) to identify the feature variable which is of the highest contribution to hypertension.

2. Materials and Methods

2.1. Subjects. A total of 929 subjects were collected from the outpatient department and the medical examination center of Shanghai Shuguang Hospital attached to Shanghai Chinese Medicine University affiliated to Shanghai University of Traditional Chinese Medicine. Among them, 450 (356 males and 94 females, average age: 44.73 ± 8.73) were diagnosed with hypertension and 479 (337 males and 142 females, average age: 44.49 ± 9.18) with no hypertension.

2.2. Inclusion and Exclusion Criteria. Inclusion criteria for hypertensive group were as follows: patients who met the diagnosis standards of hypertension [4] with the age ranging from 18 to 70, male or female.

Inclusion criteria for healthy group were as follows: (1) the examination results of B-ultrasound, electrocardiogram, biochemistry, imaging, and other subjects in physical examination indicate no disease; (2) the age ranged from 20 to 70, and gender is not limited; (3) H20 score is greater than 80 with no positive items; (4) systolic blood pressure of 90-140mmHg and diastolic pressure 60-90mmHg.

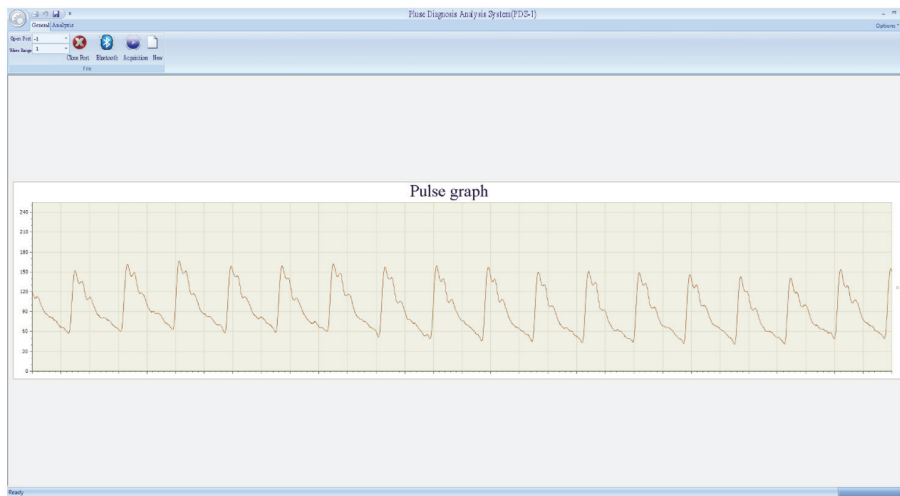
Exclusion criteria include (1) those diagnosed with other serious medical conditions, such as cardiovascular and cerebrovascular diseases, urinary tract diseases, tumors, and immune and hematological diseases; (2) those who suffered from mental illness; (3) those who refused to sign the informed consent; (4) those whose later screening data are not qualified.

2.3. Pulse Image Collection and Analysis Methods

2.3.1. Collection Instrument. In this study, a pulse diagnosis instrument (PDA-1) (Figure 1(a)) was developed by "Traditional Chinese Medicine Diagnosis Information Intelligent Processing Research Team" in Shanghai University of Traditional Chinese Medicine (SHUTCM). The device consists of a pulse wave transducer, an A/D analog-to-digital transducer, and a computer. The main technical parameters are sensitivity (0.5mV/gram force; linear range: 0-250 grams force; output impedance: 1K special symbol), AC amplifier loop (input dynamic range: 0-25mV; full-scale output $\pm 5V$), and DC amplifier loop (input dynamic range: 0-125mV; full-scale output $\pm 5V$).



(a)



(b)

FIGURE 1: Collection instrument. (a) Pulse diagnosis instrument (PDA-1). (b) Interface of pulse diagnosis and analysis system.

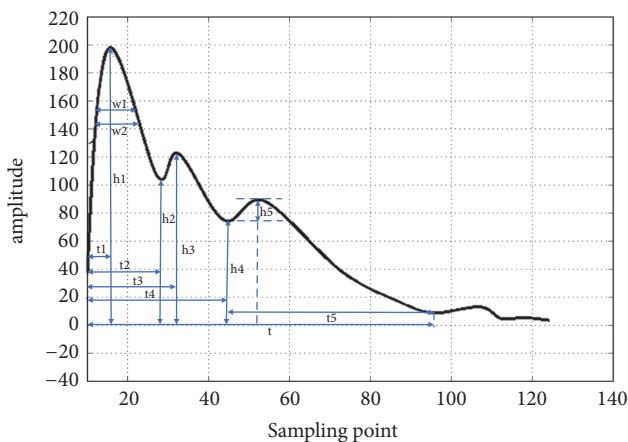


FIGURE 2: The main measurement parameters of pulse wave cycle.

2.3.2. *Collection Methods.* The pulse diagnosis instrument (PDA-1) was used to collect the pulse wave of radial artery on the left hand and the patients were asked to sit still or rest for at least 5 minutes before the acquisition. During the collecting process, patients were required to sit or stay supine, relax, not

talk, and breathe normally. If the above requirements were not satisfied, second collection was needed.

In order to obtain the pulse wave parameters, the Intelligent Information Processing Laboratory of Chinese Medicine Diagnostics of Shanghai University of TCM has developed the method into a pulse diagnosis and analysis system (PDAS) (Figure 1(b)). The system menu has “General” and “Analysis.” “General” includes “Open Port,” “New,” and “Acquisition.” “Analysis” includes “Generate Report,” “Calculate Characteristic Value,” “Export H Value,” and “Batch Analysis.” The data view window shows pulse wave information. The section at the lower right shows the pulse wave of the patient with hypertension (Figure 1(b)). The software outputs the information parameters of pulse for subsequent analysis. Time domain features of the pulse wave [10, 22, 23] including 6 duration features (t , t_1 , t_2 , t_3 , t_4 , and t_5), 5 amplitude features (h_1 , h_2 , h_3 , h_4 , h_5 , h_1/t_1 , h_3/h_1 , and h_4/h_1), 4 width features (w_1 , w_2 , w_1/t , and w_2/t), and 2 area features (A_s , A_d) were extracted by Shannon energy envelope and Hilbert transform [24]. The meaning of the features is listed in Table 1 and Figure 2.

2.4. *Study Design and Setting.* Features such as age, BMI, pulse wave parameters, and H20 score of the hypertension

TABLE 1: TD features.

No.	Features	Meaning
1	h1	Main wave amplitude. It reflects the compliance of the aorta and the cardiac ejection function of the left ventricular
2	h2	Main isthmus wave amplitude. Same physiological significance as h3.
3	h3	Heavy wave front wave amplitude. It reflects the elasticity of arterial vessels and its peripheral resistance.
4	h4	Dicrotic notch amplitude. It reflects the peripheral resistance of arterial vessels and the closure of aortic valve.
5	h5	Gravity wave amplitude. It reflects the compliance of the aorta and the function of aortic valve.
6	t1	Left ventricular rapid ejection period. The time value from the start point to the crest point of the main wave on the pulse graph.
7	t2	The duration of the beginning of the tidal wave.
8	t3	The duration of the crest of the tidal wave.
9	t4	Left ventricular systolic duration. The time value from the start point to the dicrotic notch on the pulse graph.
10	t5	Left ventricular diastolic duration. The time value from the dicrotic notch to the end point on the pulse graph.
11	t	Includes left ventricular systolic and diastolic duration. The time value from the start point to the end point on the pulse graph.
12	w1	main wave 1/3 height. The duration of maintaining high intravascular pressure.
13	w2	main wave 1/5 height. The duration of maintaining high intravascular pressure.
14	w1/t	The ratio of the width of the main wave at its 1/3 height to the entire pulse cycle. It reflects the proportion of the duration time of continuous high pressure in the aorta in the entire pulse cycle.
15	w2/t	The proportion of the duration time of continuous high pressure in the aorta in the entire pulse cycle.
16	h1/t1	cardiovascular function
17	As	Systolic area. The area on the pulse graph is related to cardiac output.
18	Ad	Diastolic area.

and healthy groups were inputted as independent variables in the pretreatment. Subjects were categorized according to whether they have hypertension (dependent variable). 60% of the samples were for training examples and 40% were for testing examples. See flow chart (Figure 3).

2.4.1. Noise Reduction. The K-means algorithm is a partition-based clustering algorithm. A dataset is divided into several groups or classes. Data with higher similarity is in the same group while dissimilar data is in different group. Cluster analysis can help to find abnormal data because similarity and dissimilarity are based on the attribute of data. Similar or neighboring data are aggregated to form each cluster set, and those data, outside these cluster sets, are to be excluded. During the process of data collection, human factors interfered with the data inevitably. Clustering analysis is used to cluster the input feature variables to obtain different levels of clustering results. In a word, cluster analysis is clustered based on different feature variables, and similar samples are gathered into one group.

Firstly the K-means algorithm determines the reference value k and then divides the N data again in the k clusters, so that clusters with a similar degree in each cluster are classified into one and clusters with a low degree of similarity are classified into another. Specific steps: First, any number of k

data in the dataset were found. The original centroid of each cluster was represented by these data. Second, the remaining datasets were divided into each cluster according to the minimization principle, which was based on the distance between each dataset and its cluster centroid. Finally, the centroid of each cluster was calculated again. The above operation was repeated and the calculation was stopped when the value of the objective function was minimum [25].

2.4.2. Feature Normalization. Owing to the differences in the magnitude of the parameters, it has a negative effect on the classification and prediction. The range is scaled and mapped from 0 to 1 (or -1 to 1 if there are negative values).

The formula of MinMaxScaler is

$$\frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where $\min(x)$ is the minimum value, $\max(x)$ is the maximum value, and x_i is the value for each feature.

2.4.3. Classifier. Random Forest (RF), Support Vector Machine (SVM), AdaBoost, Gradient Boosting (GBT), and K-Neighbor (KNN) are classical machine learning models. SVM, AdaBoost, and RF are widely used as classification

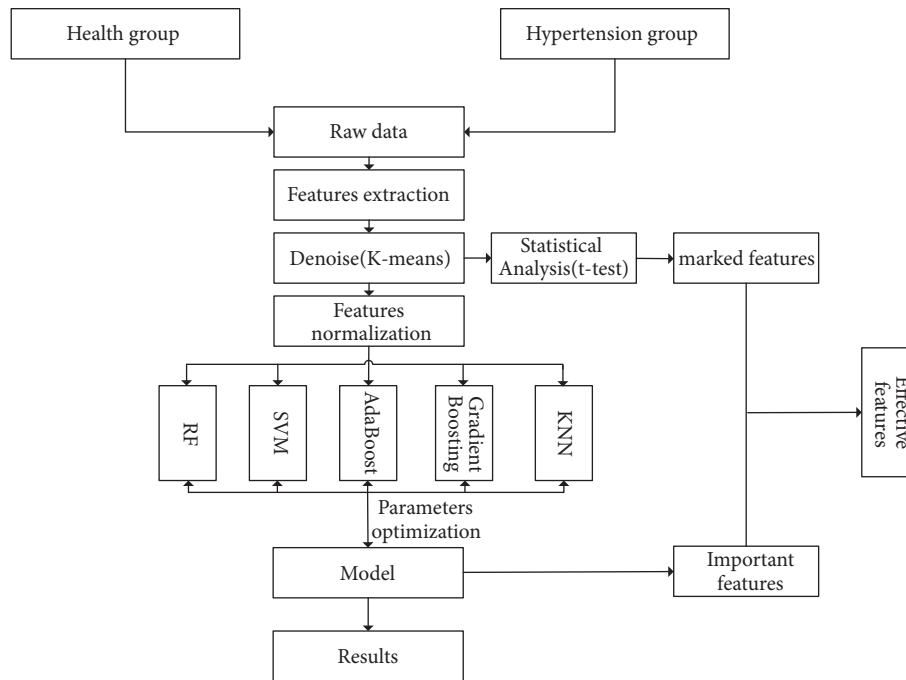


FIGURE 3: Flow chart.

predictions. Support vector machine is a sparse kernel machine, which is a model that only relies on data subset (support vector) to predict unknown class labels [26, 27]. Based on the support vector machine theory, it is pointed out that, for a nonlinear separable dataset containing two classes of points, there are many hyper planes used to classify the classes. And a common radial basis function was chosen. SVM is used to classify training samples to separate and optimize two classes of hyper planes (i.e., decision boundaries). The optimal decision boundary between support vectors is chosen by the distance of the maximum boundary M [28]. Support vector machines have good generalization capabilities. In other words, the decision surface is seen as linear in the high-dimensional space, while it is considered as nonlinear in low-dimensional feature space, which means that SVM could be applied to nonlinear separation data. In addition, in terms of overfitting problems, support vector machine is of robustness for high-dimensional data [29]. The main drawback is that it is more difficult to interpret the generated model and has a certain sensitivity to appropriate parameter adjustments.

K-nearest neighbor (KNN) is one of the easiest methods to predict classification in pattern recognition [29]. To obtain the nearest neighbor for each dataset, KNN uses measurement to calculate the distance between data pairs. In general, the measurement used is Euclidean distance. Since each new data point is classified differently, KNN can establish a local approximation of the objective function [30]. While a test example is classified, it will use a similarity function based on the Euclidean distance to find training examples of the K most recent query points [29]. Since high k results in overfitting and model instability, the appropriate values must be specifically chosen [28]. Another advantage of KNN is its

simplicity. In spite of this, the forecasting time is usually very expensive because all the training data must be reexamined [30].

Random Forest (RF) uses a majority vote to predict categories based on data partitions from multiple decision trees [28]. In each decision tree, data points fall into specific leaves according to their characteristics and are assigned a forecast. Then the data points are averaged. The maximum number of voting categories will provide the final forecast [26]. The Gini index is used to determine the “best split” threshold for a given category of input values. Compared with the parent node, the Gini index returns the measure of the heterogeneity of the child nodes [28].

AdaBoost is a supervised learning algorithm for solving classification problems [31]. In each sequence, misclassified instances are given more weight for the next sequence while correctly classified instances are given lower weight. The final model is a linear combination of all the models created in the previous sequence [32]. In addition, GBDT has very few limitations and assumptions on the input data, so it is very flexible to deal with complex nonlinear relationships [33]. In some problems, it is more stable than other learning algorithms and is less susceptible to be affected by overfitting problems. Each learning algorithm has its advantages and tends to be more suitable for certain types of problems than other types of problems, and there are usually many different parameters and configurations that need to be adjusted before achieving the best performance of the dataset [34]. Gradient Boosting could strategically combine some simple tree models to obtain optimized predictive performance while the result model could be interpreted by identifying key variables [33]. The core is that the learning objective of each subtree is the residual of the previous subtree. The sum of all

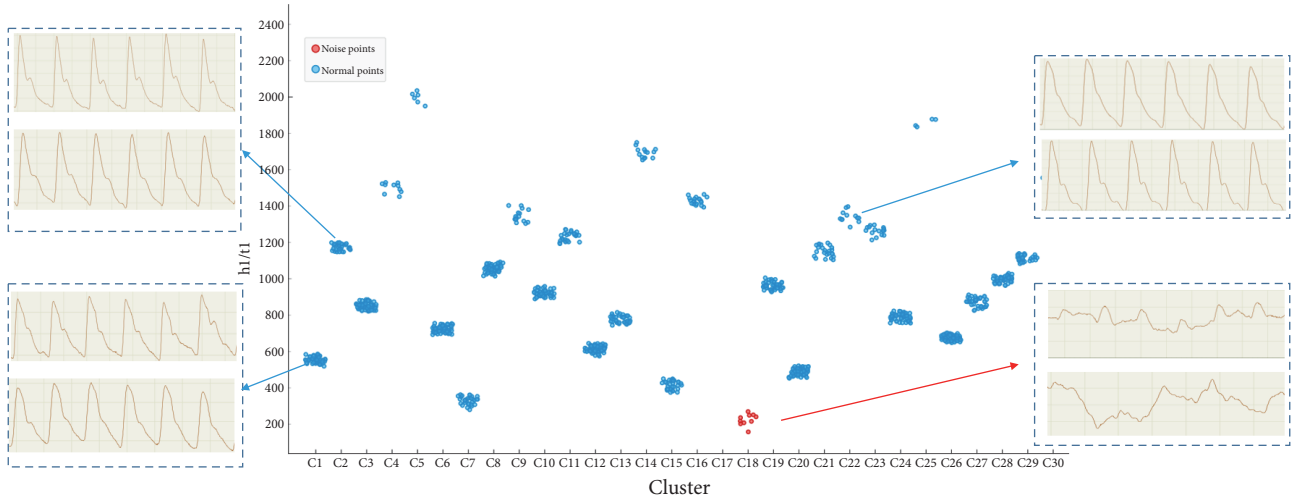


FIGURE 4: k-means clustering analysis. The x-axis represents the number of clusters while the y-axis represents the variable h1/t1. Red points include noise pulse wave. Blue points include normal pulse wave.

subtrees can be used as the final result of the model. GBT can handle different types of predictors and missing data. At the same time, it does not need to eliminate outliers and perform previous data transformations [35].

SVM is used mainly to establish a classification hyper plane as the decision surface and maximize the isolated edges of the positive and negative instances and then construct and solve the optimization problem by selecting appropriate kernel functions and appropriate penalty parameters. AdaBoost is an iterative algorithm, whose core objective is to train different classifiers for the same training set (i.e., weak classifiers) and then combine weak classifiers into a stronger classifier. Besides, K-Neighbors uses the distance calculation method. According to the new data calculated by all features and categorical distance of data point in the dataset, it will operate classification prediction by sorting them in ascending order of distance.

2.4.4. Parameters Optimization and Evaluation Criteria. Due to the different performances of different models, different classification models of hypertension were constructed, respectively. At the same time, different machine learning was performed by grid search and 10-fold cross-validation to optimize the parameters. The optimal parameters were selected to establish the model.

In order to assess the feasibility of the above methods, analysis was performed using common evaluation criteria [36], including accuracy rate (ACC), area under the ROC curve (AUC), sensitivity (ST), and specificity (SP). AUC [37], an evaluation binary model, is one of the popular methods. AUC is used in the range from 0 to 1. Moreover, the four basic statistical definitions which describe the process of classification are TP (true positive, number of positives), FP (false positive, number of negatives), TN (true negative, number of negatives), and FN (false negative, number of positives).

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$Specificity = \frac{TN}{FP + TN} \times 100\% \quad (4)$$

2.5. Development Platform. Statistical analysis was performed using SPSS 22.0 software. The BMI, age, and pulse wave parameters were analyzed in two groups by an independent samples t-test. The data were shown as mean and standard deviation. $P < 0.05$ indicates a statistical difference.

The data was collected and analyzed by python3.5 and sklearn [38] to achieve machine learning. Orange3.11 [39] was used for cluster analysis and removing unqualified pulse wave.

3. Result

3.1. Noise Reduction. Studies [40, 41] have shown that h1/t1 can reflect the ability of cardiac ejection and aortic compliance; then elasticity and compliance of vascular directly affect blood pressure. According to the results of the k-means in Figure 4, the points in the red circle below could be considered as the group with the most noise, which indicates that the pulse wave of hypertensive population was better for noise recognition. After clustering, it was found that one group contains more noise and the other group contains less noise. Pulse waves whose noise was removed are classified into one group and those whose noise was not removed are in the other group to conduct the comparative analysis. h1/t1 had greater difference in different pulse waves. The pulse wave with greater similarity was gathered into one group, whereas the pulse wave with the large noise was collected into another group in the red circle shown in Figure 4.

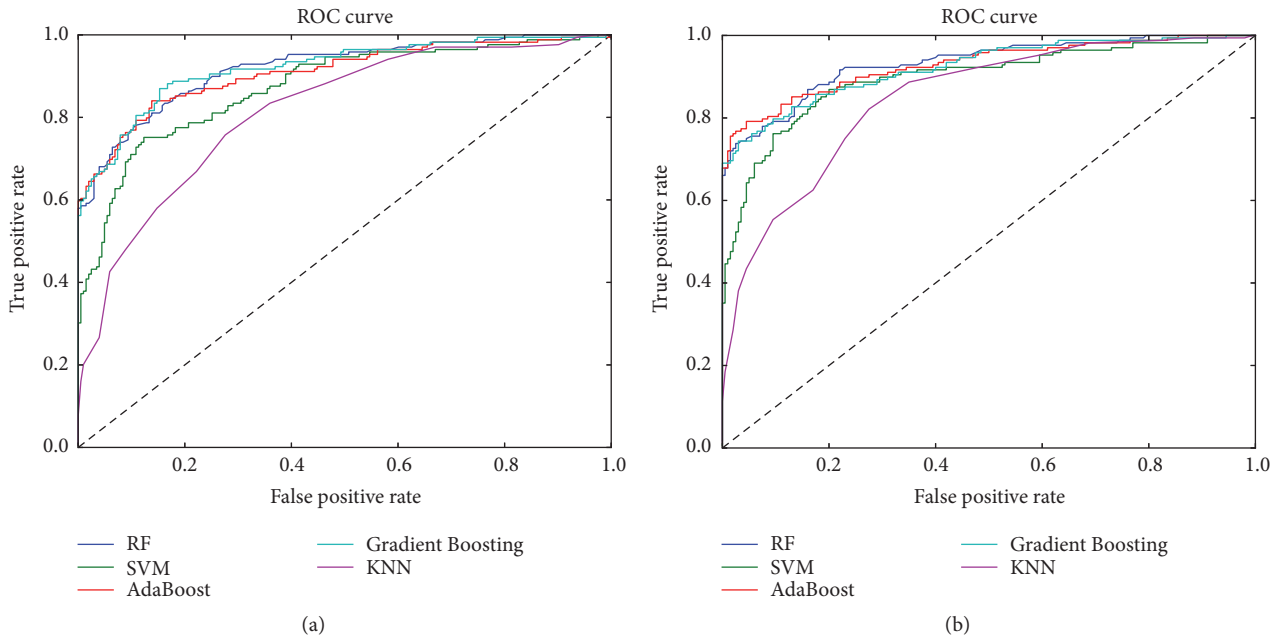


FIGURE 5: The ROC curve in different model. (a) Data without noise reduction. (b) Data with noise reduction. The x-axis denotes false positive rate. The y-axis is true positive rate. In Figure 5, the dark blue line represents Random Forest (RF). The green line represents support vector machine (SVM). The red line represents Adaboost. The light blue line represents Gradient Boosting. The purple line represents K-nearest neighbor (KNN).

3.2. *Between Hypertension and Healthy Group of the Pulse Graph Characteristics.* The results of pulse wave parameter between the healthy group and the hypertension group after noise reduction are shown in Table 2. Compared with the healthy group, the hypertension group has higher BMI, h1, h2, h3, t2, t5, h1/t1, w1/t, and HR and lower h5, t, w2/t, H20 score, and Ad ($P < 0.05$).

3.3. *Five Data Mining Algorithms Classification.* The study used RF, SVM, AdaBoost, Gradient Boosting, and K-Neighbors to establish a hypertension identification model based on pulse wave features and results were shown in Table 3. Through the comparison of different machine learning algorithms, it could be seen that the accuracy of the four kinds of machine learning models had been improved in comparison to the result of the pulse wave analysis after noise removal, and SVM had the largest increase. Although the accuracy rate of K-Neighbors had also greatly improved, the accuracy rate of prediction was the lowest. At the same time, in terms of classifier performance, the biggest increase in AUC was SVM and AdaBoost. By observing the accuracy and stability, AdaBoost, Gradient Boosting, and RF had better result, and K-Neighbors classifier had unsatisfactory result.

The ROC curve is a graph that describes the performance of the binary classifier system. In other words, the ROC curve is plotted based on true positive rate and false positive rate. Sensitivity is also known as TPR, which means that the possibility of high blood pressure is truly judged. Specificity is equal to the true negative rate, which means that there is no possibility of disease. The area under the ROC curve is

most commonly used as an accurate index. If the sensitivity and specificity reach 1, the area under the ROC curve reaches the desired accuracy. The best prediction method generates a point in the upper left corner (0, 1) of the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). In this study, sensitivity, specificity, and ROC (AUC) are used to evaluate the classifier performance. As shown in Figure 5, the ROC curve in the figure is a classifier result using a noise reduction and no noise reduction dataset, respectively. Different colored lines represented the ROC curve of different machine learning models. Among them, AdaBoost and Gradient Boosting had the most significant changes in the noise reduction and non-noise reduction curves, which indicates that the AdaBoost and Gradient Boosting classifiers have higher sensitivity and specificity after noise reduction.

3.4. *Feature Importance.* The variables output by the machine learning model were compared and analyzed to obtain the degree of contribution of the model. The results on the three models are shown in Figure 6.

Compared with the results of classification after reducing noise, the features of the hypertensive classification model varied greatly among different machine learning. AdaBoost and Gradient Boosting had the most significant changes, and RF had the smallest. After removing the noise in the AdaBoost and Gradient Boosting models, the importance of h1/t1, w1/t, t, w2, h2, t1, and t5 variables (top10) had increased. Among them, AdaBoost had the most prominent changes in w2, w1/t, t, h1/t1, h2, t5, and Ad. In RF, only BMI, t5, h2, and h1/t1 variables had increased.

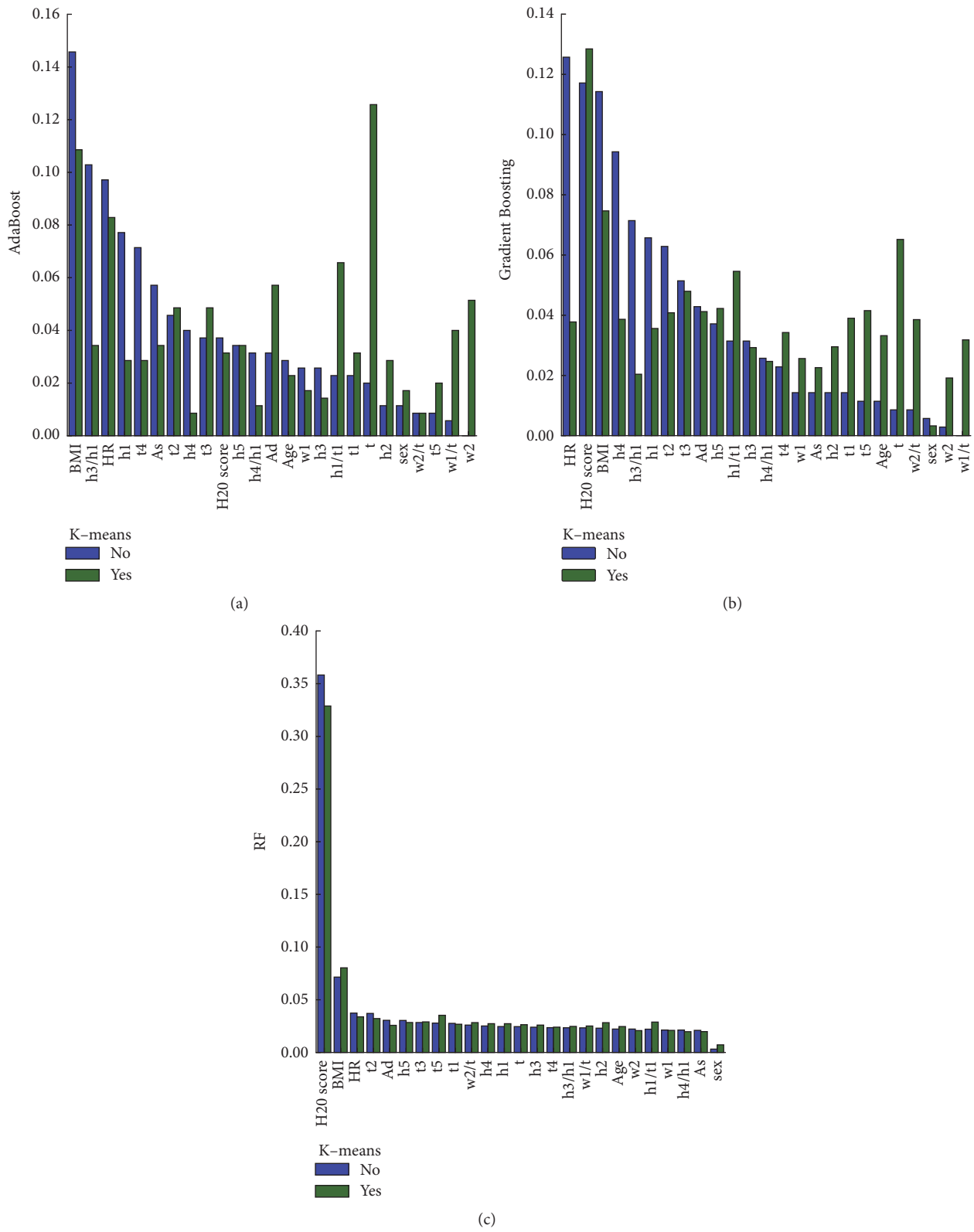


FIGURE 6: The importance of the classification variables features. The bar charts (a), (b), and (c) represent the results of feature importance of AdaBoost, Gradient Boosting, and Random Forest (RF), respectively. The y-axis represents the value of the importance for variables features. Note that “yes” represents that K-means is used to reduce noise in this model and vice versa.

TABLE 2: Comparison of the variables between hypertension group and healthy group ($\bar{X} \pm S$).

Feature	Healthy group	hypertension group	p value
Age	44.44 ± 9.204	44.7 ± 8.706	0.37
BMI	23.91 ± 2.961	25.55 ± 3.306	0.0 * *
w1	0.18 ± 0.035	0.18 ± 0.033	0.55
w2	0.13 ± 0.034	0.14 ± 0.034	0.05
As	0.22 ± 0.029	0.22 ± 0.028	1
Ad	0.11 ± 0.035	0.1 ± 0.036	0.0 * *
h1	116.0 ± 35.992	126.11 ± 44.893	0.0 * *
h2	84.98 ± 32.215	93.18 ± 40.319	0.02*
h3	78.02 ± 29.356	85.73 ± 36.019	0.01 * *
h4	44.78 ± 15.047	47.55 ± 18.016	0.09
h5	12.8 ± 4.455	12.16 ± 4.056	0.04*
t1	0.14 ± 0.021	0.14 ± 0.022	0.31
t2	0.24 ± 0.037	0.24 ± 0.041	0.0 * *
t3	0.27 ± 0.033	0.27 ± 0.038	0.08
t4	0.36 ± 0.03	0.36 ± 0.034	0.23
t5	0.41 ± 0.023	0.41 ± 0.028	0.0 * *
t	0.85 ± 0.119	0.83 ± 0.128	0.01 * *
h1/t1	838.82 ± 276.686	919.03 ± 355.812	0.0 * *
h3/h1	0.67 ± 0.128	0.68 ± 0.133	0.36
h4/h1	0.39 ± 0.082	0.38 ± 0.082	0.11
w1/t	0.21 ± 0.036	0.22 ± 0.033	0.0 * *
w2/t	0.16 ± 0.036	0.16 ± 0.035	0.0 * *
HR	77.08 ± 9.613	80.4 ± 11.839	0.0 * *
H20 score	85.74 ± 4.868	76.38 ± 10.331	0.0 * *

Compared with healthy group. * P <0.05, ** P <0.01.

TABLE 3: Results on the classification of machine learning model.

Model	ACC	AUC	SP	ST
RF	0.841	0.832	0.936	0.728
RF*	0.853	0.848	0.905	0.792
Gradient Boosting	0.852	0.843	0.941	0.746
Gradient Boosting*	0.864	0.859	0.920	0.798
SVM	0.796	0.792	0.833	0.752
SVM*	0.832	0.828	0.865	0.792
AdaBoost	0.839	0.830	0.921	0.740
AdaBoost*	0.864	0.858	0.925	0.792
KNeighbors	0.729	0.716	0.852	0.580
KNeighbors*	0.736	0.728	0.830	0.625

Models with an asterisk * mean that K-means is applied before using these models.

Furthermore, compared with the features of significant difference in traditional statistics, machine learning AdaBoost and Gradient Boosting had significant difference in the common feature rankings. Therefore, AdaBoost and Gradient Boosting were selected to analyze the important features. The variables of the top 12 are listed in Table 4. In Figure 7, the common variables among the three are h5, t, Ad, BMI, h1/t1, and t2, which indicates that hypertension may play an important role in cardiac output.

4. Discussion

Previous studies showed [12–15] that hypertension and pulse waves had a strong correlation, and statistical description of this study also confirmed this phenomenon. The comparative results of the pulse wave characteristics showed that the values such as h1/t1, h1, h3, and w1/t were higher than those in the healthy group. Therefore, the differences of characteristics in pulse wave between hypertensive and healthy group made it possible to further classify them using machine learning.

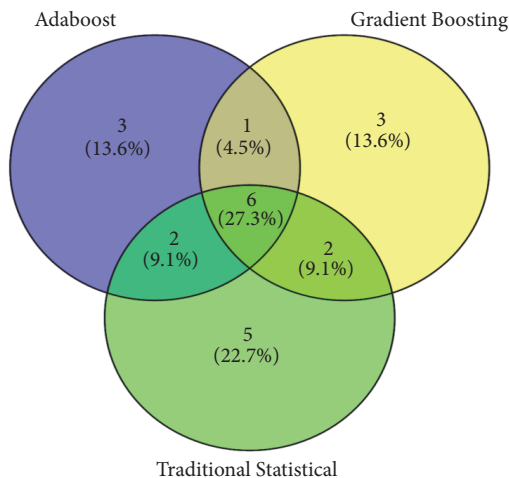


FIGURE 7: The analysis of the importance on the features of traditional statistics and machine learning.

TABLE 4: Selected (top 12) features in model.

model	Feature variable	Feature importance
Adaboost	t	0.126
	BMI	0.109
	HR	0.083
	h1/t1	0.066
	Ad	0.057
	w2	0.051
	t2	0.049
	t3	0.049
	w1/t	0.040
	h5	0.034
	h3/h1	0.034
	As	0.034
Gradient Boosting	H20 score	0.128
	BMI	0.075
	t	0.065
	h1/t1	0.055
	t3	0.048
	h5	0.042
	t5	0.042
	Ad	0.041
	t2	0.041
	t1	0.039
	h4	0.039
	w2/t	0.039

However, in terms of the analysis of raw data, filtering pulse wave lacks the methods of quality control. At present, there are human interference factors in the acquisition process of pulse wave, so it is necessary to reduce the noise. Taking advantage of pulse waves after noise reduction as the input variables of the machine learning, the study suggests that the machine learning accuracy and stability have significantly improved.

Figure 4 shows clustering analysis of data. Different sample clustering had found a certain degree of regularity. Pulse

waves with higher heat values were grouped together (typical hypertension pulse wave); pulse waves within middle-ranged heat values were collected together (mixed part of hypertension pulse wave and healthy pulse wave); pulse waves with low heat values were gathered together. Further observation revealed that the pulse wave with a lower heat value was mixed with more noise. This part of the noise was in line with the pulse wave neither in hypertensive group nor in healthy group. This noise primarily derives from respiration and muscle tension. On the one hand, the respiration can lead to abnormalities in the pulse wave. On the other hand, subjects who have the high muscle tension are likely to cause tremor of the pulse wave. Therefore, the elimination of such pulse wave is necessary.

Table 3 shows the classification results of pulse waves with noise reduction. The classifications were obtained by the RF, SVM, AdaBoost, Gradient Boosting, and K-Neighbors algorithms, respectively. Meanwhile, 10-fold cross-validation and grid optimization were performed to measure the classification performance. Evaluation indicators pointed out that, compared with the results of pulse wave classification without noise reduction, AdaBoost and Gradient Boosting had the better classification effect after noise reduction, and SVM had the larger increase.

Significant features obtained from traditional statistical analysis are specific, such as h1, h2, h3, t2, t5, h1/t1, w1/t, h5, t, w2/t, Ad, H20 score, and BMI, but traditional statistical analysis is weak in linking nonlinear relationships. Machine learning methods, however, have advantages in this regard. Through traditional analysis and machine learning analysis, the prediction results are relatively satisfactory, but machine learning cannot see the specific operation mechanism of "black box." With the development of technology over the years, some machine learnings become valuable reference on practical use. The results showed that there were significant differences in the importance ranking on features among three different machine learnings before and after noise reduction. Among them, AdaBoost and Gradient Boosting had the better changes in feature importance. The importance of variables features in t, t1, h2, h1/t1, t5, w1/t, and w2 had ascended to some extent. Therefore, the result suggests that the wave of wiry pulse is more obvious after noise reduction and this is consistent with the theory of traditional Chinese medicine [10].

The application of the combination of TCM diagnostic technology and modern information technology is promising. The development of the wearable Chinese medicine pulse wristband represented by the pulse modernization research is based on the traditional Chinese medicine theories combined with modern information technology, artificial intelligence, and other technologies. In addition, it also retains the characteristics of pulse diagnosis in TCM, including miniaturization, wearable, wireless transmission, and intelligence, and it can be widely used in Chinese medicine teaching, scientific research, medical care, health care, and many other fields, with broad domestic and international market prospects. Currently, there is a wearable health assessment technology based on pulse evaluation, which provides a new method for further disease prevention and evaluation [40]. The effective

development of wearable Chinese medicine pulse diagnosis bracelets and the transformation into products will greatly promote the development and application of modern Chinese medical diagnosis and treatment technologies and bring many positive social and economic benefits. On one hand, portable pulse acquisition provides objective results and data indicators for clinical diagnosis and therapeutic evaluation. On the other hand, it provides technical basis for the modern clinical research of TCM with Chinese characteristics and conforms to the connotation of TCM. Finally, pulse diagnosis is a TCM diagnostic technology information. This study lays a foundation for further exploration on wearable pulse diagnostic equipment.

5. Conclusion

Based on the pulse wave, this paper used cluster analysis to eliminate noise and machine learning to establish a classification model for hypertension. It shows good classification effect and indicates that removing noise has great significance in improving accuracy and stability of model. It also illustrates that it is feasible to use computer technology to conduct TCM diagnosis. Besides, it is also part of establishing the classification model to identify the factors that affect hypertension diagnosis. The results of traditional analysis and machine learning imply that the variables of h1/t1, h5, t, Ad, BMI, and t2 are likely to connect with hypertension.

Through collecting and analyzing the information of hypertension, this study explores the information and application of traditional Chinese medicine and provides a reference for the design of a more effective classification model of hypertension. In addition, combined with the symptoms and signs of the patients and the information of tongue and pulse diagnosis in Chinese medicine, the development of a more convenient and wearable pulse diagnostic instrument provides a more real-time, convenient, and quick method for further study in the prevention and prediction of hypertension.

Data Availability

The datasets generated and analyzed during the current study are not publicly available due to the confidentiality of the data, which is an important component of the National Key Technology R&D Program of the 13th five-year plan (No. 2017YFC1703301) in China, but are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors report no conflicts of interest in this work.

Authors' Contributions

Zhi-yu Luo, Ji Cui, and Xiao-juan Hu contributed equally to this work. Jia-tuo Xu and Zhi-feng Zhang conceived and designed the research. Zhi-yu Luo supervised the data collection and wrote the manuscript, while Ji Cui, Xiao-juan Hu, and Li-ping Tu assisted with the statistical analysis and

the writing of the manuscript. The rest of the coauthors were responsible for the data collection of the pulse wave. All authors have read and approved the final version of this manuscript.

Acknowledgments

The research was conducted through the National Natural Science Foundation of China (No. 81373556, No. 81102558, and No. 81503552), National Key Technology R&D program (No. 2012BAI37B06, No. 2017YFC1703301), "Shu Guang" Project Supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation (No. 12SG36), and Basic Subject Innovation Fund Supported by Shanghai Municipal Education Commission (No. A2-P1600317).

References

- [1] J. Blacher, B. I. Levy, J.-J. Mourad, M. E. Safar, and G. Bakris, "From epidemiological transition to modern cardiovascular epidemiology: hypertension in the 21st century," *The Lancet*, vol. 388, no. 10043, pp. 530–532, 2016.
- [2] E. Lurbe, L. Thijs, M. I. Torro, J. Alvarez, J. A. Staessen, and J. Redon, "Sexual dimorphism in the transition from masked to sustained hypertension in healthy youths," *Hypertension*, vol. 62, no. 2, pp. 410–414, 2013.
- [3] N. R. Campbell and X. Zhang, "Hypertension in China: Time to Transition From Knowing the Problem to Implementing the Solution," *Circulation*, vol. 137, no. 22, pp. 2357–2359, 2018.
- [4] Primary Expert Committee on Hypertension Management, "The 2017 Chinese guidelines for the management of hypertension," *Chinese Circulation Journal*, vol. 32, pp. 1041–1048, 2017.
- [5] E. Held, J. Cape, and N. Tintle, "Comparing machine learning and logistic regression methods for predicting hypertension using a combination of gene expression and next-generation sequencing data," *BMC Proceedings*, vol. 10, no. 7, pp. 141–145, 2016.
- [6] Z. Pei, J. Liu, M. Liu et al., "Risk-Predicting Model for Incident of Essential Hypertension Based on Environmental and Genetic Factors with Support Vector Machine," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 10, no. 1, pp. 126–130, 2018.
- [7] S. Sakr, R. Elshawi, A. Ahmed et al., "Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford Exercise Testing (FIT) Project," *PLoS ONE*, vol. 13, no. 4, p. e0195344, 2018.
- [8] C. Ji, T. Liping, Z. Jianfeng et al., "Research on pulse graph characteristics of 1720 cases with different health status and age gradient," *Shanghai Journal of Traditional Chinese Medicine*, 2018.
- [9] X. Y. Zheng, J. W. Luo, S. C. Wei et al., "Correlation Research of Renying Pulse Wave Intensity and Pulse Condition Parameters in the Patients of Hypertension," *World Journal of Integrated Traditional & Western Medicine*, 2015.
- [10] X.-J. Hu, L. Zhang, J.-T. Xu et al., "Pulse Wave Cycle Features Analysis of Different Blood Pressure Grades in the Elderly," *Evidence-Based Complementary and Alternative Medicine*, vol. 2018, Article ID 1976041, 12 pages, 2018.
- [11] J. Zhang, *Study on the Key Technology of TCM Health Status Identification Based on Multi-Information Characteristics*,

- Shanghai University of Traditional Chinese Medicine, Shanghai, China, 2017.
- [12] S. Jing and N.-L. Sun, "Curative effects on mild to moderate primary hypertension and influence on pulse wave velocity of slow releasing felodipine," *National Medical Journal of China*, vol. 88, no. 42, pp. 2962–2965, 2008.
 - [13] S. Mitani, M. Fujita, M. Shigeta et al., "Determinants of brachial-ankle pulse wave velocity in a Japanese population: A cohort study," *Blood Pressure*, vol. 21, no. 6, pp. 338–344, 2012.
 - [14] Y. Xue, Y. Su, C. Zhang et al., "Full-field wrist pulse signal acquisition and analysis by 3D Digital Image Correlation," *Optics and Lasers in Engineering*, vol. 98, pp. 76–82, 2017.
 - [15] Y. Zheng, Y. Zhang, Z. Ma et al., "Predicting Arterial Stiffness from radial Pulse Waveform using support vector machines," *Procedia Engineering*, vol. 7, pp. 458–462, 2010.
 - [16] R. Chen, M. Zhang, L. I. He et al., "Objective exploration of pulse diagnosis characteristics about overabundant liver-fire syndrome based on community hypertension," *Global Traditional Chinese Medicine*, 2017.
 - [17] C. Y. Tang, *Developing an objective traditional Chinese medicine pulse diagnostic model in essential hypertension*, Hong Kong Polytechnic University, 2010.
 - [18] W. Xu, H. Yan, J. Xu et al., "Objective study for pulse diagnosis of Traditional Chinese Medicine: Pulse signal analysis of patients with Coronary Heart Disease," in *Proceedings of the 11th IEEE International Conference on Control and Automation, IEEE ICCA 2014*, pp. 1118–1121, Taiwan, June 2014.
 - [19] J. Xu, L. Tu, Y. Bao, Q. Chen, H. Wu, and Z. Zhang, "Pulse graph characteristics analysis and classification of sub health state," in *Proceedings of the 2012 International Symposium on Information Technologies in Medicine and Education, ITME 2012*, pp. 677–681, Japan, August 2012.
 - [20] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithms: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, 2002.
 - [21] G.-X. Chu, Q.-G. Chen, J.-T. Xu et al., "Analysis on pulse diagram characteristics of subjects with subhealth state," *Journal of Chinese Integrative Medicine*, vol. 10, no. 10, pp. 1099–1105, 2012.
 - [22] Z. Zhang, J. Xu, E. Zhu et al., "Effect of pulsatile perfusion during cardiopulmonary bypass in terms of radial artery sphygmogram," *Journal of Traditional Chinese Medicine*, vol. 34, no. 6, pp. 673–677, 2014.
 - [23] L. Tu, J. Xu, Z. Zhang et al., "A Shiatsu pulse sensor calibration method and application to three-part pulse wave collection," in *Proceedings of the 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013*, pp. 254–255, China, December 2013.
 - [24] X. Hu, H. Zhu, J. Xu, D. Xu, and J. Dong, "Wrist pulse signals analysis based on Deep Convolutional Neural Networks," in *Proceedings of the 2014 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2014*, USA, May 2014.
 - [25] L. Jin and G. Yuefei, "A K-means clustering algorithm with noise processing," *Computer Application And Software*, vol. 32, no. 10, pp. 265–268, 2015.
 - [26] A. Decruyenaere, P. Decruyenaere, P. Peeters, F. Vermassen, T. Dhaene, and I. Couckuyt, "Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods," *BMC Medical Informatics and Decision Making*, vol. 15, no. 1, p. 83, 2015.
 - [27] J. Zhang, J. Xu, X. Hu et al., "Diagnostic method of diabetes based on support vector machine and tongue images," *BioMed Research International*, vol. 2017, Article ID 7961494, 9 pages, 2017.
 - [28] M. J. Cracknell and A. M. Reading, "Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information," *Computers & Geosciences*, vol. 63, pp. 22–33, 2014.
 - [29] J. L. M. Amaral, A. J. Lopes, J. M. Jansen, A. C. D. Faria, and P. L. Melo, "Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease," *Computer Methods and Programs in Biomedicine*, vol. 105, no. 3, pp. 183–193, 2012.
 - [30] A. C. Lorena, L. F. O. Jacintho, M. F. Siqueira et al., "Comparing machine learning classifiers in potential distribution modelling," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5268–5275, 2011.
 - [31] F. Çatak, "Classification with boosting of extreme learning machine over arbitrarily partitioned data," *Soft Computing*, vol. 21, no. 9, pp. 2269–2281, 2017.
 - [32] M. Bogaert, M. Ballings, and D. Van den Poel, "Evaluating the importance of different communication types in romantic tie prediction on social media," *Annals of Operations Research*, vol. 263, no. 1–2, pp. 501–527, 2018.
 - [33] C. Ding, D. Wang, X. Ma, and H. Li, "Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees," *Sustainability*, vol. 8, no. 11, 2016.
 - [34] M. Maniruzzaman, M. J. Rahman, M. Al-MehediHasan et al., "Accurate Diabetes Risk Stratification Using Machine Learning: Role of Missing Value and Outliers," *Journal of Medical Systems*, vol. 42, no. 5, 2018.
 - [35] B. Zhang, X. Wan, F. Ouyang et al., "Machine Learning Algorithms for Risk Prediction of Severe Hand-Foot-Mouth Disease in Children," *Scientific Reports*, vol. 7, no. 1, 2017.
 - [36] Y. Pan, D. Liu, and L. Deng, "Accurate prediction of functional effects for variants by combining gradient tree boosting with optimal neighborhood properties," *PLoS ONE*, vol. 12, no. 6, 2017.
 - [37] J. Zhang, J. Xu, X. Hu et al., "Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images," *BioMed Research International*, vol. 2017, Article ID 7961494, 9 pages, 2017.
 - [38] D. Bacciu, A. Carta, S. Gnesi, and L. Semini, "An experience in using machine learning for short-term predictions in smart transportation systems," *Journal of Logical and Algebraic Methods in Programming*, vol. 87, pp. 52–66, 2017.
 - [39] J. Demsar, T. Curk, A. Erjavec et al., "Data Mining Toolbox in Python," *Journal of Machine Learning Research*, vol. 14, pp. 2349–2353, 2013.
 - [40] J. Cui, L. Tu, X. Hu et al., "A Kind of Array Pulse Pattern Analysis Method for Wearable Health Monitoring," *Basic & Clinical Pharmacology & Toxicology*, vol. 10, 2017.
 - [41] D. Wu, C. Z. Yang, Q. Li, H. Jin, and H. Sun, "The study on continuous blood pressure information based on the pulse characteristic parameters," *Applied Mechanics and Materials*, vol. 346, pp. 103–108, 2013.

Research Article

Computer Aided Diagnosis System for Detection of Cancer Cells on Cytological Pleural Effusion Images

Khin Yadanar Win ¹, **Somsak Choomchuay**¹, **Kazuhiko Hamamoto**²,
Manasanan Raveesunthornkiat ³, **Likit Rangsirattanakul**³ and **Suriya Pongsawat**³

¹Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

²School of Information and Telecommunication Engineering, Tokai University, Tokyo, Japan

³Department of Pathology, Faculty of Medicine, Srinakharinwirot University, Nakhon Nayok, Thailand

Correspondence should be addressed to Khin Yadanar Win; kyadanarw@gmail.com

Received 4 July 2018; Revised 11 September 2018; Accepted 16 October 2018; Published 8 November 2018

Guest Editor: Fabien Scalzo

Copyright © 2018 Khin Yadanar Win 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.

Cytological screening plays a vital role in the diagnosis of cancer from the microscope slides of pleural effusion specimens. However, this manual screening method is subjective and time-intensive and it suffers from inter- and intra-observer variations. In this study, we propose a novel Computer Aided Diagnosis (CAD) system for the detection of cancer cells in cytological pleural effusion (CPE) images. Firstly, intensity adjustment and median filtering methods were applied to improve image quality. Cell nuclei were extracted through a hybrid segmentation method based on the fusion of Simple Linear Iterative Clustering (SLIC) superpixels and K-Means clustering. A series of morphological operations were utilized to correct segmented nuclei boundaries and eliminate any false findings. A combination of shape analysis and contour concavity analysis was carried out to detect and split any overlapped nuclei into individual ones. After the cell nuclei were accurately delineated, we extracted 14 morphometric features, 6 colorimetric features, and 181 texture features from each nucleus. The texture features were derived from a combination of color components based first order statistics, gray level cooccurrence matrix and gray level run-length matrix. A novel hybrid feature selection method based on simulated annealing combined with an artificial neural network (SA-ANN) was developed to select the most discriminant and biologically interpretable features. An ensemble classifier of bagged decision trees was utilized as the classification model for differentiating cells into either benign or malignant using the selected features. The experiment was carried out on 125 CPE images containing more than 10500 cells. The proposed method achieved sensitivity of 87.97%, specificity of 99.40%, accuracy of 98.70%, and F-score of 87.79%.

1. Introduction

Pleural effusion or pulmonary effusion (PE) is the pathologic accumulation of fluid in the pleural cavity, between the visceral and parietal layers surrounding the lung, as demonstrated in Figure 1 [1, 2]. Normally, the pleural space is lined by a thin layer of mesothelial cells and contains about 5-10 ml of clear fluid for lubrication during respiratory movement. When cancer cells grow or spread to the pleura, they cause malignant pleural effusion (MPE). Half of all cancer patients have a high possibility of developing MPE. Both primary and metastatic cancers can lead to a diagnosis of MPE. Mesothelioma, a rare form of cancer, is the primary

cancer of the pleura. Lung cancer and breast cancer are the most frequent metastatic cancers in male and female patients, respectively. Both malignancies are responsible for about 50-65% of MPE. Lymphoma, tumors of the genitourinary tract, and gastrointestinal tract are responsible for 25%. Tumors of unknown primary account for 7-15% of all MPE [3]. From statistics, as mentioned earlier, MPE is mostly caused by the invasion of metastatic cancer to the pleura. Metastatic cancer is the major cause of cancer morbidity and mortality. It is estimated that metastasis is responsible for about 90% of cancer deaths. Although cancer in the pleural effusion is seen in advanced stages of malignancy and leads to rapid mortality, the survival time can be prolonged by earlier diagnosis

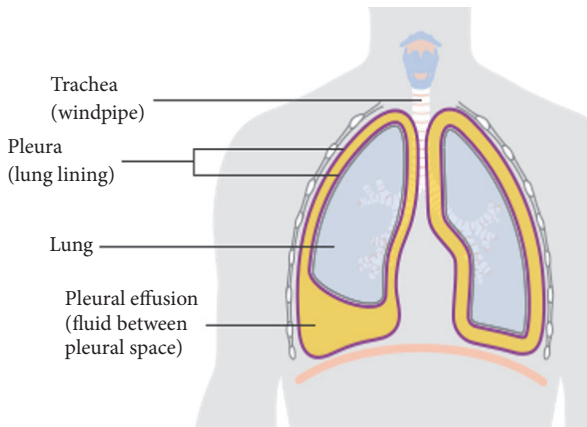


FIGURE 1: The presence of pleural effusion in the pleural cavity [2].

together with prompt and effective treatment to slow cancer progress. Currently available tools for detecting the presence of MPE in the pleura are cytology, cytometry, and imaging modalities such as X-ray, Ultrasound, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI). For the assessment of malignancy, cytological examinations are widely used by pathologists because they are simple, cheap, less invasive, and highly useful tools [4].

In a cytological examination, fluid from the malignant pleural effusion is collected and smeared on cytological glass slides using the staining methods. Then, cytologists or pathologists visually examine for morphology changes and visual abnormalities in every single cell under a microscope to determine malignancy prevalence [5]. Manual screening of cytology slides is tedious and subjective to inter- and intra-observer bias. Since the presence of MPE implies advanced malignancy and reduced survival, it is crucial to diagnose malignancy in MPE as early and speedy as possible. Thanks to recent improvements in medical technology, automated image analysis has the potential to allow for earlier and faster diagnosis with more accurate and objective diagnosis results. Hence, reliable CAD systems using CPE images are in high demand. They can serve as an essential tool to assist cytologists in the assessment of malignancy; however, complex and unusual cases still require further examination by cytologists. The benefits of CAD systems are that they accelerate the diagnosis process, make diagnosis objective, and reduce any diagnostic divergence resulting from different observers. Consequently, they allow for the early and speedy diagnosis and prognosis of cancer cells and help oncologists in making effective treatment plans promptly.

Few researchers have researched the analysis of CPE images for the automatic detection of cancerous cells from CPE specimens. In 2001, F. Chen et al. [6] proposed the automated classification of adenocarcinoma and healthy cells (especially mesothelial cells and lymphocytes) in CPE images. Morphology and wavelet features were used as inputs to a backpropagation neural network to discriminate between adenocarcinoma and benign cells. Their study was based on 60 adenocarcinoma cells and many (the number was not specified numerically) benign cells. Unfortunately, the

authors did not provide a method for segmenting nuclei nor an evaluation of classification performance. L. Zhang et al. 2006 [7] introduced a fuzzy recognition method to classify four types of cells, namely, healthy cells, cancer cells, mild dyskaryotic cells, and severe dyskaryotic cells. Otsu thresholding and fuzzy edge detection were used to segment the cells. Seven morphological features were extracted from each segmented cell and fed as input into a fuzzy recognition system to classify those four types of cells. However, there was a lack of clarity in the evaluation process in [4, 5]. This has encumbered the reproduction of these methods for practical use. A.B. Tosun et al. 2015 [8] presented the automated detection of malignant mesothelioma using nuclei chromatic distribution. Firstly, the nuclei were extracted using a semi-automatic approach in which the initial contour of cell nuclei was manually segmented under the guidance of cytologists, and level set method was utilized to finalize the contour of cell nuclei. For each extracted nucleus, its linear optimal transport (LOT) was computed and subjected to linear discriminant analysis based on k-nearest neighborhood algorithm classifier to differentiate between mesothelioma and benign cells. Their experiment was based on 1080 cell nuclei containing 590 mesotheliomas and 490 benign nuclei and obtained 100% accuracy. Unfortunately, their method was not fully automated since cell segmentation was manually performed. Moreover, none of the methods mentioned above deals with the overlapped cell problem. Decomposing overlapped cells into their constituents would enhance analysis performance and robustness. As such, the approaches mentioned thus far focus on detecting specific types of cancer cells such as adenocarcinoma or mesothelioma cells in CPE images. Meanwhile, an early and essential task in clinical practice is to differentiate between benign cells and cancer cells regardless of specific cancer types. This may then be followed by classifying cancer cells into the different types (i.e., lung carcinoma, mesothelioma, breast carcinoma, and so on). In practice, a tool that can detect malignant cells from all MPE cases is in high demand. Despite being linked with high rates of morbidity and mortality, research efforts for the automated analysis of MPE are still limited compared to other areas such as cervical cancer, breast cancer, lung cancer, and so on. Thus, automated analysis of pleural effusion samples remains to be widely researched.

To advance the utilization of MPE analysis, we propose a novel CAD system based on the analysis of CPE images which can classify cells as either benign or malignant. The main distinction of the proposed method from previous literature is that it can detect malignancy in all MPE cases. Our newly designed system is a fully automated system that addresses the overlapped cell and unbalanced-data problems which have so far been left unsolved. In addition, the proposed method takes advantage of the selection of dominant features using a hybrid metaheuristic method. Our system includes seven main stages: preprocessing, cell nuclei segmentation, postprocessing, overlapped cell nuclei isolation, feature extraction, feature selection, and classification. The preprocessing stage aims to improve the quality of the images. In the segmentation stage, our developed hybrid superpixel-driven K-Means clustering method, known as SLIC/K-Means

hybrid, was used to extract cell nuclei regions. Then, a series of morphological operations were employed to improve segmented cell nuclei boundaries and eliminate any false findings. Subsequently, the combination of the shape-based analysis and concavity analysis was applied to isolate any overlapping nuclei into individual ones. After the cell nuclei were segmented, a total of 201 features from morphometric, colorimetric, and textural features were extracted to create an initial feature set. Our novel hybrid SA-ANN feature selection approach was employed to obtain the optimal feature set that encompasses the most discerning features. The optimal feature set was fed as input to an ensemble classifier of bagged decision trees to classify benign and malignant cells.

This paper is divided into five sections. In this section we have presented an introduction to the diagnosis of malignancy in PE and outlined related works. The description of the studied dataset is given in Section 2. Section 3 describes the methodology used by the proposed CAD system. Section 4 discusses the experimental results. Section 5 concludes and presents the scope for future work.

2. Dataset Description

To date, there is no publicly available dataset of CPE images. Thus, we prepared the local dataset through the cooperation with experts from the Department of Pathology, Faculty of Medicine, Srinakharinwirot University, Thailand. The local dataset is based on the microscope images captured from the archival cytology glass slides of pleural effusion samples from the university mentioned earlier. Firstly, all samples were stained on the glass slides with a classical Papanicolaou (Pap) staining method which can provide good cellular morphology when inspected by the optical microscope [9, 10]. Then, two skilled and certified cytologists captured the digitized cytology images from the glass slides through a digital camera mounted to a light microscope with 40x magnification. Thereafter, they analyzed every single cell within the collected images and annotated the regions of the interest (i.e., cancer cells), which were used as the ground truth. The dataset with associated ground truth consists of 125 CPE images containing benign and malignant cells. The images have resolutions of 4050 x 2050 pixels and 4080 x 3702 pixels and are stored in 8-bit RGB space.

3. Methodology

The framework of the proposed CAD system is presented in Figure 2. The method involves seven major stages: (a) preprocessing, (b) nuclei segmentation, (c) postprocessing, (d) identification and isolation of overlapped cell nuclei, (e) feature extraction, (f) feature selection, and (g) classification.

3.1. Preprocessing Stage. During the staining of PE samples and digitalizing of CPE images, there is usually a degradation in quality, which includes uneven staining, uneven lighting, poor contrast, and the presence of additive noise. Therefore, preprocessing is essential in dealing with image quality prior to the main analysis. Firstly, the images were resized into

1024 x 1024 pixels in order to achieve image normalization, standardization, and computation time reduction. Then, each image was enhanced using an image intensity adjustment method that increases the contrast between the foreground (region of interests) and background [11]. In order to reduce noise without losing cell-edge clarity, R, G, and B components were separated from the original RGB image. Then, a median filter [12] was applied to each color component independently. Finally, the filtered RGB image was obtained by combining the filtered R, G, and B components together. The visual results before and after applying preprocessing to different images are depicted in Figures 3(a) and 3(b).

3.2. Segmentation of Cell Nuclei Using a Novel Hybrid SLIC/K-Means Algorithm. Segmentation is one of the most essential processes in biomedical image analysis. Most of the image analysis in cytology and histology is focused on nuclei segmentation since cell nuclei provide more significant diagnostic value than other cell parts. To determine cell malignancy, the cell nucleus needs to be segmented from the background (i.e., cytoplasm, red blood cells). Then, malignancy is predicted based on certain features extracted from each nucleus. Since the results of nuclei segmentation have a high impact on all subsequent analysis, it is crucial that the nuclei are accurately extracted.

Few researchers have studied the automated segmentation of cells or nuclei in CPE images. E. Baykal et al. 2017 [13] introduced an active appearance model to segment nuclei from the background in CPE images and compared it with color thresholding, clustering, and graph-based methods. They obtained 98.77% accuracy. However, their approach was designed to segment an image with only one cell. It is hard to use this in practice since there may be up to a million cells in one image. In [14], they investigated the detection of cell nuclei using supervised learning approach. The approach is based on the combination of Haar filter and AdaBoost classifier. Three images with a total of 178 nuclei were used for testing. A True Positive Rate of 89.32% and False Positive Rate of 5.05% were obtained. Their framework performed well with an independent cell nucleus; however, it showed limitations when it came to segmenting overlapped cell nuclei. Moreover, it required extensive prior knowledge to train the classifier. In our previous works [15], we have proposed several alternative nuclei segmentation methods such as Otsu thresholding approach, K-Means clustering approach [16], and supervised pixel classification using ANN [17] on a small dataset (24 CPE images). Recently, we collected more images and built a new dataset containing 35 CPE images. Using that new dataset, we employed twelve segmentation methods: (1) the Otsu method, (2) an ISODATA thresholding method, (3) a maximum entropy thresholding method, (4) cross-entropy thresholding, (5) minimum error thresholding, (6) fuzzy entropy thresholding, (7) adaptive thresholding, (8) K-Means clustering, (9) fuzzy C-means clustering, (10) mean shift clustering, (11) Chan-Vese level set, and (12) graph cut methods to extract the cell nuclei from CPE images, and we compared the results attained [18]. From the comparison results, Otsu, K-Means, mean shift clustering, graph cut

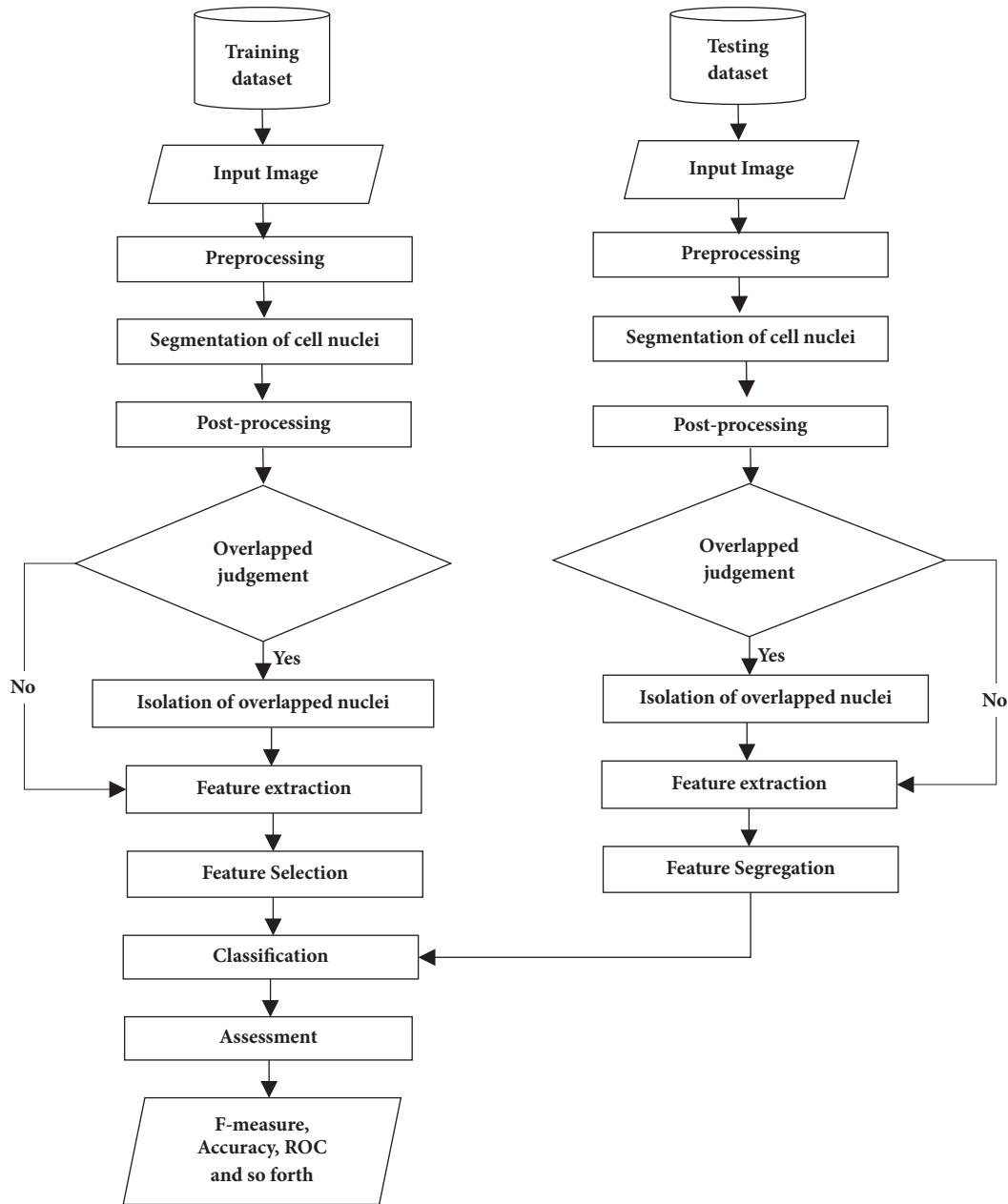


FIGURE 2: System framework of the proposed CAD system.

method, and a Chan-Vese level set method provided promising segmentation results. Although Otsu provided promising results with low computational time, the segmentation accuracy of Otsu showed degradation in images with a high level of noise because Otsu is sensitive to noise. The images in the studied dataset (124 images) have a great deal of noise. K-Means, mean shift, Chan-Vese, and graph cut methods were found to be computationally expensive especially with images containing a high population of cells. For machine learning based segmentation methods, prior knowledge is required to train a learning model. Thus, there are still opportunities for further enhancements in the nuclei segmentation of CPE images. Reliable nuclei segmentation stays challenging due

to the high population of cells and high diversity of cell appearance. In this study, we present a hybrid novel SLIC/K-Means based nuclei segmentation method in which SLIC superpixels are used as a presegmentation step to minimize the computational time of K-means clustering.

The first step of the hybrid SLIC/K-Means method is to perform superpixel segmentation as a presegmentation step. Superpixels fragment the image into a set of structurally meaningful segments where the boundaries of each segment take into the consideration the edge information from the original image. Superpixels are used in the preprocessing stage for object recognition and medical image segmentation. Among the various superpixel segmentation techniques, we

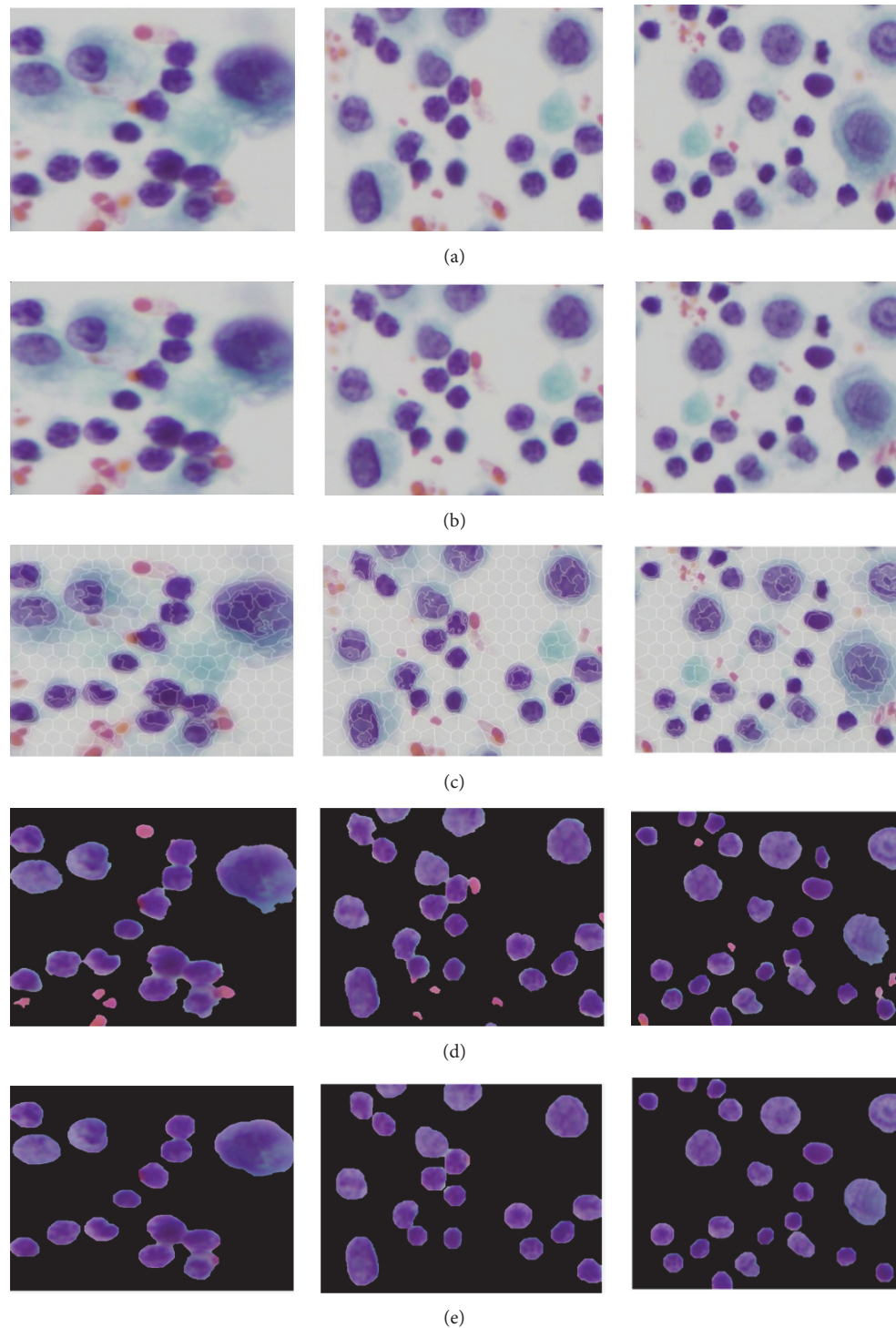


FIGURE 3: Visual results of segmenting cell nuclei from CPE images: (a) original image, (b) preprocessed image, (c) superpixels segmentation using SLIC, (d) K-Means based unsupervised color segmentation on SLIC superpixels, and (e) postprocessed image (refinement of nuclei boundary and elimination of false findings).

opted for a SLIC algorithm because SLIC generates compact superpixels with a more regular shape (R. Achanta et al. [19]). By breaking the image into regularly shaped superpixels, it is easier to distinguish between the nuclei and background depending on the superpixel shape. Moreover, SLIC is simple

to implement. It requires only the number of desired superpixels as the input parameter and needs a low computation time compared to other superpixel techniques [20]. SLIC generates compact, uniform superpixels by clustering pixels based on their color similarity and proximity. This is done

by using a combined 5-dimensional space [labxy], where l, a, b constitute the pixel color vector in LAB color model and xy denotes the x and y positional coordinates of the pixel position (x, y coordinates). SLIC takes as input the desired number of approximately uniform superpixels. Once SLIC generated the superpixels, we determined the median color feature of each superpixel region in the L*a*b* color space. K-Means clustering [21] was then utilized to classify the color feature of each compact superpixel into nuclei or non-nuclei, rather than having to perform clustering over the full original image pixels. Since representing the image as SLIC superpixels can give more accurate boundary information than representing the image by pixels, performing presegmentation using SLIC superpixels before K-Means clustering allows us to preserve the natural shape of cell nuclei. Also, it can reduce the complexity of the algorithm dramatically. This happens because the number of superpixels is much smaller than the number of pixels. Hence, applying K-Means clustering on SLIC superpixels, rather than on pixels, can improve the algorithm efficiency and lead to rapid computation. The visual results of nuclei segmentation on different images are illustrated in Figures 3(c) and 3(d).

3.3. Postprocessing Stage (Boundary Refinement of Cell Nuclei and False Findings Elimination). After the segmentation stage, spurious regions such as blood cells or artifacts still existed in the image. It is essential to remove these false findings for better accuracy and robustness. A series of morphological operations (MO) were used to eliminate these false findings as well as to refine the boundaries of the segmented nuclei. A morphological opening method was applied to eliminate false findings that were smaller than a predetermined structuring element (SE). After performing this opening operation, the boundaries of cell nuclei often hold an irregular shape. A morphological closing operation was subsequently utilized to refine the shape or boundary of the cell nuclei.

An important consideration when applying MO is the size and shape of SE. SE identifies the pixels in the image being processed and also designates the neighborhood to be employed in the processing of each pixel. There are two parameters (shape and radius) of SE to be specified. In our algorithm, both opening and closing operations are achieved by using a disk shape with an SE radius of “n”. The SE radii “n” should be determined according to the size of the undesired objects to be removed [22]. However, it is difficult to set SE radii of “n” that can work well across all images in a dataset or across different nuclei within an image. The optimal radius should be closely related to the size of the false findings that need to be eliminated. Setting too large structuring element size oversimplifies the image, while using too small SE undersupplies the images (blood cells or noise remain). Hence, we applied a multiscale approach. This means that each image was processed with different SE radii. For the opening operation, we adapted the SE radii range to be $n \{7, 8, \dots, 15\}$, which corresponds approximately to the expected range of undesired objects in the pleural effusion cell nuclei. For the closing operation, a

small SE (half the SE radii of the opening operation) size was adopted. The morphological opening and closing operations are mathematically formulated as follows:

$$Seg_{bi} \cdot SE = (Seg_{bi} \ominus SE) \oplus SE \quad (1)$$

$$Seg_{bi} \cdot SE = (Seg_{bi} \oplus SE) \ominus SE \quad (2)$$

where Seg_{bi} and SE denote the binary image and structuring element, respectively. \ominus and \oplus represent erosion and dilation, respectively. The visual results of this postprocessing are given in Figure 3(e).

3.4. Identification and Isolation of Overlapped Cell Nuclei.

Most of the pleural effusion images in this study contain nuclei that overlap to different degrees. Isolation of overlapped cell nuclei is essential for optimal segmentation performance since the size and shape of cell nuclei need to be determined accurately for quantitative analysis. To the best of our knowledge, the isolation of overlapped cell nuclei in CPE images has only previously been addressed in our previous works mentioned above. In our previous studies, we employed watershed variants such as marker-controlled and distance transform watershed methods to split overlapped cell nuclei. Unfortunately, these methods suffered from oversplitting and did not perform well on images with a great deal of overlapped cells. Existing splitting methods for overlapped objects can be broadly grouped into watershed methods and contour concavity analysis. With these methods, the points to be separated are searched across all objects in an image, and it is then determined whether to split them or not. In contrast, we now propose the integration of shape analysis and concavity analysis to identify and split overlapped nuclei for better accuracy and robustness. The proposed method contains two substages: the identification of overlapped cell nuclei and their separation into individual ones, the details for which are given in Sections 3.4.1 and 3.4.2. Before any splitting process occurs, shape analysis is performed to judge whether nuclei are single or overlapped. If any overlapped nuclei are detected, a splitting process based on concavity analysis is carried out only on overlapped cell nuclei rather than on all nuclei in the image. This process can reduce computation time and also prevent oversplitting and undersplitting.

3.4.1. Identification of Overlapped Cell Nuclei Using Shape-Based Analysis. During this step, we aimed to develop a shape-based predetermination mechanism to identify the presence of overlapped cell nuclei. Identification of overlapped cell nuclei was performed in two consecutive steps: (i) key features were extracted from cell nuclei and (ii) the cell nuclei were classified into two classes: single nucleus or overlapped nuclei. It is our general observation that shape features are useful in helping to differentiate between individual and overlapped cell nuclei. Hence, we extracted a set of shape features, containing solidity, eccentricity, equivalent diameter, major axis length, and minor axis length. The formulation of shape-based features is explained and shown in Figure 4. The extracted key features given in



FIGURE 4: Formula of extracted features.

TABLE 1: Extracted shape based features and their equations.

No.	Features	Formula
(1)	Solidity	$\frac{Area}{ConvexArea}$
(2)	Eccentricity	$\frac{F1}{a}$
(3)	Equivalent Diameter	$\sqrt{\frac{4 * Area}{\pi}}$
(4)	Major axis length	$2 * a$
(5)	Minor axis Length	$2 * b$

Table 1 were utilized as input to SVM classifier [23] to classify and discriminate between single and overlapped cell nuclei. SVM classifier is a supervised learning mechanism that requires training with pre-labeled training data. A trained SVM classifier was applied to identify overlapped cell nuclei in the image.

3.4.2. Splitting Overlapped Cell Nuclei Using Concavity Analysis. When overlapped nuclei were identified via shape analysis, we separated the overlapped nuclei regions from the single nucleus regions. Then, contour concavity analysis (CCA), introduced in [24], was utilized to isolate the overlapped cell nuclei into individual ones. CCA includes contour evidence extraction and contour estimation. Contour evidence extraction involves two subprocesses: contour segmentation and grouping. In contour segmentation, canny edge method was utilized to extract the edge map. Then, curvature scale space (CSS) method based on curvature analysis was applied to detect the concave points representing the corner points of the object boundaries. Once the contour segments were obtained through the detection of concave points, the contour segments belonging to the same object were merged through a grouping process. The grouping process was performed using the properties of fitted ellipse. It groups contour segments of objects composed of an elliptical shape. When contour evidence was acquired, the contour estimation was carried out using a stable direct least square fitting method. The visual result of identification and isolation of overlapped cell nuclei is illustrated in Figure 5.

3.5. Features Extraction. After the cell nuclei were accurately delineated, feature extraction was established to extract the features that reflect the observation of cytologists. In the literature of cytology and histology image analysis, the

dominant features for the diagnosis of malignancy used by cytologists are related to morphometric, colorimetric, and textural features [25–29]. In keeping with other cytological images, CPE images are also rich in various features like color, shape, and texture. In this study, 201 features related to the morphometric, colorimetric, and textural features were extracted and combined to obtain a robust, information-rich, and discerning feature set.

3.5.1. Morphometric Features. There are certain differences in morphology between benign and cancer cell nuclei in CPE images. For instance, excessive growth of cell nuclei size and a significant variation of cell nuclei size in an image are suggestive of malignancy. Moreover, cell nuclei shape irregularities such as unsmooth nuclei margins occur in malignant cases. Thus, in this study, 14 morphometric features were extracted to evaluate nucleus size and shape irregularity. The description of these features is given in Table 2 and coded as F1-F14.

3.5.2. Colorimetric Features. The usage of colorimetric features has tremendously increased in computer vision tasks due to their discriminative ability across different types of objects. Color provides useful information to determine malignancy. According to the cytological study, if any particular nuclei are affected by disease, the nucleus region changes in color. For instance, malignant cell nuclei become darker in color. In order to capture color features, means of R, G, B, H, S, and V components were extracted independently from RGB and HSV models. These features were coded in the range of F15 to F20.

3.5.3. Textural Features. In cytological pleural effusion images, malignant and cancer cell nuclei differ heavily in their distribution of color and chromatin. For instance, the frequent appearance of a distinct mass in a nucleus may be suggestive of malignancy. Texture features have been widely adopted in literature to exploit color and chromatin distribution. In this study, three statistical textural descriptors: first order statistics (FOS), gray level occurrence matrix (GLCM), and gray level run-length matrix (GLRLM) were employed to extract the textural features.

(1) Color Component Based First Order Statistics (CCFOS). FOS describes the distribution of pixel intensities within a nucleus region [30]. In related literature, the combination of color and FOS features has achieved better accuracy compared to conventional FOS features [31, 32]. Thus, seven FOS features for seven color components (namely, gray, R, G, B, H, S, and V from RGB and HSV model) were extracted for each nucleus. The extracted features were named by color component based on FOS (CCFOS) and encoded from F21 to F69. The reason for extracting seven color components was to obtain FOS textures from the view of different color components. Different color components describe the different defined textures as given in Figure 6. The details of these extracted features are given in Table 3 and coded from F21 to F69.

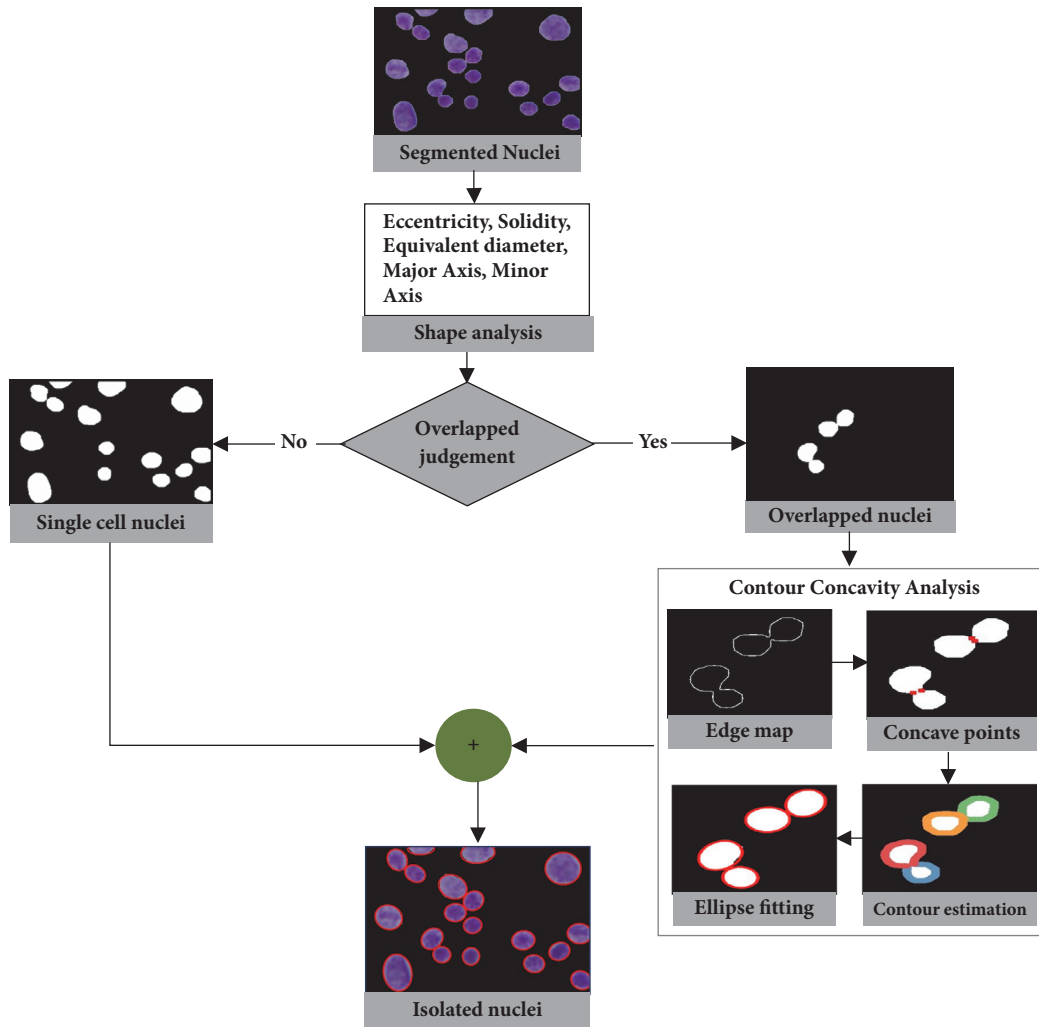


FIGURE 5: Visual demonstration of identification and splitting of overlapped cell nuclei.

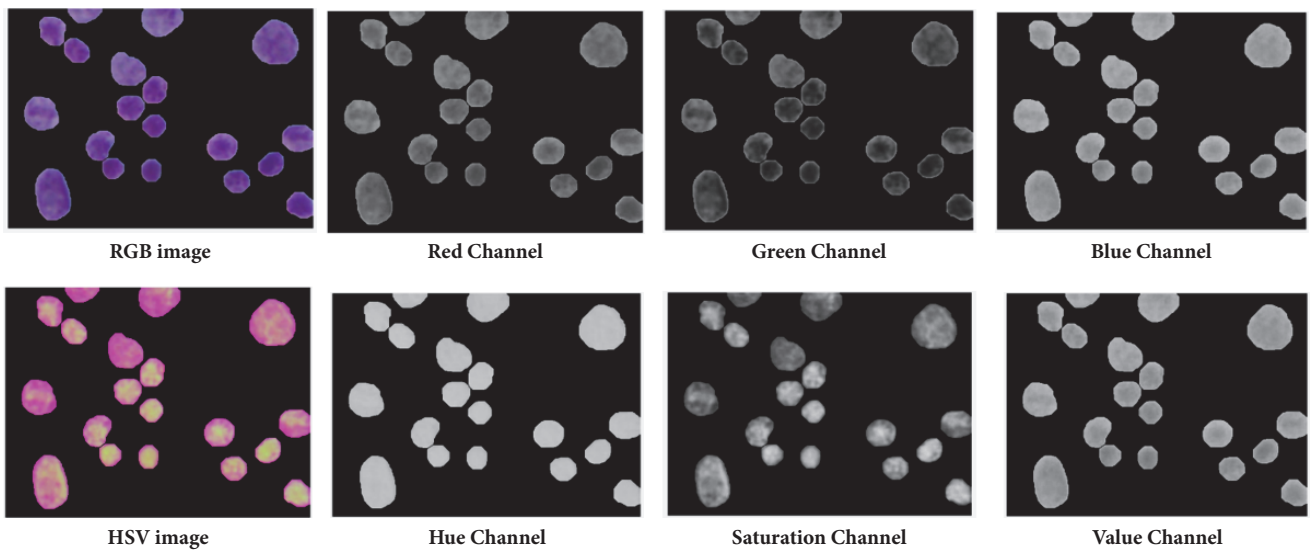


FIGURE 6: Individual color components of RGB and HSV color models in the segmented cell nuclei of CPE images.

TABLE 2: List of morphometric features and their associated equations.

Code	Feature Name	Equation
F1	Area	$\sum_{i=1}^n \sum_{j=1}^m S(i, j)$
F2	Perimeter	$Even\ count + \sqrt{2} (oddcount)$
F3	Roundness, circularity	$\frac{4\pi * Area}{Perimeter^2}$
F4	Solidity	$\frac{Area}{ConvexArea}$
F5	Equivalent circular diameter	$\sqrt{\frac{4 \times Area}{\pi}}$
F6	Compactness	$\frac{Area}{Perimeter^2}$
F7	Eccentricity	$2 * \frac{(\sqrt{(ma/2)^2 - (mi/2)^2})}{ma}$
F8	Diameter	$\frac{Perimeter}{2\pi}$
F9	Major axis length (<i>ma</i>)	$\sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2}$
F10	Minor axis length (<i>mi</i>)	$\sqrt{(x_2 - x_1)^2 - (y_2 - y_1)^2}$
F11	Elongation	$ma/perimeter$
F12	MaxIntensity	$\max(pixelValues)$
F13	MinIntensity	$\min(pixelValues)$
F14	MeanIntensity	$\text{mean}(pixelValues)$

$S(i, j)$ is the segmented image of rows i and columns j . ma and mi are the major axis and minor axis of the nucleus, respectively. x_1, y_1 and x_2, y_2 are the end points of the major axis and minor axis.

TABLE 3: List of CCFOS features and their associated equations.

Feature Name	Equation
Mean (μ)	$\sum_{i=0}^{L-1} ip(i)$
Standard deviation(σ)	$\sum_{i=0}^{L-1} (i - \mu)^2 \cdot p(i)$
Smoothness	$1 - \left(\frac{1}{L} + \sigma^2\right)$
Variance	$\sum_{i=0}^{L-1} (i - \mu)^2 p(i)$
Skewness	$\sigma^{-3} \sum_{i=0}^{L-1} (i - \mu)^3 p(i)$
Kurtosis	$\sigma^{-4} \sum_{i=0}^{L-1} (i - \mu)^4 p(i) - 3$
Energy	$\sum_{i=0}^{L-1} p(i)^2$

$p(i)$ is the number of pixels with gray level i , and L represents the number of gray-level bins set for p .

(2) *GLCM and GLRLM*. FOS captures features only on individual pixels. It ignores the spatial relationship between neighboring pixels. In order to capture texture features that take into account the spatial relationship between neighboring pixels, GLCM [33, 34] and GLRLM [35] based higher

order statistic features were considered. GLCM represents the distribution of cooccurring intensities in a nucleus at a specific given distance and orientation. When extracting GLCM features, it is required to define three parameters: distance (d) and orientations (θ) that determine the offset and angle between adjacent pixels, and the number of gray levels (NG) in the image. In this study, d and NG were set to 1 and 8, respectively. θ was adopted for four orientations $0^\circ, 45^\circ, 90^\circ, 135^\circ$ in order to take into account the rotation of the image. Thus, 22 GLCM features for four different orientations were extracted. GLRLM represents the length of homogeneous runs for each gray level in a definite direction. Similar to GLCM, GLRLM is constructed at four orientations and 8 gray levels. 11 GLRLM features in four different orientations ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) were extracted. Tables 4 and 5 describe the lists of GLCM and GLRLM feature and their associated equations. Finally, a feature vector was generated by combining 14 features of form morphology and 6 color features and 181 textural features from CCFOS, GLCM, and GLRLM. The list of extracted features is given in Table 6. The class of each nucleus is labeled as either positive or negative class under the guidance of cytologists.

3.6. *Feature Selection*. The initial feature set contains 201 features related to morphometry, colorimetry, and texture. Directly utilizing all candidate features for classification may cause redundancy and irrelevancy. Redundancy can lengthen

TABLE 4: List of GLCM features and their associated equations.

Features	Equations
Autocorrelation	$\sum_i \sum_j (i \cdot j) p(i, j)$
Contrast	$\sum_i \sum_j i - j ^2 p(i, j)$
Correlation I	$\sum_i \sum_j \frac{(i - \mu_x)(j - \mu_y) p(i, j)}{\sigma_x \sigma_y}$
Correlation II	$\sum_i \sum_j \frac{(i \cdot j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Cluster Prominence	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i, j)$
Cluster Shade	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i, j)$
Dissimilarity	$\sum_i \sum_j i - j \cdot p(i, j)$
Energy	$\sum_i \sum_j p(i, j)^2$
Entropy	$-\sum_i \sum_j p(i, j) \cdot \log(p(i, j))$
Homogeneity I	$\sum_i \sum_j \frac{p(i, j)}{1 + i - j }$
Homogeneity II	$\sum_i \sum_j \frac{p(i, j)}{1 + i - j ^2}$
Maximum Probability	$\max_{i,j} p(i, j)$
Sum of square	$\sum_i \sum_j (i - \nu)^2 p(i, j)$
Sum average	$\sum_{i=2}^{2L} i \cdot p_{x+y}(i)$
Sum energy	$-\sum_{i=2}^{2L} p_{x+y}(i) \cdot \log(p_{x+y}(i))$
Sum variance	$\sum_{i=2}^{2L} (i - \text{Sum energy})^2 \cdot p_{x+y}(i)$
Difference variance	$\sum_{i=0}^{L-1} i^2 \cdot p_{x-y}(i)$
Difference entropy	$-\sum_{i=0}^{L-1} p_{x-y}(i) \cdot \log(p_{x-y}(i))$
Information measure of correlation I	$\frac{(-\sum_i \sum_j p(i, j) \cdot \log(p(i, j))) - (-\sum_i \sum_j p(i, j) \cdot \log(p_x(i) p_y(j)))}{\max(-\sum_i p_x(i) \cdot \log(p_x(i)), -\sum_i p_y(i) \cdot \log(p_y(i)))}$
Information measure of correlation II	$\left(1 - \exp \left[-2 \left(\left(-\sum_i \sum_j p_x(i) p_y(j) \cdot \log(p_x(i) p_y(j)) \right) - \left(-\sum_i \sum_j p(i, j) \cdot \log(p(i, j)) \right) \right) \right] \right)^{1/2}$
Inverse Difference Normalized	$\sum_i \sum_j \frac{p(i, j)}{1 + i - j /L}$
Inverse difference moment normalized	$\sum_i \sum_j \frac{p(i, j)}{1 + (i - j)^2/L}$

$p(i, j)$ is the $(i, j)^{th}$ entry of the cooccurrence probability matrix, and L represents the number of gray levels used, while μ_x, μ_y and σ_x, σ_y are the mean and standard deviation of the p .

computation time. In turn, irrelevancy may cause poor predictive accuracy. To handle these problems, feature selection was performed in advance of classification. Feature selection is often applied in computer vision when many features

get extracted. It improves the prediction performance and generalization capability and provides a faster and more cost-effective model. Feature selection is generally divided into two techniques: filter and wrapper [36]. In filter techniques,

TABLE 5: List of GLRLM features and their associated equations.

Features	Equations
Short run emphasis (SRE)	$\frac{1}{n^r} \sum_{i=1}^G \sum_{j=1}^R \frac{g(i, j)}{j^2}$
Long run emphasis (LRE)	$\frac{1}{n^r} \sum_{i=1}^G \sum_{j=1}^R g(i, j) * j^2$
Low gray-level run emphasis (LGRE)	$\frac{1}{n^r} \sum_{i=1}^G \sum_{j=1}^R \frac{g(i, j)}{i^2}$
High gray-level run emphasis (HGRE)	$\frac{1}{n^r} \sum_{i=1}^G \sum_{j=1}^R g(i, j) * i^2$
Short run low gray-level emphasis (SRLGE)	$\frac{1}{n^r} \sum_{i=1}^G \sum_{j=1}^R \frac{g(i, j)}{i^2 * j^2}$
Short run high gray-level emphasis (SRHGE)	$\frac{1}{n^r} \sum_{i=1}^G \sum_{j=1}^R \frac{g(i, j) * i^2}{j^2}$
Long run Low gray-level emphasis (LRLGE)	$\frac{1}{n^r} \sum_{i=1}^G \sum_{j=1}^R \frac{g(i, j) * j^2}{i^2}$
Long run high gray-level emphasis (LRHGE)	$\frac{1}{n^r} \sum_{i=1}^G \sum_{j=1}^R g(i, j) * i^2 * j^2$
Gray level nonuniformity (GNU)	$\frac{1}{n^r} \sum_{i=1}^G \left[\sum_{j=1}^R g(i, j) \right]^2$
Run length nonuniformity (RNU)	$\frac{1}{n^r} \sum_{j=1}^{MG} \left[\sum_{i=1}^R g(i, j) \right]^2$
Run percentage (RP)	$\frac{n_r}{n_p}$

$g(i, j)$ denotes the number of runs of pixels of gray level i and the run length j , G is the number of gray levels in the image, R is the number of different run lengths in the image, n_r is the total number of runs, and n_p is the number of pixels in the image.

TABLE 6: List of various features extracted from each nucleus.

Name of Feature sets	Number of Features	Ranges
Morphometric Features	14	F1-F14
Colorimetric Features	6	F15-F20
CCFOS (Textural Features)	49	F21-F69
GLCM (Textural Features)	88	F70-F157
GLRLM (Textural Features)	44	F158-201
Combined Feature Set	201	F1-F201

the features are chosen depending on their relevance ability with respect to the target. Filter methods are computationally fast and easy to implement. However, there is a possibility that the chosen features might contain redundant information since the selection process is carried out on the statistical measure of each feature. Unlike the filter approach, the wrapper approach depends on learning methods. It utilizes the estimated accuracy of the learning method as a performance measure to evaluate the usefulness of a feature. As an extension of the wrapper approach, the hybrid approach, which combines metaheuristics methods and supervised learning methods as integral components of feature selection, has been widely utilized in medical image analysis [37–39]. Experiments have found that hybrid methods are more

efficient in finding optimal solutions compared to filter and wrapper methods. The main benefit of the hybrid methods is the ability to avoid being stuck in the local optima. In this study, a novel hybrid feature selection method based on hybridizing simulated annealing, one of the metaheuristics methods, with an artificial neural network, one of the popular machine learning methods, was developed to select the most relevant and informative features. The proposed method is known as a hybrid simulated annealing coupling artificial neural network (SA-ANN) feature selection. The details of SA-ANN are given in the subsection below.

3.6.1. Hybrid SA-ANN Feature Selection. Simulated annealing is a global optimization algorithm that is inspired by the natural annealing process in metallurgy. It models the annealing process of heating material and then gradually cooling it by lowering the temperature at a controlled rate, thus minimizing system energy [40]. It is typically used to search for the global minimum in a high-dimensional data space. The main advantage of SA is that it allows up-hill moves in the iteration to avoid being stuck at a local minimum. SA has been widely used as a supervised or unsupervised feature subset selection method in data mining techniques, especially for microarray gene classification in biomedical data analysis [41–43]. Inspired by those works, in this study, we developed

```

Input: Features_set, MaxIt, Temp, alpha
Output:  $S_{best}$ 
 $Initial_{subset} \leftarrow CreateInitialSolution(Features\_set)$ 
 $Cost(S_{current}), S_{current} \leftarrow CostFunction\_ANN(Initial_{subset})$ 
 $S_{best} \leftarrow S_{current}$ 
For ( $i = 1 : MaxIt$ )
   $New_{subset} \leftarrow CreateNeighbourSolution(S_{current})$ 
   $Cost(S_i), S_i \leftarrow CostFunction\_ANN(New_{subset})$ 
  if ( $Cost(S_i) \leq Cost(S_{current})$ )
     $S_{current} \leftarrow S_i$ 
  elseif ( $Exp\left(\frac{CostS_{current} - CostS_i}{Temp}\right) > Rand()$ )
     $S_{current} \leftarrow S_i$ 
  End
  if ( $Cost(S_{current}) \leq Cost(S_{best})$ )
     $S_{best} \leftarrow S_{current}$ 
  End
 $Temp = Temp * alpha$ 
End
Return( $S_{best}$ )

```

ALGORITHM 1: The main loop of hybrid SA-ANN based feature selection.

a novel hybrid feature selection method by hybridizing SA with an artificial neural network (ANN). ANN is a machine learning algorithm that mimics the structure of the biological brain. During feature selection via hybrid SA-ANN, the cost value of SA based search space was computed depending on the number of samples correctly predicted by ANN. Firstly, the random initial feature subsets were created. These subsets were assessed using a 3-layer ANN trained by a Levenberg-Marquardt (LM) backpropagation algorithm [44] containing a fixed number of hidden neurons. The features with the most minimal cost were initialized as the best feature set. At each iteration of SA, the neighboring subset was randomly generated by implementing a neighborhood function. Then, in a similar manner to the first stage, a 3-layer ANN trained by LM backpropagation algorithm was used to evaluate the cost of the neighboring subset. If the neighboring subset had a lower cost than the initial subset, we would then change the initial subset to its neighboring subset. Alternatively, if the neighboring subset had a higher cost, then the individual would move to that subset only if the acceptance probability condition was fulfilled. Otherwise, the individual remained in the initial subset. By accepting individuals that increase the cost, the algorithm avoids getting stuck by a local minimum in early iterations and explores globally for better solutions. As the algorithm progresses, the temperature is reduced causing individuals to converge towards the subset with a minimum cost and hence an optimal point. Hybrid SA-ANN feature selection can be summarized using the pseudocode in Algorithm 1, wherein feature set, MaxIt, Temp, and alpha are the candidate features, maximum numbers of iteration, initial temperature, and the temperature reduction rate, respectively. S_{best} is the output that represents the corresponding optimal feature set. The selected features in the optimal feature set were utilized for training and testing the

classifier. The code implementation of proposed hybrid SA-ANN feature selection is based on the Matlab implementation available in [45] and modified as necessary.

3.7. Classification. The selected features were utilized as input to the classifier to differentiate between benign and malignant cells. In cytology and histology image analysis, classification models revolve around Support Vector Machine (SVM) [26, 27], Naïve Bayes (NB) [27], artificial neural network (ANN) [28], K-nearest neighborhood (KNN) [8, 27], Logistic Regression (LR) [29], Linear Discriminant Analysis (LDA) [8], Decision Tree (DT) [46], and Ensemble Classifier (EC) [31]. The selection of a classification model for medical image analysis depends on the type and size of the dataset to be classified. Our dataset of cell nuclei was large and highly unbalanced wherein the class of cancer nuclei was limited while the class of benign nuclei was abundant. Ensemble classification has yielded preferable results for classification of skewed data [47, 48]. Thus, to deal with the unbalanced-data distribution, we adopted an ensemble classifier that employs bootstrap aggregation (bagging) decision trees and is termed as ECBDT [49, 50]. The core idea of using ECBDT was to develop multiple bootstrap data-samples and to build multiple base classifiers for each bootstrapped sample. One hundred decision trees were used as the base classifiers. The final prediction of ECBDT was obtained through major voting. The block diagram of the ECBDT classifier is depicted in Figure 7. The classifier was trained in 5-fold cross-validation.

4. Experiments

4.1. Experimental Setup. The proposed CAD system presented here was developed in a Matlab environment using a PC with Intel® Core i7, CPU@3.40 GHz, RAM@16.0 GB. The study was based on 125 cytology pleural effusion images containing around 10500 cells. The studied dataset was randomly partitioned into training and testing sets in an 80-20% ratio. 80% of the images were allocated to the training dataset to train the classifier and 20% to the testing dataset to validate the trained classifier. Training and testing datasets were disjointed (i.e., the same image was not assigned to represent both training and testing datasets). It is noteworthy that all the experiments carried out in this study are based on the same experimental setting and environment.

4.2. Experimental Results and Discussion. To obtain a comprehensive discussion, the experimental results are discussed in two phases. The first phase is the segmentation phase, which encompasses preprocessing, the segmentation of cell nuclei, postprocessing, and the isolation of cell nuclei. The second phase is the classification phase, which comprises feature extraction, feature selection, and classification.

4.2.1. Segmentation Phase. Intensity adjustment and median filter methods were employed to enhance image contrast and suppress the noises, respectively. Then, a novel hybrid SLIC/K-Means segmentation method was developed to segment the cell nuclei from the entire image. In SLIC/K-Means,

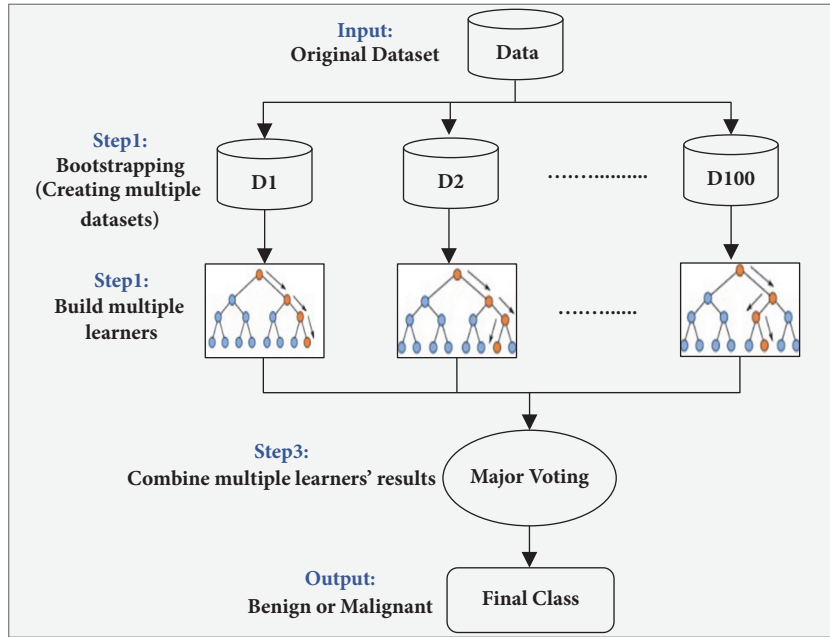


FIGURE 7: Block diagram of ensemble classifier of bagged decision trees (ECBDT) used in this study.

TABLE 7: Comparison of time complexity in segmentation methods using testing images.

Segmentation methods	Average processing time
Classical K Means	66.6 seconds
Proposed Method	5.8 seconds

TABLE 8: Comparison of time complexity in splitting methods using testing images.

Splitting methods	Average processing time
Concavity analysis	10.2 seconds
Proposed method	6.8 seconds

the SLIC method is firstly performed to presegment the image into the small compact superpixels. Then, K-Means clustering is carried out to cluster each superpixel into two groups by using the extracted features from superpixels. Features extracted over the uniform and compact SLIC superpixels tend to be more discriminative, helping K-Means to produce better segmentation. Good adherence to the image boundaries exhibited by SLIC superpixels results in smoother and more accurate segmentation. Utilizing K-Means clustering on superpixels can shorten computation because the number of superpixels is significantly lower than the number of pixels. It scales up linearly in computational cost and memory usage. The proposed segmentation method extracts cell nuclei at a lower computational cost and preserves the natural shape of the cell nuclei while achieving excellent segmentation results. In the hybrid SLIC/K-Means segmentation method, we need to specify two parameters: the number of superpixels for SLIC and the k clusters for K-Means. The desired number of superpixels was set to 500. According to our previous work, k was set at 2 because cell nuclei are segmented in a straightforward way when k is 2. False findings such as artifacts or blood cells may present obstacles to accurate segmentation. These undesired regions were filtered out with a series of morphological operations. Subsequently, the boundaries of cell nuclei were further refined. The visual results of the proposed SLIC/K-Means and classical

K-Means, supplemented by the same preprocessing and postprocessing approaches, are demonstrated in Figure 8. Compared to classical K-Means clustering based segmentation, the proposed method performs better in preserving the natural shape of the cell nuclei. Moreover, it is significantly faster than classical K-Means in computation, as given in Table 7.

Almost all the images in the studied dataset possessed an overlapped cell nucleus to different degrees. Separating them into individual ones was hence essential. In almost all related literature, cell splitting is applied directly on the entire segmented image. This means that the splitting method is processed not only on overlapped regions but also on single cell nuclei regions. Such an attempt can lengthen computation time. In contrast, we propose a sequential combination of shape-based analysis and concavity analysis to identify overlapped areas and isolate them into individual ones. First, shape-based analysis was performed to determine the overlapped cell nuclei and separate them from single cell nuclei regions. Then, contour concavity analysis based splitting is applied only on the identified overlapped nuclei, rather than on all nuclei in the image. By identifying overlapped regions before applying the splitting method, one can not only prevent over- and undersplitting but also shorten computation time, as tabulated in Table 8. The visual results of splitting overlapped cell nuclei are illustrated in Figure 9.

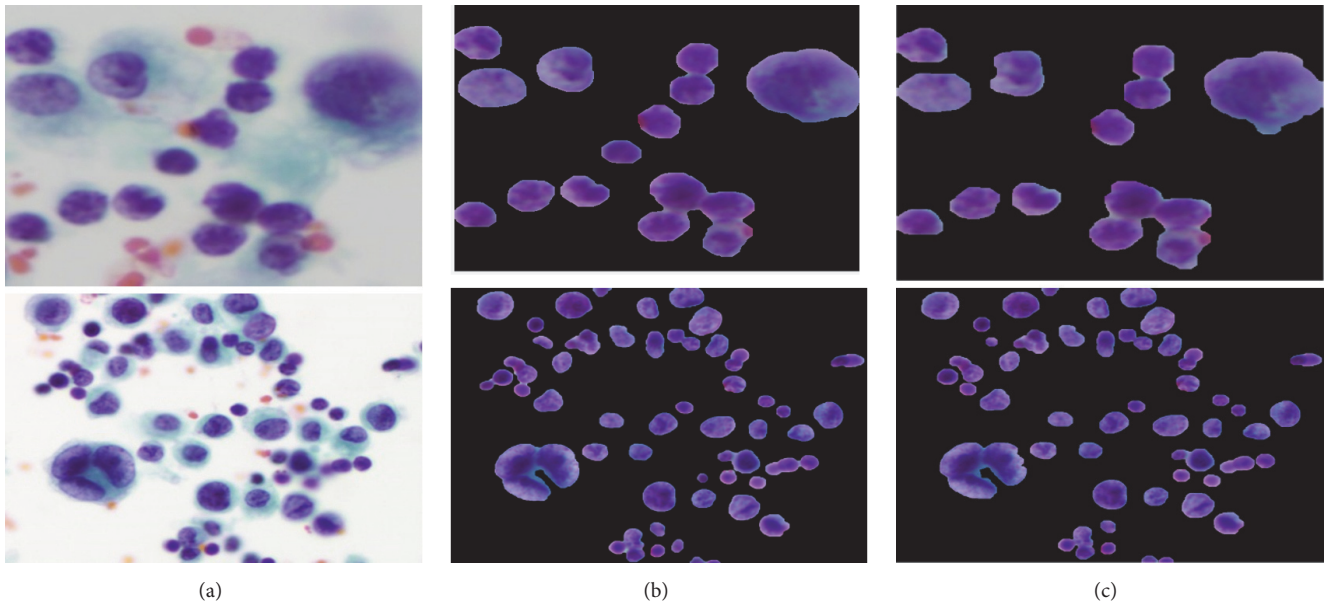


FIGURE 8: Comparison results of nuclei segmentation methods: (a) original image, (b) proposed method (SLIC + K-Means), and (c) K-Means clustering based segmentation.

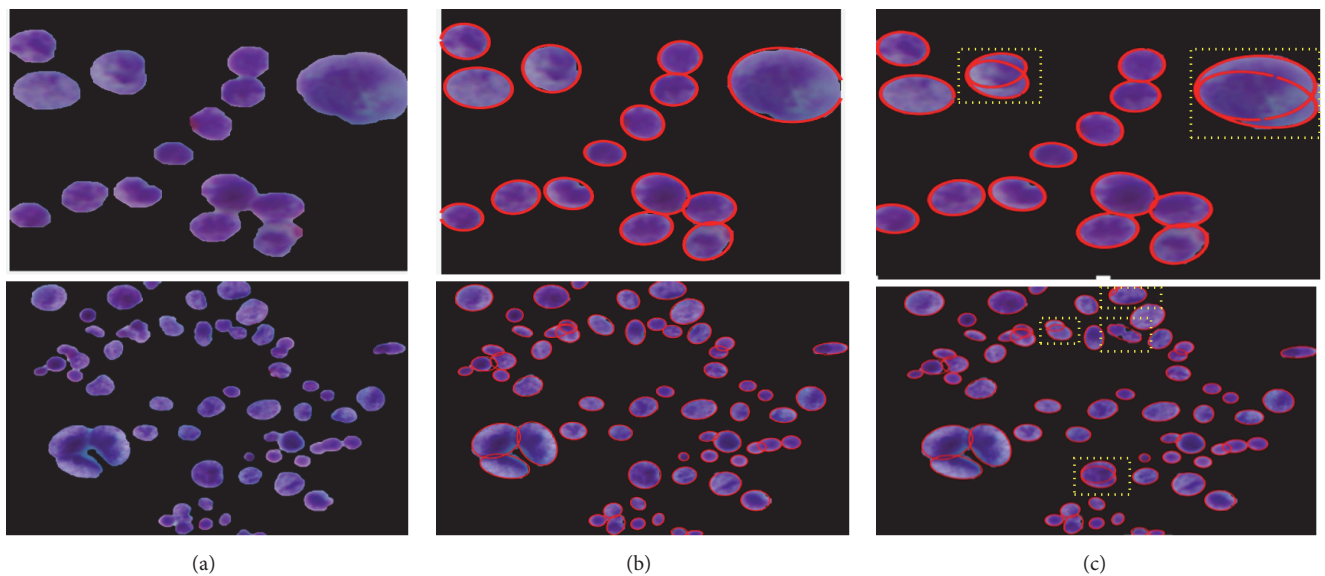


FIGURE 9: Comparison results of overlapped nuclei splitting methods: (a) segmented nuclei (**input**), (b) proposed splitting method based on the combination of shape analysis and concavity analysis, and (c) contour concavity analysis (**note that the yellow rectangular box indicates the over- and undersplitting**).

Figure 9(a) shows the segmented nuclei image. Figure 9(b) represents the resulting images from our proposed splitting methods (i.e., the combination of shape analysis and contour concavity analysis) and Figure 9(c) depicts the resulting images from classical contour concavity analysis. As shown in Figure 9(b), employing a splitting method only on the identified overlapped region can prevent the single cell nuclei from oversplitting and overlapped cell nuclei from undersplitting. This happens because the splitting method is focused solely on the overlapped area. The yellow shading

box in Figure 9(c) is illustrated to highlight the over- and undersplitting which result from using the classical concavity analysis based splitting method.

4.2.2. Classification Phase. Once the nuclei were accurately delineated, 201 features representing the morphometric, colorimetric, and textural features were extracted from each nucleus. In order to avoid redundancy and irrelevancy, hybrid SA-ANN feature selection was developed to choose the most discerning and informative features. Promising features that

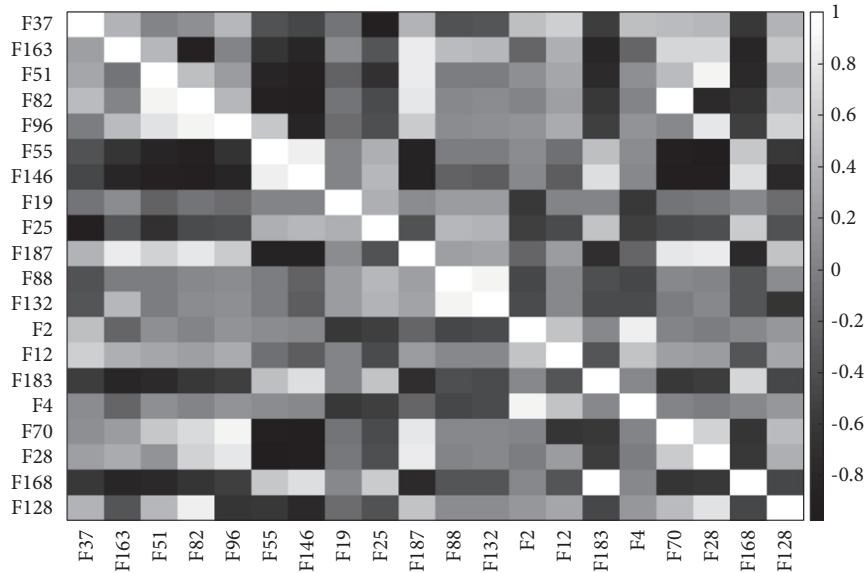


FIGURE 10: Correlation matrix for the selected features using hybrid SA-ANN feature selection (note that correlation =1 (white) means the highest correlation, -0 (black) no correlation).

TABLE 9: Description of selected features through hybrid SA-ANN feature selection.

No.	Feature Code	Feature Name	Feature Set
(1)	F37	Smoothness of B component	CCFOS
(2)	F163	Short run high gray-level emphasis	GLRLM0
(3)	F 51	Smoothness of S component	CCFOS
(4)	F 82	Sum of square	GLCM0
(5)	F 96	Cluster Prominence	GLCM45
(6)	F 55	Energy of S component	CCFOS
(7)	F 146	Homogeneity II	GLCM 135
(8)	F 19	Mean color of S component	Colorimetric
(9)	F 25	Skewness of R component	CCFOS
(10)	F 187	Long run high gray-level emphasis	GLRLM 90
(11)	F 88	Information Measure of Correlation	GLCM0
(12)	F 132	Difference Entropy	GLCM 90
(13)	F 2	Perimeter	Morphometric
(14)	F 12	MaxIntensity	Morphometric
(15)	F 183	High gray-level run emphasis	GLRLM 90
(16)	F4	Solidity	Morphometric
(17)	F 70	Autocorrelation	GLCM 0
(18)	F 28	Mean from G component	CCFOS
(19)	F 168	Run percentage	GLRLM0
(20)	F 128	Sum Entropy	GLCM0

correctly map to the target are identified by supervised ANN and used in the annealing process. The SA-ANN algorithm was iterated 50 times with an initial temperature (temp=10) and temperature reduction rate (alpha=0.99). The algorithm was adapted to select a different desired number of features (nf) such as 15, 20, 25, 30, 35, and 40. Based on the experimental results obtained, it was deduced that

selecting more than 20 features resulted in slightly decreased classification accuracy. Thus, the SA-ANN algorithm was fixed to select 20 features out of 201 features. The list of selected features and their correlation matrix are described in Table 9 and Figure 10, respectively. By analyzing the selected features, it was revealed that they included one or more representative features from each group of features given in

Section 3.5. Among 20 selected features, 16 features were textural features. Thus, it is reasonable to conclude that textural features supply more diagnostic information than other features. Moreover, the correlation matrix demonstrates that proposed hybrid SA-ANN feature selection selected the most significant features with less redundant information. The selected features were used as input to the classification model to predict malignancy. Classification model choice depends on the size and the type of data to be predicted. Our data is highly skewed, wherein the cell nuclei, belonging to malignant (positive), were limited, and the cell nuclei

belonging to benign (negative) were abundant. Thus, we adopted ensemble classification which provides preferable results to the classification of unbalanced data. As mentioned in Section 3.6, the dataset was firstly bootstrapped randomly, and 100 decision trees were used as the base classifiers to classify the bagged datasets. The final classification result was obtained through major voting. To evaluate classification performance, we compared the ground truth and classification results with respect to four performance metrics: sensitivity, specificity, F-score, and accuracy. These four performance measures are formulated in (3)-(8).

$$\text{Sensitivity} = \frac{\text{TruePositive}}{(\text{TruePositive} + \text{FalseNegative})} \quad (3)$$

$$\text{Specificity} = \frac{\text{TrueNegative}}{(\text{TrueNegative} + \text{FalsePositive})} \quad (4)$$

$$\text{Precision} = \frac{\text{TruePositive}}{(\text{TruePositive} + \text{FalsePositive})} \quad (5)$$

$$\text{Recall} = \text{Sensitivity} \quad (6)$$

$$F - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (7)$$

$$\text{Accuracy} = \frac{(\text{TruePositive} + \text{TrueNegative})}{(\text{TruePositive} + \text{FalsePositive} + \text{FalseNegative} + \text{TrueNegative})} \quad (8)$$

To make a fair and objective comparison, a common public dataset is required. By far, we are not aware of any common publicly available dataset. Also, the diagnosis schemes of CPE images in related literature are different from the proposed diagnosis scheme. Thus, we built our own experimental setup wherein the impact of using different feature selection methods and different classification models on classification performance was observed. In the first three experimental scenarios, we compared the classification accuracy achieved with and without features using the proposed classifier (i.e., ECBDT). In the first scenario, we compared the results between our proposed SA-ANN approach and an “all features” approach (i.e., without feature selection). Secondly, the result of the SA-ANN approach was compared with the results of the SA approach. In the third scenario, we established a comparison between the SA-ANN approach and other robust hybrid feature selection methods: PSO-ANN and GA-ANN approach. Furthermore, in the fourth experimental scenario, we employed seven alternative classifiers, namely, SVM [23], ANN [51], NB [52], KNN [53], LR [52], LDA [54], and DT [55] classifiers, and coupled them with the feature selection approaches. The result achieved by the proposed synergy between SA-ANN feature selection and ECBDT classification was compared with the results obtained through various pairings. Therefore, for each feature selection approach, the experimental results are presented with respect to four performance measures and eight classification models (including ECBDT). The results

from four experimental scenarios are shown in Table 10. We clarify that hybrid SA-ANN coupling with an ECBDT classifier (shaded in bold) is our proposed method. As reported in Table 10, utilizing the feature selection methods (i.e., SA-ANN, SA, PSO-ANN, GA-ANN, or SA) provided better accuracy compared to the all features approaches (i.e., without feature selection) for all classifiers. The results also demonstrate that, with the exception of coupling with SVM, KNN, and LR classifiers, the proposed SA-ANN selection marginally improves accuracy compared to the SA based approach and yields better accuracy compared to PSO-ANN and GA-ANN approaches when coupling with ANN, NB, LD, DT, and proposed ECBDT classifiers. When coupling with an SVM classifier, the PSO-ANN approach yields better results compared to other selection approaches. Similarly, the GA-ANN approach yields better accuracy compared to other feature selection methods when coupling with KNN classifier. Likewise, the SA approach yields better accuracy compared to other feature selection methods when coupling with LR. The superior feature selection method for each classifier is shown in italic. It was observed that different classifiers perform differently for different selected features. However, regardless of the feature selection methods utilized, ECBDT (ensemble classifier) consistently provided better accuracy compared to other single classifiers. From the experimental results, it is inferred that the synergy of hybrid SA-ANN coupling with an ECBDT classifier outperformed other pairs of feature selection approaches and classification models described

TABLE 10: Comparison of classification performance achieved by different synergy between feature selection methods and classification models.

Feature Selection (FS)	Performance Metrics	Classifiers							
		SVM	ANN	NB	KNN	LR	LDA	DT	Proposed ECBDDT
All features (No FS)	Sensitivity	72.18%	75.19%	66.17%	72.93%	71.43%	75.19%	71.43%	74.48%
	Specificity	95.47%	94.48%	93.41%	95.51%	94.82%	95.12%	94.10%	96.11%
	F-score	57.31%	55.25%	46.93%	57.91%	54.44%	57.64%	51.91%	61.73%
	Accuracy	94.21%	93.44%	91.95%	94.29%	93.57%	94.05%	92.88%	94.98%
PSO-ANN	Sensitivity	73.65%	70.91%	69.16%	74.29%	69.23%	69.16%	71.83%	76.47%
	Specificity	96.64%	96.11%	95.67%	96.72%	96.11%	95.67%	96.32%	97.09%
	F-score	76.29%	77.33%	74.30%	75.96%	69.96%	74.30%	80.42%	80.28%
	Accuracy	97.21%	97.25%	96.64%	97.21%	97.29%	96.64%	97.73%	97.73%
GA-ANN	Sensitivity	87.97%	86.47%	64.66%	87.97%	87.22%	64.66%	86.47%	86.47%
	Specificity	97.22%	97.09%	99.44%	98.20%	97.31%	99.44%	98.63%	98.93%
	F-score	74.29%	72.78%	74.14%	80.14%	74.36%	74.14%	82.14%	84.25%
	Accuracy	96.72%	96.52%	97.57%	97.65%	96.76%	97.57%	97.98%	98.26%
SA	Sensitivity	85.71%	86.47%	90.23%	84.21%	84.96%	90.23%	84.21%	87.22%
	Specificity	97.22%	97.73%	97.60%	97.52%	98.37%	97.60%	99.14%	99.27%
	F-score	73.08%	76.41%	77.67%	73.93%	79.58%	77.67%	84.53%	87.22%
	Accuracy	96.60%	97.13%	97.21%	96.80%	97.65%	97.21%	98.34%	98.62%
Proposed SA-ANN	Sensitivity	85.71%	72.93%	72.93%	84.21%	79.70%	72.93%	86.47%	87.97%
	Specificity	97.22%	99.70%	99.66%	97.52%	98.16%	99.66%	99.27%	99.40%
	F-score	73.08%	81.86%	81.51%	73.93%	75.18%	81.51%	86.79%	87.79%
	Accuracy	96.60%	98.26%	98.22%	96.80%	97.17%	98.22%	98.58%	98.70%

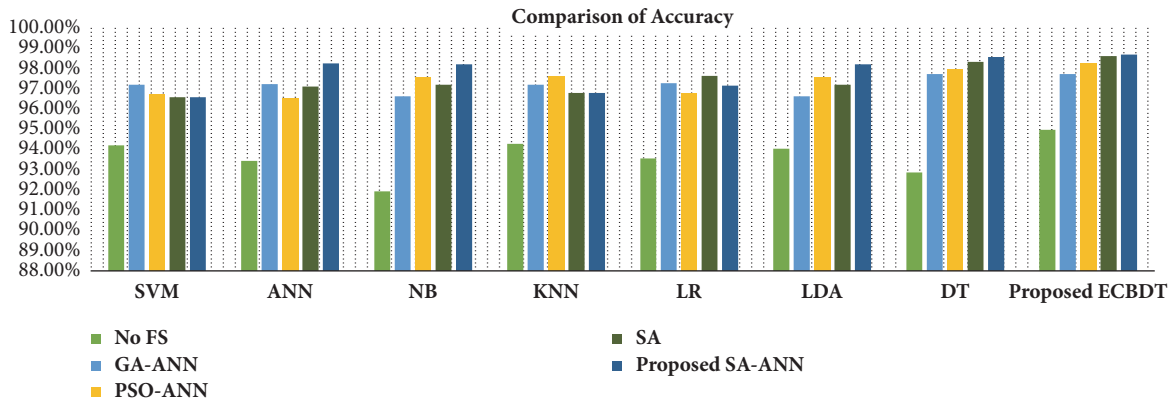


FIGURE 11: Comparison of accuracy using different pairs of feature selection methods and classifiers.

above in terms of classifying cells in CPE images. To get clear comparison results, we further plotted the comparison of accuracy and F-score as illustrated in Figures 11 and 12, respectively. Moreover, a Receiver Operating Characteristics (ROC) curve for different classifiers coupling with SA-ANN feature selection is depicted in Figure 13. It shows that the ROC curve of the proposed method is on the left upper corner and has higher classification rate stability when compared to other methods in the study. The visual results of detected malignant nuclei (both correct and failed cases) are depicted in Figure 14. Figure 14(a) shows annotated malignant cell nuclei labeled by two experts in which blue and green represent the two experts. Figure 14(b) describes the

diagnostic results of the proposed CAD system wherein the red bounding boxes represent the detected malignant cells. Even though the proposed method yields promising results, there are still some failures especially when the malignant characteristics of a cell occur in the cytoplasm. Therefore, it remains for future work to detect for malignancy based on the combined analysis of cell nuclei and cytoplasm.

5. Conclusion

In this study, we presented a novel CAD system to detect cancer cells on CPE images. Firstly, intensity adjustment

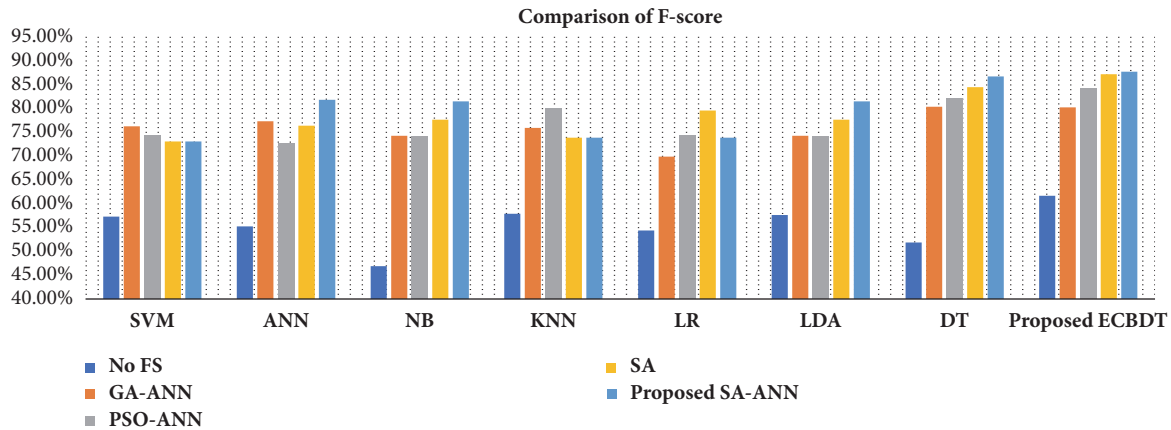


FIGURE 12: Comparison of F-score using different pairs of feature selection methods and classifiers.

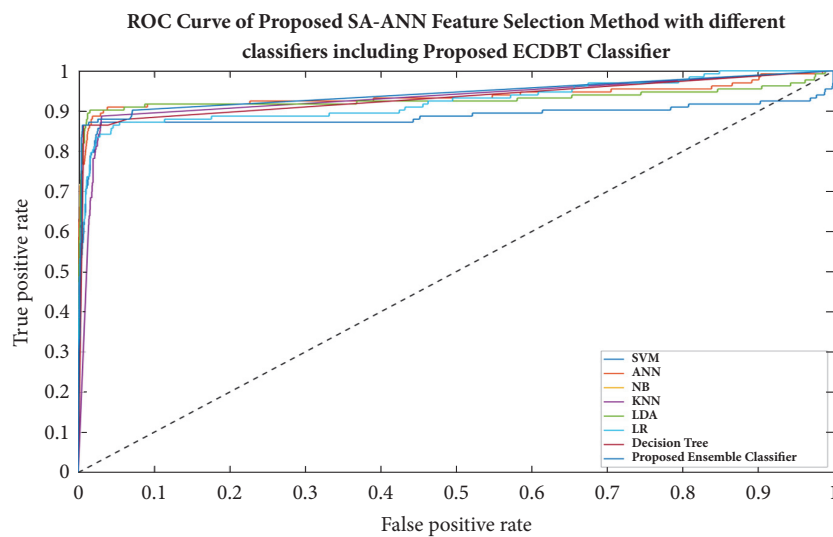


FIGURE 13: ROC curve for the performance of SA-ANN feature selection by blending with eight different classifiers.

and median filter methods were employed to enhance image contrast and suppress noise, respectively. Then, the cell nuclei were extracted using a novel hybrid SLIC/K-Means segmentation method followed by postprocessing. Overlapped nuclei regions were then identified through shape-based analysis. Subsequently, concavity analysis was utilized to isolate the detected overlapping regions into individual ones. After the cell nuclei were accurately delineated, 201 features that comprise the morphometric, colorimetric, and textural features were extracted from each nucleus. A feature selection framework based on a hybrid SA-ANN was developed to select the most significant and informative features from the initial feature set containing those 201 features. The chosen features were used as input into ECBDT classifier to predict for malignancy. The proposed method can achieve 87.97% sensitivity, 99.40% specificity, 98.70% accuracy, and 87.80% F-score. The results achieved were compared with the results gained through an “all features”, SA, PSO-ANN, and GA-ANN approaches by coupling with eight different classifiers, namely, ECBDT, SVM, ANN, NB, KNN, LR, LDA, and DT.

The comparison results demonstrated that the hybrid SA-ANN approach significantly improves accuracy compared to the “all features” approach for all classifiers. It marginally improves accuracy compared to the PSO-ANN, GA-ANN, and SA methods for most classifiers. Furthermore, the ECBDT classifier consistently improves classification performance compared to other individual classifiers: SVM, ANN, NB, KNN, LR, LDA, and DT. The proposed CAD system based on the synergy between SA-ANN feature selection and an ensemble classifier outperformed all other combinations conducted in this study. Nevertheless, there were still some failures, especially when the malignant characteristics of a cell occur in the cytoplasm. Hence, the future work of this research is to extend the combined analysis of cytoplasm and nuclei and further classify the detected malignant cells into different types, such as lung carcinoma, breast carcinoma, mesothelioma, and lymphoma. There is also a potential of adapting the proposed CAD system to the same kind of cytopathology images captured from other body fluid types such as the peritoneal, cerebrospinal, and synovial fluid.

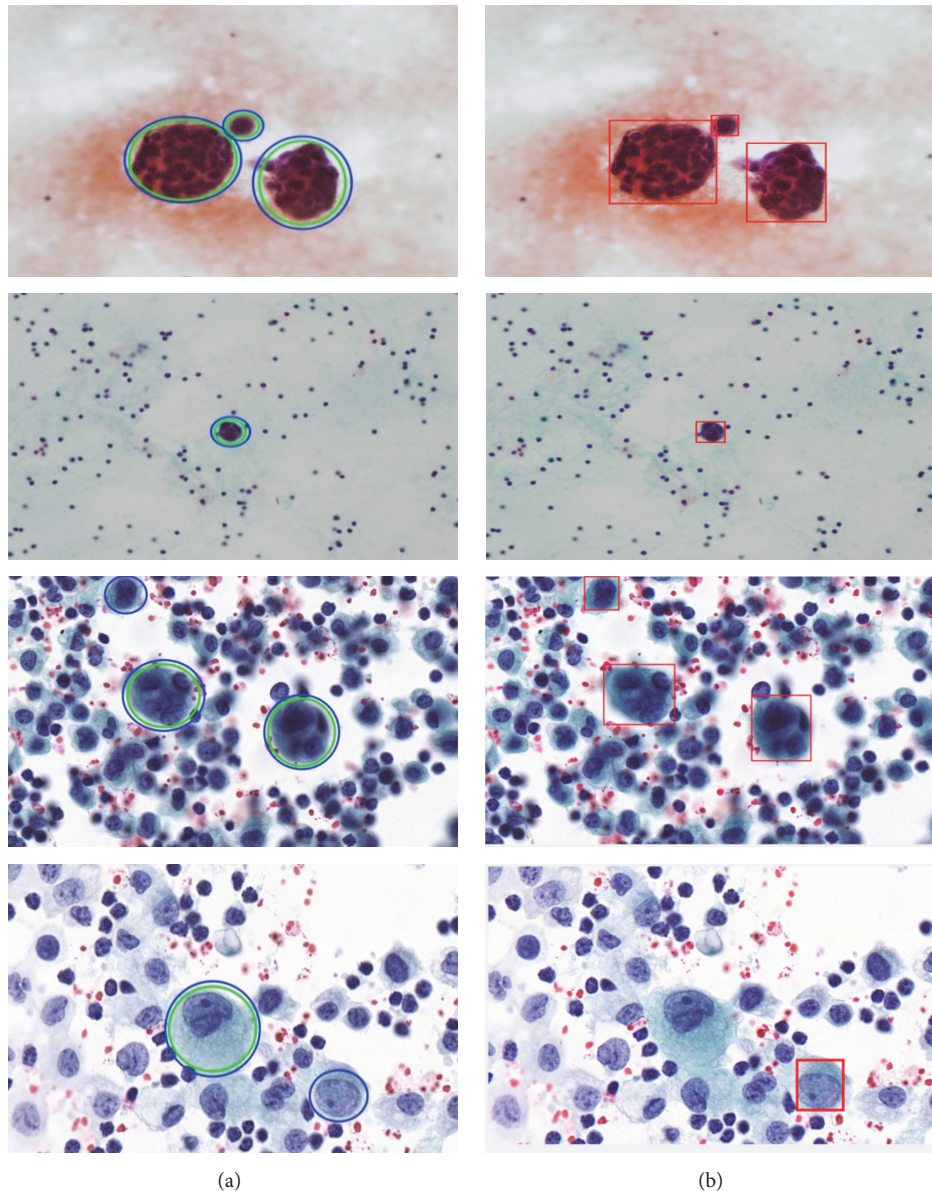


FIGURE 14: Visual demonstration of diagnostic results using the proposed CAD system to detect malignant cells in CPE images: (a) the original image with ground truth malignant cells annotated by two experts (blue and green circles represent the two experts) and (b) detected malignant cells through the proposed CAD system.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors would like to express their gratitude to ASEAN University Network/Southeast Asia Engineering Education

Development Network (AUN/SEED-Net) and Japan International Cooperation Agency (JICA) for the financial support. They also greatly acknowledge the Department of Pathology, Faculty of Medicine, Srinakharinwirot University, Thailand, for the insightful suggestions and the cooperation on the datasets. They also sincerely thank Mr. Mark O'Carroll, Office of International Affairs, KMITL, for his diligent proofreading of this paper.

References

- [1] V. S. Karkhanis and J. M. Joshi, "Pleural effusion: Diagnosis, treatment, and management," *Open Access Emergency Medicine*, vol. 4, pp. 31–52, 2012.

- [2] *Treatment for fluid on the lung (pleural effusion)*, Cancer Research UK. Available online: <https://www.cancerresearchuk.org/about-cancer/coping/physically/breathing-problems/treatment/fluid-on-the-lung-treatment> (Assessed on 10th September 2018).
- [3] E. Penz, K. N. Watt, C. A. Hergott, N. M. Rahman, and I. Psalidas, "Management of malignant pleural effusion: Challenges and solutions," *Cancer Management and Research*, vol. 9, pp. 229–241, 2017.
- [4] J. E. Heffner and J. S. Klein, "Recent advances in the diagnosis and management of malignant pleural effusions," *Mayo Clinic Proceedings*, vol. 83, no. 2, pp. 235–250, 2008.
- [5] E. S. Cibas and B. S. Ducatman, *Cytology E-Book: diagnostic principles and clinical correlates*, Elsevier Health Sciences, 2013.
- [6] F. Chen, J. Xie, H. Zhang, and D. Xia, "A technique based on wavelet and morphology transform to recognize the cancer cell in pleural effusion," in *Proceedings of the International Workshop on Medical Imaging and Augmented Reality, MIAR 2001*, pp. 199–203, Hong Kong, June 2001.
- [7] L. Zhang, Q. Wang, and J. Qi, "Research based on fuzzy algorithm of cancer cells in pleural fluid microscopic images recognition," in *Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia*, pp. 211–214, Pasadena, CA, USA, December 2006.
- [8] A. B. Tosun, O. Yergiyev, S. Kolouri, J. F. Silverman, and G. K. Rohde, "Detection of malignant mesothelioma using nuclear structure of mesothelial cells in effusion cytology specimens," *Cytometry Part A*, vol. 87, no. 4, pp. 326–333, 2015.
- [9] A. J. Mach, O. B. Adeyiga, and D. Di Carlo, "Microfluidic sample preparation for diagnostic cytopathology," *Lab on a Chip*, vol. 13, no. 6, pp. 1011–1026, 2013.
- [10] K. Raju, "Evolution of pap stain," *Biomedical Research and Therapy*, vol. 3, no. 2, pp. 490–500, 2016.
- [11] MathWorks - Image Adjustment- <https://www.mathworks.com/help/images/ref/imadjust.html>.
- [12] J. S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 1990.
- [13] E. Baykal, H. Dogan, M. Ekinici, M. E. Ercin, and S. Ersoz, "Automated cell nuclei segmentation in pleural effusion cytology using active appearance model," in *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, pp. 59–69, Springer International Publishing, Cham, Switzerland, 2017.
- [14] E. Baykal, H. Dogan, M. Ekinici, M. E. Ercin, and S. Ersoz, "Automated nuclei detection in serous effusion cytology based on machine learning," in *Proceedings of the 2017 25th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, Antalya, Turkey, May 2017.
- [15] K. Y. Win, S. Choomchuay, and K. Hamamoto, "Automated segmentation and isolation of touching cell nuclei in cytopathology smear images of pleural effusion using distance transform watershed method," in *Proceedings of the Second International Workshop on Pattern Recognition*, vol. 10443, p. 104430Q, Singapore, June 2017.
- [16] K. Y. Win, S. Choomchuay, and K. Hamamoto, "K mean clustering based automated segmentation of overlapping cell nuclei in pleural effusion cytology images," in *Proceedings of the 10th International Conference on Advanced Technologies for Communications, ATC 2017*, pp. 265–269, Viet Nam, October 2017.
- [17] K. Y. Win, S. Choomchuay, K. Hamamoto, and M. Raveesunthornkiat, "Artificial neural network based nuclei segmentation on cytology pleural effusion images," in *Proceedings of the 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, pp. 245–249, IEEE, Okinawa, November 2017.
- [18] K. Y. Win, S. Choomchuay, K. Hamamoto, and M. Raveesunthornkiat, "Comparative study on automated cell nuclei segmentation methods for cytology pleural effusion images," *Journal of Healthcare Engineering*, vol. 2018, Article ID 9240389, 14 pages, 2018.
- [19] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels," Tech. Rep., 2010, No. EPFL-REPORT-149300.
- [20] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2281, 2012.
- [21] N. Dhanachandra, K. Mangle, and Y. J. Chanu, "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm," in *Proceedings of the 11th International Conference on Communication Networks, ICCN 2015*, pp. 764–771, India, August 2015.
- [22] C. G. Rafael and E. W. Richard, *Digital Image Processing*, 3rd edition, 2008.
- [23] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [24] S. Zafari, T. Eerola, J. Sampo, H. Kälviäinen, and H. Haario, "Segmentation of partially overlapping nanoparticles using concave points," in *Proceedings of the International Symposium on Visual Computing*, pp. 187–197, Springer International Publishing, Cham, Switzerland, 2015.
- [25] K. Rodenacker and E. Bengtsson, "A feature set for cytometry on digitized microscopic images," *Analytical Cellular Pathology*, vol. 25, no. 1, pp. 1–36, 2003.
- [26] P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," *Signal Processing*, vol. 122, pp. 1–13, 2016.
- [27] P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies," *IEEE Transactions on Medical Imaging*, vol. 32, no. 12, pp. 2169–2178, 2013.
- [28] M. A. Devi, S. Ravi, J. Vaishnavi, and S. Punitha, "Classification of cervical cancer using artificial neural networks," in *Proceedings of the 12th International Conference on Communication Networks, ICCN 2016*, pp. 465–472, India, August 2016.
- [29] J. Su, X. Xu, Y. He, and J. Song, "Automatic detection of cervical cancer cells by a two-level cascade classification system," *Analytical Cellular Pathology*, vol. 2016, Article ID 9535027, 11 pages, 2016.
- [30] G. N. Srinivasan and G. Shobha, "Statistical texture analysis," in *Proceedings of World Academy of Science, Engineering and Technology*, vol. 36, pp. 1264–1269, 2008.
- [31] S. Rathore, M. Hussain, M. Aksam Iftikhar, and A. Jalil, "Ensemble classification of colon biopsy images based on information rich hybrid features," *Computers in Biology and Medicine*, vol. 47, no. 1, pp. 76–92, 2014.
- [32] A. Sengur, "Color texture classification using wavelet transform and neural network ensembles," *Arabian Journal for Science and Engineering*, vol. 34, pp. 491–502, 2009.
- [33] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

- [34] W. Gómez, W. C. A. Pereira, and A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1889–1899, 2012.
- [35] B. V. Dasarathy and E. B. Holder, "Image characterizations based on joint gray level-run length distributions," *Pattern Recognition Letters*, vol. 12, no. 8, pp. 497–502, 1991.
- [36] I. Iguyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [37] R. K. Sivagaminathan and S. Ramakrishnan, "A hybrid approach for feature subset selection using neural networks and ant colony optimization," *Expert Systems with Applications*, vol. 33, no. 1, pp. 49–60, 2007.
- [38] D. L. Tong and A. C. Schierz, "Hybrid genetic algorithm-neural network: Feature extraction for unpreprocessed microarray data," *Artificial Intelligence in Medicine*, vol. 53, no. 1, pp. 47–56, 2011.
- [39] E. Alba, J. García-Nieto, L. Jourdan, and E.-G. Talbi, "Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms," in *Proceedings of the IEEE Congress on Evolutionary Computation (CEC '07)*, pp. 284–290, IEEE, Singapore, September 2007.
- [40] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [41] S.-W. Lin, Z.-J. Lee, S.-C. Chen, and T.-Y. Tseng, "Parameter determination of support vector machine and feature selection using simulated annealing approach," *Applied Soft Computing*, vol. 8, no. 4, pp. 1505–1512, 2008.
- [42] M. Filippone, F. Masulli, and S. Rovetta, "Simulated annealing for supervised gene selection," *Soft Computing*, vol. 15, no. 8, pp. 1471–1482, 2011.
- [43] F. F. Gonzalez-Navarro and L. A. Belanche-Muñoz, "Feature selection for microarray gene expression data using simulated annealing guided by the multivariate joint entropy," *Computación y Sistemas*, vol. 18, no. 2, 2014.
- [44] B. M. Wilamowski and H. Yu, "Improved computation for levenberg-marquardt training," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 21, no. 6, pp. 930–937, 2010.
- [45] Yarpiz, <http://yarpiz.com>.
- [46] P. Yang, L. Xu, B. B. Zhou, Z. Zhang, and A. Y. Zomaya, "A particle swarm based hybrid system for imbalanced medical data sampling," *BMC Genomics*, vol. 10, no. 3, p. S34, 2009.
- [47] S. Huda, J. Yearwood, H. F. Jelinek, M. M. Hassan, G. Fortino, and M. Buckland, "A hybrid feature selection with ensemble classification for imbalanced healthcare data: A case study for brain tumor diagnosis," *IEEE Access*, vol. 4, pp. 9145–9154, 2016.
- [48] S. Nagi and D. K. Bhattacharyya, "Classification of microarray cancer data using ensemble approach," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 2, no. 3, pp. 159–173, 2013.
- [49] T. G. Dietterich, "Ensemble methods in machine learning," in *Proceedings of the International workshop on multiple classifier systems*, pp. 1–15, Springer, Heidelberg, Berlin, Germany, 2000.
- [50] A. M. Prasad, L. R. Iverson, and A. Liaw, "Newer classification and regression tree techniques: bagging and random forests for ecological prediction," *Ecosystems*, vol. 9, no. 2, pp. 181–199, 2006.
- [51] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural Network Design*, vol. 20, Pws Pub, Boston, USA, 1996.
- [52] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.
- [53] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [54] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, vol. 544 of *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*, John Wiley and Sons, New York, NY, USA, 1992.
- [55] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, Mass, USA, 1984.

Research Article

A Deep Learning Approach to Vascular Structure Segmentation in Dermoscopy Colour Images

Joanna Jaworek-Korjakowska 

Department of Automatic Control and Robotics, AGH University of Science and Technology, Cracow, Poland

Correspondence should be addressed to Joanna Jaworek-Korjakowska; jaworek@agh.edu.pl

Received 2 July 2018; Revised 3 October 2018; Accepted 14 October 2018; Published 1 November 2018

Guest Editor: Fabien Scalzo

Copyright © 2018 Joanna Jaworek-Korjakowska. 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.

Background. Atypical vascular pattern is one of the most important features by differentiating between benign and malignant pigmented skin lesions. Detection and analysis of vascular structures is a necessary initial step for skin mole assessment; it is a prerequisite step to provide an accurate outcome for the widely used 7-point checklist diagnostic algorithm. *Methods.* In this research we present a fully automated machine learning approach for segmenting vascular structures in dermoscopy colour images. The U-Net architecture is based on convolutional networks and designed for fast and precise segmentation of images. After preprocessing the images are randomly divided into 146516 patches of 64×64 pixels each. *Results.* On the independent validation dataset including 74 images our implemented method showed high segmentation accuracy. For the U-Net convolutional neural network, an average DSC of 0.84, sensitivity 0.85, and specificity 0.81 has been achieved. *Conclusion.* Vascular structures due to small size and similarity to other local structures create enormous difficulties during the segmentation and assessment process. The use of advanced segmentation methods like deep learning, especially convolutional neural networks, has the potential to improve the accuracy of advanced local structure detection.

1. Introduction

Melanoma is the deadliest form of skin cancer which develops when skin cells multiply rapidly as a consequence of mutations in their DNA caused by the sun's ultraviolet (UV) radiation (Figure 1). Melanoma is a cancer that starts in the melanocytes which are cells that make a brown pigment called melanin, which gives the skin its tan or brown colour [1]. Other names for this cancer include malignant melanoma and cutaneous melanoma. Most melanoma cells still make melanin, so melanoma tumours are usually brown or black but can appear pink, tan, or even white [1]. The introduction of dermoscopy has improved the accuracy of diagnosis of melanoma.

According to the World Health Organization about 132,000 new cases of malignant melanoma are diagnosed worldwide each year. In some parts of the world, especially in New Zealand and Australia, melanoma is becoming more common every year and has more than doubled in the past 30 years [2]. For 2018 it is predicted that 14,320 new cases of

melanoma skin cancer will be diagnosed in Australia which is estimated to be 10,4% of all new diagnosed cancer cases [3].

One of the main goals in prevention of malignant melanoma is awareness, early diagnosis and surgical excision. The introduction of dermoscopy has improved the accuracy of diagnosis of melanoma. Digital dermoscopy is currently the most used technology, although novel methods, such as confocal microscopy, show promising result. Nowadays, dermatologists work in order to define the different skin lesion types based on dermoscopic features to improve early detection (Figure 2) [4].

1.1. Motivation. Recently, the most commonly used diagnostic method is the 7-point checklist algorithm. The 7-point checklist is a diagnostic method that requires the identification of only 7 dermoscopic criteria, thus enabling even less experienced clinicians to perform the skin mole examination. One of the most important dermoscopic criteria is the atypical vascular pattern which has the second highest odds ratio of 7.42 and is among the 3 most important features that

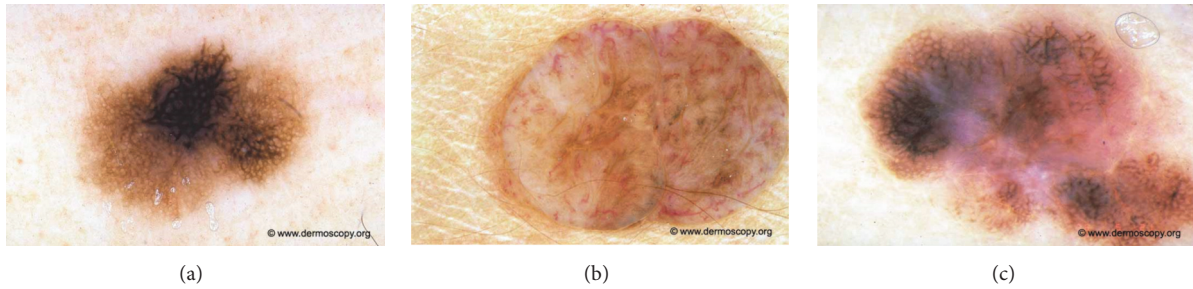


FIGURE 1: Melanocytic lesion analysis is one of the most difficult areas in modern dermatology. The challenges fall into two broad categories, namely, the recognition of rare but characteristic entities and the much more common problem of where to place an unusual lesion on the spectrum of melanocytic lesions. Examples of melanocytic lesions: (a) melanoma in situ, (b) dermal nevus, Unna type, and (c) early invasive melanoma.

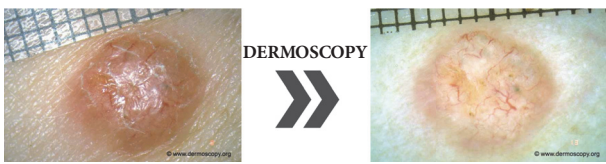


FIGURE 2: Examination of pigment mole with a dermatoscope. Diagnostic instrument commonly used for dermoscopic examination is a hand-held dermatoscope. Dermoscopy is a noninvasive method that allows evaluation of colours and microstructures of the epidermis, the dermoepidermal junction, and the papillary dermis not visible to the naked eye.

indicate the malignancy of the skin mole. Atypical vascular pattern appears as linear-irregular or dotted vessels not clearly combined with regression structures and associated with other local structures [6, 7]. Dermoscopy images show only the horizontal inspection of the skin lesion. Vascular structures that are located in parallel to the skin's surface will appear as a line, while those located vertically to the skin's surface will become visible as dots and nodes. In this respect, we observe a strong connection between the vascular structure and tumour progression and volume. Flat and superficial amelanotic/hypomelanotic melanoma and basal cell carcinoma will display different vascular structures than those of their thick or nodular counterparts (Figure 3) [8].

Based on the information above exact detection and classification of vascular structures is a crucial step in early and accurate diagnosis of malignant melanoma. In this paper we present a new and of the first approaches to the detection of vessels in dermoscopic colour images. In this study, deep learning methods (CNNs) have been used to fully automatically localize and segment vascular structure. To evaluate the performance of deep learning segmentation, we compared the outcome with the manual segmentation. Deep learning technologies have the potential to improve the accuracy and speed up the diagnosis of skin structure in clinical settings.

1.2. Related Works. To the best of our knowledge only few attempts have been made to detect vascular structures in dermoscopic colour images. Recently, Kharazmi et al. in work [9] proposed a data-driven feature learning framework based

on stacked sparse autoencoders (SSAE) for comprehensive detection of cutaneous vessels. The proposed framework demonstrated performance of 95.4% detection accuracy over a variety of vessel patterns. Betta et al. in work [10] presented a method for the identification of atypical vessels. Due to the difficulty to obtain a relevant number of epiluminescence microscopy (ELM) images with the occurrence of this local structure, the training set was constituted by pixels selected as vascular pattern in a set of images containing occurrences of this criterion. The Hue, Saturation, and Luminance components were evaluated and the frequency histograms corresponding to the three colour planes were determined. The pixel classification depended on the value of the particular HSL component. However, the authors warned that in some cases the algorithm misclassified the area, evidencing a low specificity [4, 11]. In 2014, Fabbrocini et al. proposed an automatic detection algorithm combining colour segmentation and structural analysis [12]. The skin lesion area was matched with the texture descriptors (entropy, inverse difference moment, and correlation) based on the gray level cooccurrence matrix in order to exclude texture areas. Then, a statistical analysis of the segments was performed. The system has been tested on 200 medical images and achieved 80% sensitivity, and 78% specificity. More recently Kharazmi et al. [13] proposed a new approach to vessel segmentation. Authors firstly decompose the image using independent component analysis into melanin and hemoglobin. Using k-means clustering, the hemoglobin component is clustered and a vessel mask is generated as a result of global thresholding. The segmentation sensitivity and specificity of 90% and 86% were achieved on a set of 500 000 manually segmented pixels provided by an expert. Recently, advances have been observed in retinal vessel segmentation, which is another medical area, where vessel segmentation is crucial for accurate diagnosis and early treatment. In [14] authors present the implementation of the neural network structure derived from the U-Net architecture. The algorithm obtained an AUC-ROC accuracy of 0.98.

2. Material and Methods

Artificial intelligence research has been around for more than half a century but in recent years a huge progress is observed

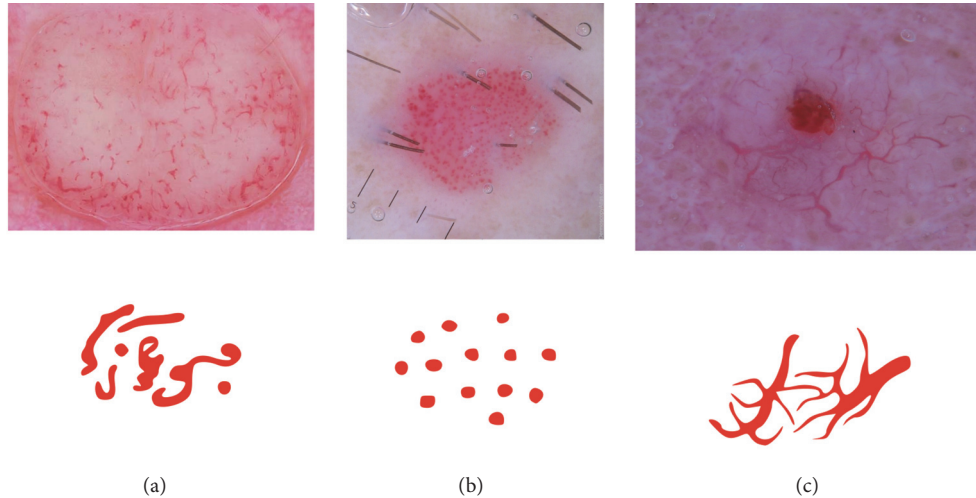


FIGURE 3: Examples of vascular patterns that can be identified in melanocytic lesions: (a) polymorphous vessels: different vascular morphologies in the same mole, (b) dotted vessels: small, reddish vessels that resemble a pinhead, and (c) arborizing vessels: large vessels that branch into finer, small vessels [5].

in widely understood machine learning [15]. Advanced statistical techniques, known as deep learning models, have been exploited with impressive results. Convolutional neural networks applied in this research learn feature representation automatically from the training data [16]. Deep learning in general and convolutional neural networks in particular have been used in variety of pattern recognition problems like retinal vessel segmentation, lung area detection, or breast cancer classification [17].

2.1. Overview. As illustrated in Figure 4 the implemented application is divided into four stages: preprocessing (image enhancement), patch extraction, training, and validation.

In this research we use the U-Net convolutional network architecture introduced by Ronneberger et al. in 2015 [18]. This method was chosen because it is one of the most promising for the addressed problem, especially, advanced image segmentation. In this section, the preprocessing step is described shortly, based on previous works, while the network architecture and training phase are presented in detail. The preprocessing and patch extraction stages have been implemented in Matlab 2017a while the neural architecture and classification process has been performed in Python using U-Net implementation as proposed by Ronneberger *et al.* and developed with TensorFlow [18, 19].

2.2. Image Preprocessing. The first step in every medical image processing system is the image acquisition, which aims at obtaining an image of the highest quality. After dermoscopic image is acquired, it may not have the expected quality to perform the diagnostic analysis. Dermoscopic images are inhomogeneous and complex and furthermore include extraneous artifacts, such as skin lines, air bubbles, and hairs, which appear in virtually every image. The preprocessing stage consists of two parts. The first step is the removal of black frame that is introduced during the digitization

process. The second step is a hair-removal algorithm which comprises two parts: hair detection and inpainting. These steps have been precisely described in our previous works [20, 21].

2.3. Patches Extraction. After preprocessing we extract N small patches $x_{m,n}$ of size 64×64 from the dermoscopy image I and the corresponding annotation G at the same position. The so-called ground-truth mask contains zeros and ones, where $(x_i, y_j) = 1$ informs that there is a vessel area at this location. Patch extraction can be performed in few different ways: nonoverlapping, overlapping, and randomly extracted. Each of these solutions has its advantages and disadvantages. To avoid the problem of class imbalance, patches have been extracted randomly around pixels pointing to vessel area both from dermoscopy image as well as from the accompanying masks. For 74 dermoscopy images with different resolution we have obtained 146516 patches. The size of the patches has been chosen experimentally and concatenated with the U-Net architecture. Larger patches require more max-pooling layers that reduce the localization accuracy, while small patches allow the network to see only little context [18].

2.4. Network Architecture. U-Net convolutional network is a popular architecture in the class of encoder-decoders, where the encoder reduces the spatial dimension of objects with pooling layers while decoder recovers the object details with upsampling layers. U-Net is a modified and extended version of fully convolutional network. Figure 5 presents the overview of a U-Net architecture.

U-Net consists of contracting path (left side) and an expansive path (right side). U-Net is like a combination of convolutional and deconvolution layers. Contracting path structurally repeats a typical 3×3 convolutional layer (unpadded convolutions) followed by a Rectified Linear Unit and a 2×2 max-pooling operation with stride 2 for down-sampling. On the expansive path, information is merged from

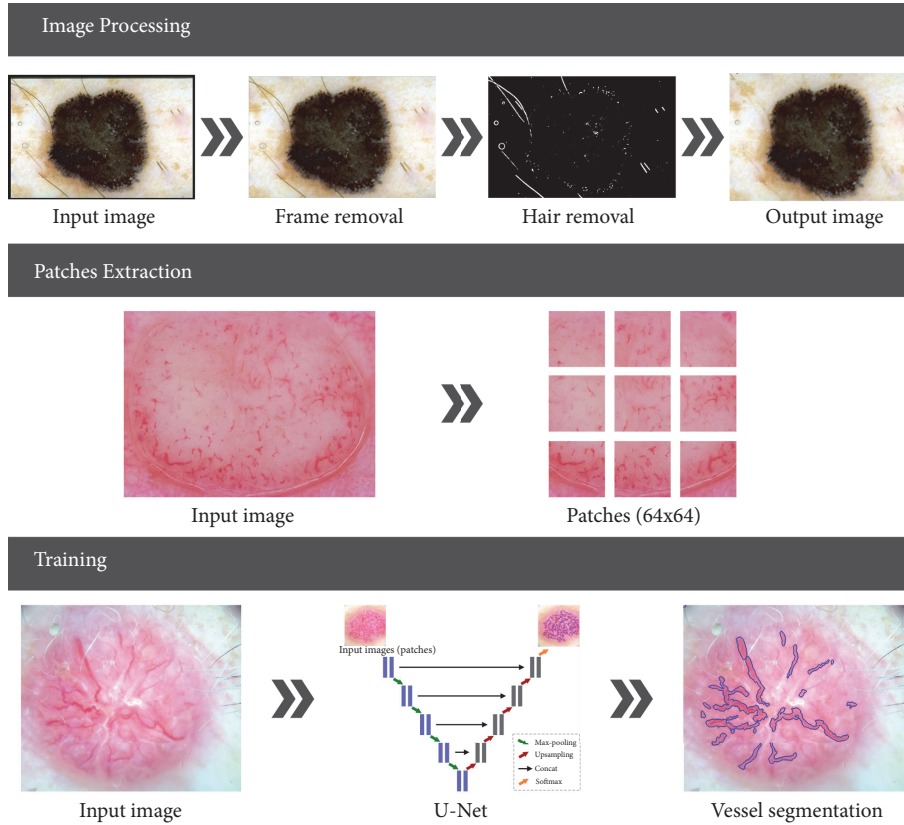


FIGURE 4: Illustration of the proposed system. (a) Image preprocessing involves black frame, hair and air bubble detection, removal or inpainting, and furthermore image normalization. (b) Patch extraction is carried out in a randomly sliding window with some overlap. (c) U-Net architecture construction and neural network training.

layers of contracting path of appropriate resolution and layers of expansive path of lower resolution, so that a whole network recognizes patterns at several scales. The final layer of a 1×1 convolution is used to map each 64-component feature vectors to the desired number of classes [18]. Input image is firstly passed through a convolutional layer with Rectified Liner Unit (ReLU) activation function:

$$f(x) = \max(0, x) \quad (1)$$

The rectifier is an activation function defined as the positive part of its argument where x is the input to a neuron. Our first layer has 64-feature map. Afterwards the max-pooling operation is applied. The convolutional layers downsample the spatial dimension from 64×64 to 8×8 . The expansive path consists of an upsampling of the feature map followed by upconvolutional and convolution layers with ReLU [17]. As we have only 2 classes (present or absent) we use the softmax classifier that calculates the cross-entropy loss for every single example:

$$L_i = -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) \quad (2)$$

The encoder and decoder comprise five layers with (64 - 128 - 256 - 512 - 1024) and (1024 - 512 - 256 - 128 - 64) filters of

size (3×3) pixels, respectively. We have chosen the Adam optimization algorithm that is an extension to stochastic gradient descent that has recently seen broader adoption for deep learning applications in computer vision and natural language processing.

3. Results

3.1. Image Database. The proposed segmentation method based on the U-Net neural network architecture has been tested on colour dermoscopic images from a widely used Interactive Atlas of Dermoscopy [6]. Images for this repository have been provided by two university hospitals (University of Naples, Italy, and University of Graz, Austria) and stored on a CD-ROM in the JPEG format. The documentation of each dermoscopic image was performed using a Dermaphot apparatus (Heine, Optotechnik, Herrsching, Germany) and a photo camera (Nikon F3) mounted on a stereomicroscope (Wild M650, Heerbrugg AG, Switzerland) in order to produce digitized ELM images of skin lesions. All the images have been assessed manually by a dermoscopic expert with an extensive clinical experience. Additionally, colour images have been used from the PH^2 database [22], and images available in online libraries [23, 24]. The database

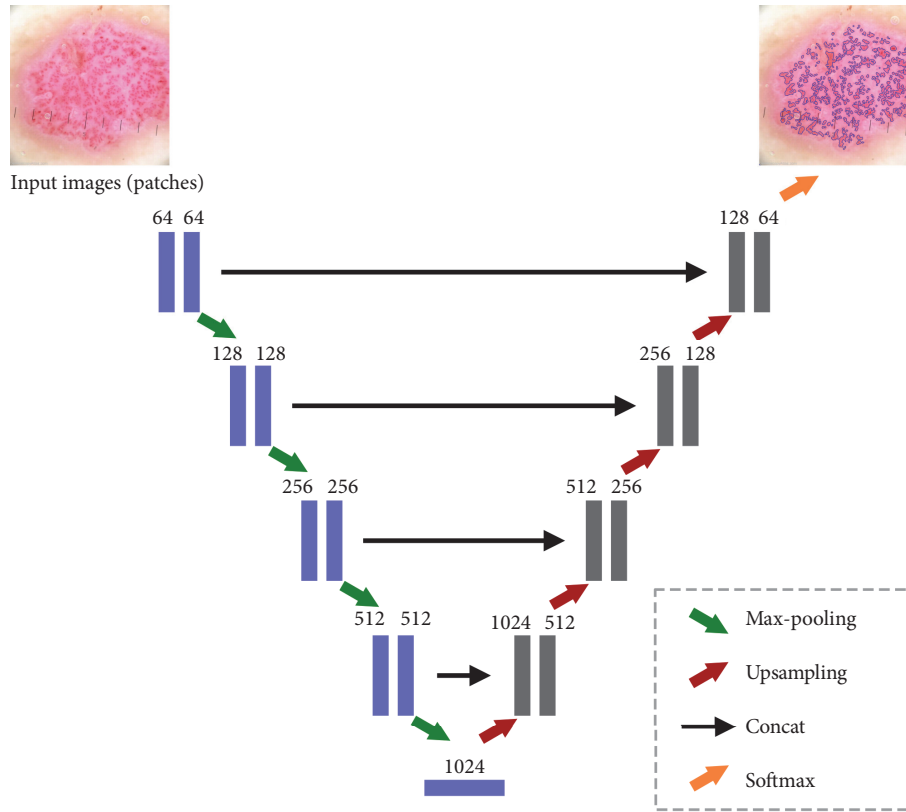


FIGURE 5: Original layers of U-Net: an encoder-decoder architecture. Each box corresponds to a multichannel feature map. The horizontal arrow denotes transfer residual information form early stage to later stage.

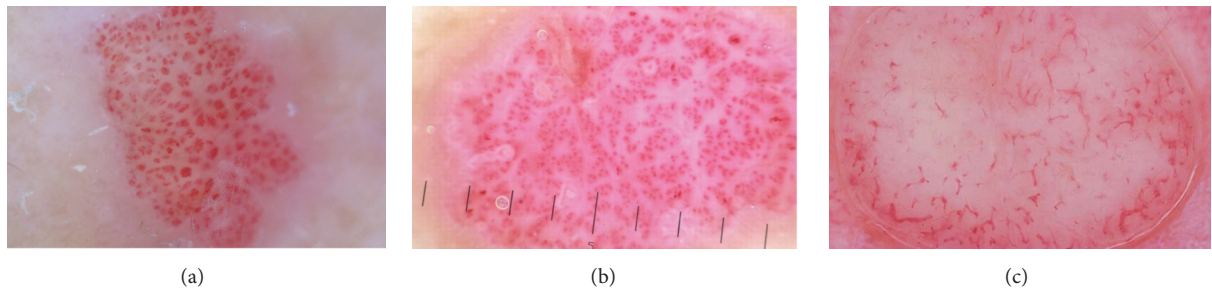


FIGURE 6: Database examples.

included 74 cases with different types of vascular pattern including linear, dotted, comma, hairpin, and polymorphous. Dermoscopy colour images have different resolutions, ranging from 0.033 to 0.5mm/pixels. Figure 6 presents samples from our database.

3.2. Deep Network (CNN) Training. The database set was divided into training set (80%) and test set (20%). The training set was used for train the U-Net, while the test set was used to analyse the training versus test error in case of overfitting. In the training stage the input image (patch) and the corresponding ground-truth mask are used to train the implemented U-Net network. The softmax layer

at the end of the network creates a probabilistic two-channel output, just like a binary segmentation problem. However, the ground truth here is a probabilistic map, not a binary segmentation map. Training is performed for 100 epochs. As training continuous (seen by epoch) we can see that the generated mask becomes more precise (Figure 7). Grayscale images in Figure 7 present predictions as a grayscale map, where light colours display values near 1, while dark colours display values near 0.

For the U-Net architecture the patch-size, the batch-size, and the weighted pixel-wise cross entropy were proved to be of high importance. The patch-size proved to be best at 64×64 pixels with a large batch-size. When the batch-size is too low,

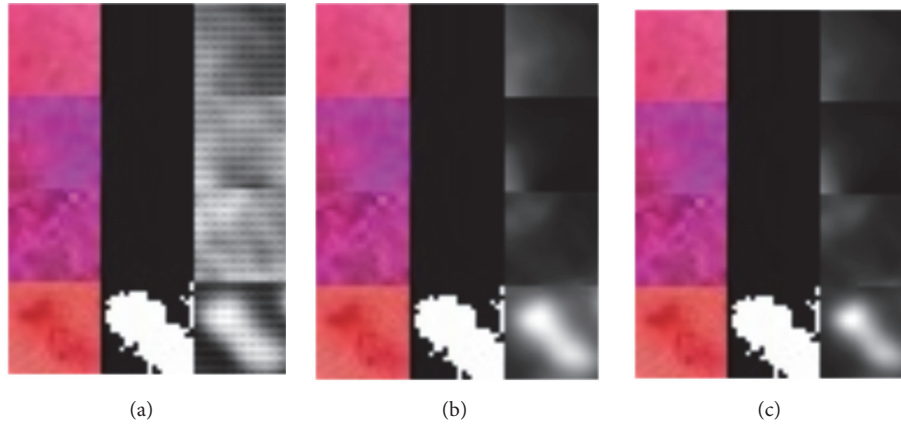


FIGURE 7: Visualization of segmentation precision over 100 epochs. Colour vascular structures in training patch slices (left), accompanying ground-truth masks (middle), and the predictions (right): (a) 10 epochs, (b) 76 epochs, and (c) 95 epochs.

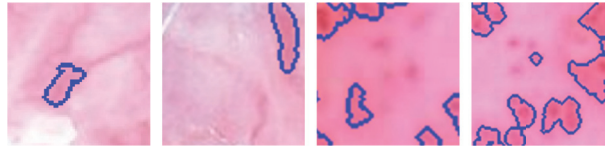


FIGURE 8: Examples of the incorrect segmentation of vascular structures. Different unmarked vascular patterns are still visible.

it might be unable to learn, thus negatively impacting total computation time. The predictions were thresholded at 0.5 and are displayed as a binary masks (Figure 8).

3.3. Analysis of CNN Segmentation Method. The performance of the U-Net neural network vascular structure segmentation approach can be assessed based on the analysis of *Sørensen index* also known as *dice similarity coefficient* which is a statistic used for comparing the similarity of two samples [25].

Given two binary sets, X and Y , the Sørensen's formula is defined as

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3)$$

Using the definition of true positive (TP), false positive (FP), and false negative (FN), it can be written as

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (4)$$

where TP denotes vascular structure pixels, FP denotes vascular structure pixels not detected, FN denotes background pixels classified as vascular structures.

The DSC is a statistical measure that calculates the degree of overlapping between the experimental segmentation and the manual segmentation and is frequently used to compare segmentations.

Furthermore, sensitivity and specificity are calculated using following equation:

$$SE = \frac{TP}{TP + FN} \quad (5)$$

$$SP = \frac{TN}{TN + FP} \quad (6)$$

The proposed algorithm achieved an average DSC of 0.84, sensitivity 0.85, and specificity 0.81. Possible values of DSC range from 0.0 to 1.0. A perfect classifier or segmentation model achieves a DSC of 1.0. The mean DSC scored 0.84 in range of 0.54-0.92. Taking into account the fact that the vascular structures have been segmented manually for the ground-truth mask the achieved DSC score is very promising. We observed that the algorithm misclassified areas which were on the boarder between skin lesion and vascular structures as well as the red surrounding between skin mole and healthy skin (Figure 8).

Figure 9 presents few segmentation results with corresponding ground-truth masks and the predictions of the previously unseen test data. The figure contains three columns. From the left to the right, each one represents the original image, the ground-truth, and the segmentation result using the generated map, respectively. Ground-truth masks and predictions have been inverted, where white represents vessel tissue and black nonvessel tissue.

Section *Related Works* presented the state of the art of previous studies concerning the segmentation of vascular patterns in dermoscopy images. In paper [13] authors presented a method that achieved sensitivity and specificity of 90% and 86%. However, it is difficult to compare these results

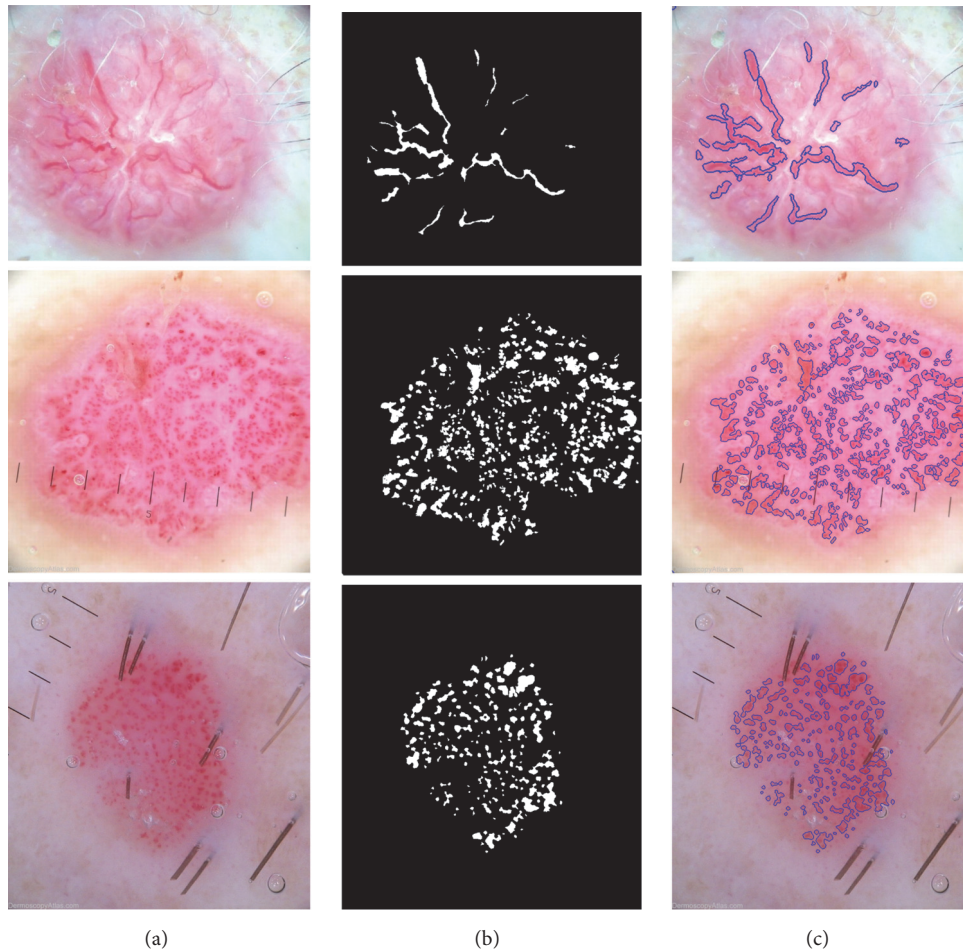


FIGURE 9: Example of dermoscopy images, accompanying segmentation masks and the predictions.

because our method has been tested on a much broader database with more complex images. For the evaluation of the advantages and correctness of the implemented algorithm, we present several outcomes with segmented structures (Figure 9).

4. Conclusions

In this research we have obtained accurate and comprehensive results showing that the applying of U-Net neural networks for local structure detection in dermoscopy colour images brings a valuable alternative to vascular structure detection. We believe that this solution can be implemented as part of a vascular pattern classification algorithm or furthermore a computer-aided diagnostic system for early detection of melanoma. Our technique shows a clear advantage over other implemented and stated in the related works section algorithms including detection accuracy, insensitive to different dermoscopy image acquisition methods.

4.1. Discussion. Starting from the described framework, further research efforts will be firstly addressed to compare and

integrate the very promising approaches reported in the most recent literature, in order to improve the neural network and optimize layers and parameters. We will also conduct a follow-up study after collecting more data with different vascular patterns. Future research will concentrate on the possibility of vascular structure classification.

Data Availability

Previously reported dermoscopy images were used to support this study and are available at 10.1016/j.jaad.2003.07.029. These prior studies (and datasets) are cited at relevant places within the text as [6].

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This scientific work was financially supported by AGH University of Science and Technology Status Funds on Decision no. 11.11.120.714.

References

- [1] The American Cancer Society medical and editorial content team, "What is melanoma skin cancer?" 2016, <https://www.cancer.org/cancer/melanoma-skin-cancer>.
- [2] AIM At Melanoma Foundation, "Understanding melanoma: About melanoma," 2017, <https://www.aimatmelanoma.org/about-melanoma/melanoma-stats-facts-and-figures/>.
- [3] Australian Institute of Health and Welfare, "Melanoma of the skin," 2017, www.aihw.gov.au/acim-books.
- [4] M. E. Celebi, J. S. Marques, and T. Mendonca, *Dermoscopy Image Analysis*, CRC Press LLC, 1st edition, 2016.
- [5] J. Martín, R. Bella-Navarro, and E. Jordá, "Vascular Patterns in Dermoscopy," *Actas Dermo-Sifiliográficas (English Edition)*, vol. 103, no. 5, pp. 357–375, 2012.
- [6] G. Argenziano, P. H. Soyer, V. D. Giorgio et al., *Interactive Atlas of Dermoscopy*, Edra Medical Publishing and New Media, 2000.
- [7] R. Marks, "Epidemiology of melanoma," *Clinical and Experimental Dermatology*, vol. 25, no. 6, pp. 459–463, 2000.
- [8] E. Ayhan, D. Ucmak, and Z. M. Akkurt, "Vascular structures in dermoscopy," *Anais Brasileiros de Dermatologia*, vol. 90, no. 4, pp. 545–553, 2015.
- [9] P. Kharazmi, J. Zheng, H. Lui, Z. Jane Wang, and T. K. Lee, "A Computer-Aided Decision Support System for Detection and Localization of Cutaneous Vasculature in Dermoscopy Images Via Deep Feature Learning," *Journal of Medical Systems*, vol. 42, no. 2, 2018.
- [10] G. Betta, G. Di Leo, G. Fabbrocini, A. Paolillo, and P. Sommella, "Dermoscopic image-analysis system: Estimation of atypical pigment network and atypical vascular pattern," in *Proceedings of the IEEE International Workshop on Medical Measurement and Applications, MeMeA 2006*, pp. 63–67, Italy, April 2006.
- [11] M. Emre Celebi, W. V. Stoecker, and R. H. Moss, "Advances in skin cancer image analysis," *Computerized Medical Imaging and Graphics*, vol. 35, no. 2, pp. 83–84, 2011.
- [12] G. Fabbrocini, V. D. Vita, S. Cacciapuoti et al., "Automatic diagnosis of melanoma based on the 7-point checklist," in *Computer Vision Techniques for the Diagnosis of Skin Cancer*, J. Scharcanski and M. E. Celebi, Eds., Series in BioEngineering, pp. 71–107, Springer, Berlin, Heidelberg, 2014.
- [13] P. Kharazmi, M. I. Aljasser, H. Lui, Z. J. Wang, and T. K. Lee, "Automated Detection and Segmentation of Vascular Structures of Skin Lesions Seen in Dermoscopy, with an Application to Basal Cell Carcinoma Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1675–1684, 2017.
- [14] D. Cortinovis, *Retina blood vessel segmentation with a convolution neural network (U-net)*, 2016, <https://github.com/qtim-lab/retinaunet>.
- [15] L. Ogiela, R. Tadeusiewicz, and M. R. Ogiela, "Cognitive computing in intelligent medical pattern recognition systems," *Lecture Notes in Control and Information Sciences*, vol. 344, pp. 851–856, 2006.
- [16] S. Trebeschi, J. J. M. Van Griethuysen, D. M. J. Lambregts et al., "Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR," *Scientific Reports*, vol. 7, no. 1, 2017.
- [17] B. Ait Skourt, A. El Hassani, and A. Majda, "Lung CT Image Segmentation Using Deep Neural Networks," *Procedia Computer Science*, vol. 127, pp. 109–113, 2018.
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention and Medical Image Computing and Computer-Assisted Intervention (MICCAI '15)*, vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241, November 2015.
- [19] J. Akeret, C. Chang, A. Lucchi, and A. Refregier, "Radio frequency interference mitigation using deep convolutional neural networks," *Astronomy and Computing*, vol. 18, pp. 35–39, 2017.
- [20] J. Jaworek-Korjakowska and P. Kleczek, "Automatic classification of specific melanocytic lesions using artificial intelligence," *BioMed Research International*, vol. 2016, Article ID 8934242, 17 pages, 2016.
- [21] J. Jaworek-Korjakowska, "Computer-aided diagnosis of micro-malignant melanoma lesions applying support vector machines," *BioMed Research International*, vol. 2016, Article ID 4381972, 8 pages, 2016.
- [22] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH2 - A dermoscopic image database for research and benchmarking," in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2013*, pp. 5437–5440, Japan, July 2013.
- [23] The Skin Cancer Society of Australia, "Dermoscopy atlas," 2007, <http://www.dermoscopyatlas.com/>.
- [24] E. Ehrsam, *Dermoscopy*, 2016, <http://dermoscopic.blogspot.com>.
- [25] T. Sorensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons," *Kongelige Danske Videnskabernes Selskab*, vol. 5, pp. 1–34, 1948.

Research Article

PCG Classification Using Multidomain Features and SVM Classifier

Hong Tang ¹, Ziyin Dai ¹, Yuanlin Jiang ¹, Ting Li,² and Chengyu Liu ³

¹Department of Biomedical Engineering, Dalian University of Technology, Dalian, China

²College of Information and Communication Engineering, Dalian Minzu University, Dalian, China

³School of Instrument Science and Engineering, Southeast University, Nanjing, China

Correspondence should be addressed to Hong Tang; tanghong@dlut.edu.cn and Chengyu Liu; bestlcy@sdu.edu.cn

Received 20 March 2018; Accepted 12 June 2018; Published 9 July 2018

Academic Editor: Magdalena Kasprowicz

Copyright © 2018 Hong Tang 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.

This paper proposes a method using multidomain features and support vector machine (SVM) for classifying normal and abnormal heart sound recordings. The database was provided by the PhysioNet/CinC Challenge 2016. A total of 515 features are extracted from nine feature domains, i.e., time interval, frequency spectrum of states, state amplitude, energy, frequency spectrum of records, cepstrum, cyclostationarity, high-order statistics, and entropy. Correlation analysis is conducted to quantify the feature discrimination abilities, and the results show that “frequency spectrum of state”, “energy”, and “entropy” are top domains to contribute effective features. A SVM with radial basis kernel function was trained for signal quality estimation and classification. The SVM classifier is independently trained and tested by many groups of top features. It shows the average of sensitivity, specificity, and overall score are high up to 0.88, 0.87, and 0.88, respectively, when top 400 features are used. This score is competitive to the best previous scores. The classifier has very good performance with even small number of top features for training and it has stable output regardless of randomly selected features for training. These simulations demonstrate that the proposed features and SVM classifier are jointly powerful for classifying heart sound recordings.

1. Introduction

Heart sounds are a series of mechanical vibrations produced by the interplay between blood flow and heart chambers, valves, great vessels, etc. [1–3]. Heart sounds provide important initial clues in heart disease evaluation for further diagnostic examination [4]. Listening to heart sounds plays an important role in early detection for cardiovascular diseases. It is practically attractive to develop computer-based heart sound analysis. Automatic classification of pathology in heart sounds is one of the hot problems in the past 50 years. But accurate classification is still an open challenge question. To the authors' knowledge, Gerbarg et al. were the first to publish automatic classification of pathology in heart sounds [5].

Automatic classification of PCG recording in clinical application typically consists of four steps: preprocessing, segmentation, feature extraction, and classification. Over the past decades, features and methods for the classification have been widely studied. In summary, features may be

wavelet features, time-domain features, frequency domain features, complexity-based features, and joint time-frequency domain features. Methods available for classification may be artificial neural network [6–10], support vector machine [11, 12], and clustering [13–16]. Unfortunately, comparisons between previous methods have been hindered by the lack of standardized database of heart sound recordings collected from a variety of healthy and pathological conditions. The organizers of the PhysioNet/CinC Challenge 2016 set up a large collection of recordings from various research groups in the world. In the conference, many methods were proposed for this discrimination purpose, like deep learning methods [17–19], tensor based methods [20], support vector machine based methods [21, 22], and others [23–27]. Generally, time and/or frequency domain features were used in these papers. The reported top overall scores were 89.2% by [27], 86.2% by [28], 85.9% by [29], and 85.2% by [30]. In this paper, the authors extend their previous study [31] and extracted a total of 515 features for normal/abnormal PCG classification. The

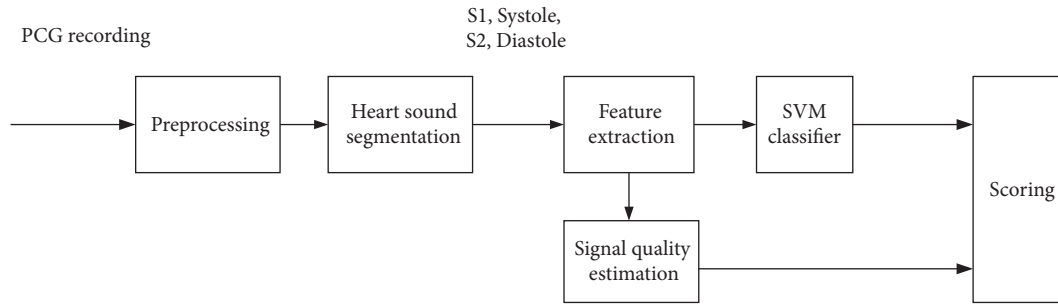


FIGURE 1: Flow diagram of the proposed classification.

difference of the proposed method to the existing methods is that these features are from multidomains, such as time interval, state amplitude, energy, high-order statistics, cepstrum, frequency spectrum, cyclostationarity, and entropy. To the authors' knowledge, the proposed method in this paper perhaps uses the most number of features. Correlation analysis shows the contribution of each feature. A SVM classifier is used to discriminate abnormal/uncertain/normal types. Cross validation shows that the proposed features have excellent generation ability. The mean overall score based on 20% data training is up to 0.84. It rises to 0.87 based on 50% data training and rises to 0.88 based on 90% data training. The results demonstrate that the method is competitive comparison to previous approaches.

2. Methods

2.1. Database. The database used in this paper is provided by the international competition PhysioNet/CinC Challenge 2016, which can be freely downloaded from the website [32]. The database includes both PCG recordings of healthy subjects and pathological patients collected in either clinical or nonclinical environments. There are a total of 3,153 heart sound recordings, given as “*.wav” format, from 764 subjects/patients, lasting from 5 s to 120 s. The recordings were divided into two classes: normal and abnormal records with a confirmed cardiac diagnosis. Label “1” was used to present abnormal (665 recordings) and “-1” to present normal cases (2488 recordings). A skilled cardiologist was also invited to evaluate the signal quality for each recording. As he believed that a recording had good signal quality, it was labeled “1”. Otherwise, it was labeled “0”. There are 279 recordings which were labeled as bad signal quality and the rest of 2874 were labeled as good quality. Details about the database can be found in [33].

2.2. Flow Diagram of the Proposed Method. The flow diagram of the proposed method to classify PCG recordings is shown in Figure 1. Each step will be described in the following subsections.

2.3. Preprocessing. Each PCG recording is high-pass filtered with a cut-off frequency of 10 Hz to remove baseline drift. The

spike removal algorithm is applied to the filtered recording [34]. Then, the recording is normalized to zero mean and unit standard deviation.

2.4. Heart Sound Segmentation by Springer's Algorithm. Springer's hidden semi-Markov model (HSMM) segmentation method [35] is used to segment a PCG recording into four states, i.e., S1, systole, S2, and diastole. Figure 2 shows an example of this segmentation. Hence, the following signals can be defined and further used for feature extraction. $x(n)$ is a digital PCG recording where n is the discrete time index. $s_{1i}(n)$ and $s_{2i}(n)$ are S1 and S2 signals occurring in the i th cardiac cycle, respectively. $sys_i(n)$ and $dia_i(n)$ are the signals of systolic interval and diastole interval in the i th cardiac cycle, respectively. $c_i(n)$ is the signal of the i th cardiac cycle. Hence, $c_i(n)$ consists of the digital sequence of $[s_{1i}(n)sys_i(n)s_{2i}(n)dia_i(n)]$.

2.5. Features Extracted in Multidomains

2.5.1. Time-Domain Features (20 Features). After the segmentation operation, a PCG recording is divided into many states in the order of S1, systole, S2, and diastole. The time interval of each state can be measured by the time difference between the beginning and the end. The cardiac cycle period can be measured by the time difference between the beginnings of two adjacent S1s. Since the intervals have physiological meanings in view point of heart physiology, Liu et al. [33] proposed 16 features from the intervals as shown in Table 1. Another 4 features from time-domain intervals are added in this paper.

2.5.2. Frequency Domain Features for States (77×4=308 Features). Frequency spectrum is estimated for the S1 state of each cardiac cycle using a Gaussian window and discrete Fourier transform. The mean frequency spectrum over cycles can be further computed. The spectrum magnitudes from 30 Hz to 790 Hz with 10 Hz interval are taken as features. The maximum frequency 790 Hz is considered to adapt possible murmurs. So, 77 features for S1 state are obtained. Similar operation is done to S2, systole, and diastole state. So, the total number of features obtained from frequency domain for states is 77 features × 4 = 308 features.

TABLE 1: Summary of time-domain features.

Feature index	Feature name	Physical meaning
1	m_RR	mean value of RR intervals
2	sd_RR	standard deviation (SD) of RR intervals
3	m_IntS1	mean value of S1 intervals
4	sd_IntS1	SD of S1 intervals
5	m_IntS2	mean value of S2 intervals
6	sd_IntS2	SD of S2 intervals
7	m_IntSys	mean value of systolic intervals
8	sd_IntSys	SD of systolic intervals
9	m_IntDia	mean value of diastolic intervals
10	sd_IntDia	SD of diastolic intervals
11	m_Ratio_SysRR	mean value of the ratio of systolic interval to RR interval of each heart beat
12	sd_Ratio_SysRR	SD of the ratio of systolic interval to RR interval of each heart beat
13	m_Ratio_DiaRR	mean value of the ratio of diastolic interval to RR interval of each heart beat
14	sd_Ratio_DiaRR	SD of the ratio of diastolic interval to RR interval of each heart beat
15	m_Ratio_SysDia	mean value of the ratio of systolic to diastolic interval of each heart beat
16	sd_Ratio_SysDia	SD of the ratio of systolic to diastolic interval of each heart beat
17	m_Ratio_S1RR*	mean value of the ratio of S1 interval to RR interval of each heart beat
18	sd_Ratio_S1RR*	SD of the ratio of S1 interval to RR interval of each heart beat
19	m_Ratio_S2RR*	mean value of the ratio of S2 interval to RR interval of each heart beat
20	sd_Ratio_S2RR*	SD of the ratio of S2 interval to RR interval of each heart beat

Note: * means the new added features in this study.

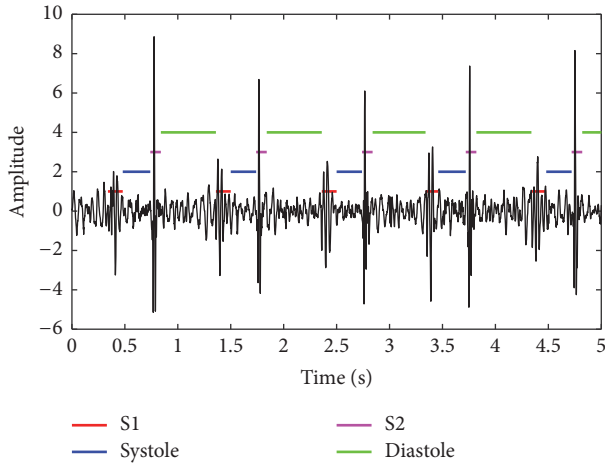


FIGURE 2: Illustration of the HSMM segmentation.

2.5.3. Normalized Amplitude Features (12 Features). Previous physiological findings in amplitude of heart sound [1–3] disclosed that the amplitude is related to heart hemodynamics. So, it is reasonable to extract features from amplitude of heart sounds. To eliminate the difference between subjects and records, no absolute amplitude is considered. Relative ratios of amplitude between states are extracted as given in Table 2.

2.5.4. Energy-Domain Features (47 Features). The features in energy domain consist of two parts: the energy ratio of a

band-pass signal to the original one and the energy ratio of one state to another.

For the first part, various frequency bands are considered with initial value of 10 Hz and increment bandwidth of 30 Hz; i.e., the 27 frequency bands are [10 40] Hz, [40 70] Hz, [70 100] Hz, . . . , and [790 820] Hz, respectively. The previous studies disclosed that murmurs' frequency is hardly higher than 800 Hz. In order to reflect murmurs' properties, the maximum frequency considered in this domain is 820 Hz. In this paper, each band-pass filter is designed by a five-order Butterworth filter. The output of the i th filter is y_i :

$$y_i = \text{filter}(\mathbf{b}_i, \mathbf{a}_i, \mathbf{x}), \quad i = 1, \dots, 27, \quad (1)$$

where \mathbf{b}_i (numerator) and \mathbf{a}_i (denominator) are the Butterworth IIR filter coefficient vectors. Hence, the energy ratio is defined as

$$\text{Ratio_band_energy}_i = \frac{\sum |y_i|^2}{\sum |\mathbf{x}|^2}, \quad i = 1, \dots, 27. \quad (2)$$

It is known that a normal heart sound signal generally has a frequency band below 200 Hz. However, the frequency band may extend to 800 Hz if it contains murmurs. So, the energy ratio reflects signal energy distribution along frequency band. These features are helpful to discriminate a PCG records with murmurs or not.

TABLE 2: Summary of normalized amplitude features.

Feature index	Feature name	Physical meaning
1	m_Amp_SysS1	mean value of the ratio of the mean absolute amplitude during systole to that during S1 in each heart beat
2	sd_Amp_SysS1	SD of m_Amp_SysS1
3	m_Amp_DiaS2	mean value of the ratio of the mean absolute amplitude during diastole to that during S2 in each heart beat
4	sd_Amp_DiaS2	SD of m_Amp_DiaS2
5	m_Amp_S1S2	mean value of the ratio of the mean absolute amplitude during S1 to that during S2 in each heart beat
6	sd_Amp_S1S2	SD of m_Amp_S1S2
7	m_Amp_S1Dia	mean value of the ratio of the mean absolute amplitude during S1 to that during diastole in each heart beat
8	sd_Amp_S1Dia	SD of m_Amp_S1Dia
9	m_Amp_SysDia	mean value of the ratio of the mean absolute amplitude during systole to that during diastole in each heart beat
10	sd_Amp_SysDia	SD of m_Amp_SysDia
11	m_Amp_S2Sys	mean value of the ratio of the mean absolute amplitude during S2 to that during systole in each heart beat
12	sd_Amp_S2Sys	SD of m_Amp_S2Sys

For the second part, the relative energy ratio is investigated between any two states, resulting in another 20 features. The energy ratio of S1 to the cycle period is defined as

$$Ratio_state_energy_{S1_cycle} = \frac{\sum_n |s_{i_i}(n)|^2}{\sum_n |c_i(n)|^2}, \quad (3)$$

$$i = 1, \dots, N,$$

where N is the number of cycles in a PCG recording and n is the discrete time index. The authors consider average of $Ratio_state_energy_{S1_cycle}$ and standard deviation of $Ratio_state_energy_{S1_cycle}$ as two features. Similarly, another 18 features can be obtained from the averages and the standard deviations. The 47 proposed features in this domain are listed in Table 3.

2.5.5. Spectrum-Domain for Records Features (27 Features). As is mentioned in Section 2.5.4, the frequency band is divided into 27 bands with start from 10 Hz to 30 Hz increment. Fast Fourier transform is performed for every record. The ratio of spectrum magnitude sum in a band to spectrum magnitude sum in whole band is taken as a feature. So, 27 features can be produced for a record. These features can discriminate murmurs because murmurs generally have higher frequency than those of normal heart sounds.

2.5.6. Cepstrum-Domain Features (13 Features \times 5 = 65 Features). The cepstrum of a PCG recording is calculated and the first 13 cepstral coefficients are taken as features [36]. Additionally, all S1 states from a PCG recording are joined together to create a new digital sequence. Then, the cepstrum can be calculated and the first 13 cepstral coefficients are taken as features. Similarly, the same operation is done to S2, systole, and diastole states. So, another 13 features \times 3 = 39 features

are obtained. The cepstrum of a signal $p(n)$ is computed as follows:

$$P(k) = DFT[p(n)], \quad (4)$$

$$\hat{P}(k) = \log[P(k)], \quad (5)$$

$$\hat{p}(n) = |IDFT[\hat{P}(k)]|, \quad (6)$$

where the operator $DFT[\cdot]$ is the discrete Fourier transform, $IDFT[\cdot]$ is the inverse DFT, $\log[\cdot]$ is the natural logarithm, and $|\cdot|$ is the absolute operation. It is known that the cepstrum coefficient decays quickly. So, it is reasonable to select the first 13 coefficients as features. The cepstrum-domain features are listed in Table 4.

2.5.7. Cyclostationary Features (4 Features). (1) $m_cyclostationarity_1$ is mean value of the degree of cyclostationarity. The definition of “degree of cyclostationarity” can be found in [37]. This feature indicates the degree of a signal repetition. It will be infinite if the events which occurred in heart beating were exactly periodic. However, it will be a small number if the events are randomly alike. Let us assume $\gamma(\alpha)$ is the cycle frequency spectral density (CFSD) of a heart sound signal at cycle frequency α , as shown in Figure 3. This feature is defined as

$$d(\eta) = \frac{\gamma(\eta)}{\int_0^\beta \gamma(\alpha) d\alpha}, \quad (7)$$

where β is the maximum cycle frequency considered and η is the basic cycle frequency indicated by the main peak location of $\gamma(\alpha)$. A heart sound signal is equally divided into subsequences. The feature can be estimated for each subsequence; then the mean value and standard deviation can be obtained.

TABLE 3: Summary of energy-domain features.

Feature index	Feature name	Physical meaning
1-27	$Ratio_band_energy_i$	Ratio of a given band energy to the total energy
28	$m_Ratio_state_energy_{S1_cycle}$	Mean of $Ratio_state_energy_{S1_cycle}$
29	$SD_Ratio_state_energy_{S1_cycle}$	standard deviation of $Ratio_state_energy_{S1_cycle}$
30	$m_Ratio_state_energy_{S2_cycle}$	Mean of $Ratio_state_energy_{S2_cycle}$
31	$SD_Ratio_state_energy_{S2_cycle}$	standard deviation of $Ratio_state_energy_{S2_cycle}$
32	$m_Ratio_state_energy_{sys_cycle}$	Mean of $Ratio_state_energy_{sys_cycle}$
33	$SD_Ratio_state_energy_{sys_cycle}$	standard deviation of $Ratio_state_energy_{sys_cycle}$
34	$m_Ratio_state_energy_{dia_cycle}$	Mean of $Ratio_state_energy_{dia_cycle}$
35	$SD_Ratio_state_energy_{dia_cycle}$	standard deviation of $Ratio_state_energy_{dia_cycle}$
36	$m_Ratio_state_energy_{S1_S2}$	Mean of $Ratio_state_energy_{S1_S2}$
37	$SD_Ratio_state_energy_{S1_S2}$	standard deviation of $Ratio_state_energy_{S1_S2}$
38	$m_Ratio_state_energy_{S1_sys}$	Mean of $Ratio_state_energy_{S1_sys}$
39	$SD_Ratio_state_energy_{S1_sys}$	standard deviation of $Ratio_state_energy_{S1_sys}$
40	$m_Ratio_state_energy_{S1_dia}$	Mean of $Ratio_state_energy_{S1_dia}$
41	$SD_Ratio_state_energy_{S1_dia}$	standard deviation of $Ratio_state_energy_{S1_dia}$
42	$m_Ratio_state_energy_{S2_sys}$	Mean of $Ratio_state_energy_{S2_sys}$
43	$SD_Ratio_state_energy_{S2_sys}$	standard deviation of $Ratio_state_energy_{S2_sys}$
44	$m_Ratio_state_energy_{S2_dia}$	Mean of $Ratio_state_energy_{S2_dia}$
45	$SD_Ratio_state_energy_{S2_dia}$	standard deviation of $Ratio_state_energy_{S2_dia}$
46	$m_Ratio_state_energy_{dia_sys}$	Mean of $Ratio_state_energy_{dia_sys}$
47	$SD_Ratio_state_energy_{dia_sys}$	standard deviation of $Ratio_state_energy_{dia_sys}$

TABLE 4: Summary of cepstrum-domain features.

Feature index	Feature name	Physical meaning
1-13	Cepstrum coefficients	Cepstrum coefficients of a PCG recording
14-26	Cepstrum coefficients	Cepstrum coefficients of jointed S1 state
27-39	Cepstrum coefficients	Cepstrum coefficients of jointed systolic state
40-52	Cepstrum coefficients	Cepstrum coefficients of jointed S2 state
53-65	Cepstrum coefficients	Cepstrum coefficients of jointed diastole state

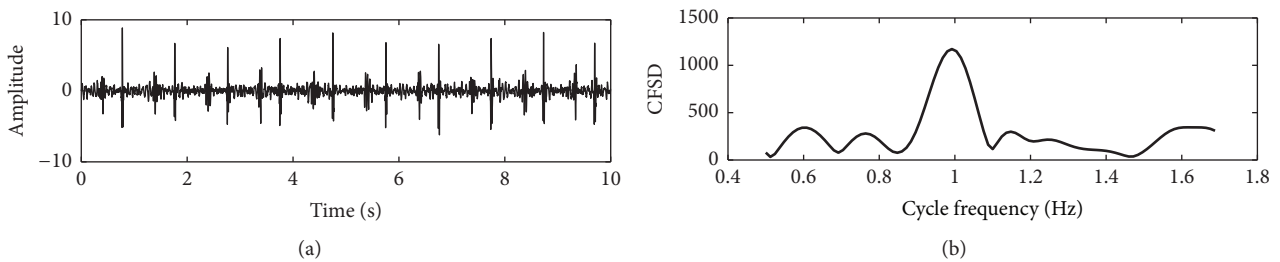


FIGURE 3: An example of cycle frequency spectral density. (a) A subsequence of a PCG recording and (b) cycle frequency spectral density of the subsequence.

(2) $sd_cyclostationarity_1$ is SD of the degree of cyclostationarity.

(3) $m_cyclostationarity_2$ is mean value of the sharpness measure. The definition of this indicator is the sharpness of the peak of cycle frequency spectral density. It is

$$peak_sharpness = \frac{\max(\gamma(\alpha))}{\text{median}(\gamma(\alpha))}. \quad (8)$$

The operators $\max(.)$ and $\text{median}(.)$ are the maximum and median magnitude of the cycle frequency spectral density. It is obvious that the sharper the peak is, the greater the feature is. Similarly, the feature can be calculated for each subsequence of the heart sound signal and then get the mean value and SD.

(4) $sd_cyclostationarity_2$ is SD of the sharpness measure. The four features are listed in Table 5.

TABLE 5: Summary of cyclostationary features.

Feature index	Feature name	Physical Meaning
1	m_cyclostationarity_1	mean value of the degree of cyclostationarity
2	sd_cyclostationarity_1	SD of the degree of cyclostationarity
3	m_cyclostationarity_2	mean value of the sharpness measure
4	sd_cyclostationarity_2	SD of the sharpness measure

TABLE 6: Summary of high-order statistics features.

Feature index	Feature name	Physical Meaning
1	m_S1_skewness	mean value of the skewness of S1
2	sd_S1_skewness	SD of the skewness of S1
3	m_S1_kurtosis	mean value of the kurtosis of S1
4	sd_S1_kurtosis	SD of the kurtosis of S1
5	m_S2_skewness	mean value of the skewness of S2
6	sd_S2_skewness	SD of the skewness of S2
7	m_S2_kurtosis	mean value of the kurtosis of S2
8	sd_S2_kurtosis	SD of the kurtosis of S2
9	m_sys_skewness	mean value of the skewness of systole
10	sd_sys_skewness	SD of the skewness of systole
11	m_sys_kurtosis	mean value of the kurtosis of systole
12	sd_sys_kurtosis	SD of the kurtosis of systole
13	m_dia_skewness	mean value of the skewness of diastole
14	sd_dia_skewness	SD of the skewness of diastole
15	m_dia_kurtosis	mean value of the kurtosis of diastole
16	sd_dia_kurtosis	SD of the kurtosis of diastole

2.5.8. *High-Order Statistics Features (16 Features)*. In probability theory and statistics, skewness is a measure of the asymmetry of the probability distribution of real-valued random numbers about its mean. It is a three-order statistics. Kurtosis is a measure of “tailedness” of the probability distribution of real-valued random numbers. It is a four-order statistics. The skewness and kurtosis of each state are considered here. There are sixteen related features, as listed in Table 6.

2.5.9. *Entropy Features (16 Features)*. Sample entropy (SampEn) and fuzzy measure entropy (FuzzyMEn) have the ability to measure the complexity of a random sequence [38, 39]. Sample entropy and fuzzy measure entropy are both computed to measure the complexity of every state segmented by Springer’s algorithm. Then, the average and standard deviation are used as the features. The detailed algorithm to calculate sample entropy and fuzzy measure entropy can be found in [38, 39]. So, 16 features in entropy are listed in Table 7.

2.5.10. *Summary*. This paper considers 515 features in nine domains. They are listed in Table 8 for reference. To the authors’ knowledge, the features extracted from entropy and cyclostationarity are new for heart sound classification. On the other hand, the combination of the features in the nine

domains is novel for this classification. Seldom previous study has considered so many features simultaneously.

2.6. *SVM-Based Model for Signal Quality Estimation and Classification*. The signal quality classification is typically two-category classification problem in this study. The SVM-based model has yielded excellent results in many two-class classification situations. Given a training sample set $\{\mathbf{x}_i, y_i\}$, $i = 1, \dots, K$, where \mathbf{x}_i is the feature vector $\mathbf{x}_i \in R^d$, y_i is the label. So, SVM-based model is applicable for both signal quality estimation and classification. For the quality estimation, the label is $y_i \in \{1, 0\}$, which means good quality and bad quality. For the classification, the label is $y_i \in \{1, -1\}$, which means abnormal and normal cases. The aim of SVM classifier is to develop optimal hyperplane between two classes besides distinguishing them. The optimal hyperplane can also be constructed by calculating the following optimization problem.

$$\min \phi(\mathbf{w}) = \frac{1}{2} (\mathbf{w}^T \mathbf{w}) + C \sum_{i=1}^K \xi_i \quad (9)$$

$$\text{subject to } y_i ((\mathbf{w}^T \varphi(\mathbf{x}_i)) + b) \geq 1, \quad i = 1, \dots, K.$$

Here ξ_i is a relaxation variable and $\xi_i \geq 0$, C is a penalty factor, and \mathbf{w} is the coefficient vector. $\varphi(\mathbf{x}_i)$ is introduced to get a

TABLE 7: Summary of entropy features.

Feature index	Feature name	Physical meaning
1	<i>m_SampEn_S1</i>	Mean value of SampEn of S1 state
2	<i>SD_SampEn_S1</i>	SD value of SampEn of S1 state
3	<i>m_SampEn_S2</i>	Mean value of SampEn of S2 state
4	<i>SD_SampEn_S2</i>	SD value of SampEn of S2 state
5	<i>m_SampEn_sys</i>	Mean value of SampEn of systolic state
6	<i>SD_SampEn_sys</i>	SD value of SampEn of systolic state
7	<i>m_SampEn_dia</i>	Mean value of SampEn of diastolic state
8	<i>SD_SampEn_dia</i>	SD value of SampEn of diastolic state
9	<i>m_FuzzyMEn_S1</i>	Mean value of FuzzyMEn of S1 state
10	<i>SD_FuzzyMEn_S1</i>	SD value of FuzzyMEn of S1 state
11	<i>m_FuzzyMEn_S2</i>	Mean value of FuzzyMEn of S2 state
12	<i>SD_FuzzyMEn_S2</i>	SD value of FuzzyMEn of S2 state
13	<i>m_FuzzyMEn_sys</i>	Mean value of FuzzyMEn of systolic state
14	<i>SD_FuzzyMEn_sys</i>	SD value of FuzzyMEn of systolic state
15	<i>m_FuzzyMEn_dia</i>	Mean value of FuzzyMEn of diastolic state
16	<i>SD_FuzzyMEn_dia</i>	SD value of FuzzyMEn of diastolic state

TABLE 8: Summary of the proposed features.

Index	Domain	Num. of features	Motivation
1	Time interval	20	The time interval of each state has physiological meaning based on heart physiology.
2	Frequency spectrum of state	308	To reflect the frequency spectrum within state.
3	State amplitude	12	The amplitude is related to the heart hemodynamics.
4	Energy	47	To reflect energy distribution with respect to frequency band
5	Frequency spectrum of records	27	To reflect frequency spectrum within records
6	Cepstrum	65	To reflect the acoustic properties.
7	Cyclostationary	4	To reflect the degree of signal repetition.
8	High-order statistics	16	To reflect the skewness and kurtosis of each signal state.
9	Entropy	16	To reflect the PCG signal inherent complexity.
Total	--	515	--

nonlinear support vector machine. The optimization problem can be equally transformed into

$$\max L(\alpha) = \sum_{i=1}^K \alpha_i - \frac{1}{2} \sum_{i,j=1}^K \alpha_i \alpha_j y_i y_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \tag{10}$$

$$\text{subject to } \sum_{i=1}^K \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, K$$

where $\kappa(\mathbf{x}_i, \mathbf{y}_i)$ is a kernel function. The authors use RBF kernel function in this paper. And the parameter sigma is empirically set as 14. The discussions about the selection of kernel function and the influence of sigma are given in the Section 4.3.

2.7. *Scoring.* The overall score is computed based on the number of records classified as normal, uncertain, or abnormal, in each of the reference categories. These numbers are denoted by $Nn_k, Nq_k, Na_k, An_k, Aq_k,$ and Aa_k in Table 9.

Weights for the various categories are defined as follows (based on the distribution of the complete test set):

$$wa_1 = \frac{\text{clean abnormal records}}{\text{total abnormal records}}, \tag{11}$$

$$wa_2 = \frac{\text{noisy abnormal records}}{\text{total abnormal records}}. \tag{12}$$

$$wn_1 = \frac{\text{clean normal records}}{\text{total normal records}}, \tag{13}$$

$$wn_2 = \frac{\text{noisy normal records}}{\text{total normal records}}. \tag{14}$$

The modified sensitivity and specificity are defined as (based on a subset of the test set)

$$Se = wa_1 \frac{Aa_1}{Aa_1 + Aq_1 + An_1} + wa_2 \frac{Aa_2 + Aq_2}{Aa_2 + Aq_2 + An_2}, \tag{15}$$

TABLE 9: Variables to evaluate the classification.

		Classification Results		
		Normal (-1)	Uncertain (0)	Abnormal (1)
Reference label	Normal, clean	Nn_1	Nq_1	Na_1
	Normal, noisy	Nn_2	Nq_2	Na_2
	Abnormal, clean	An_1	Aq_1	Aa_1
	Abnormal, noisy	An_2	Aq_2	Aa_2

TABLE 10: Summary of the correlation coefficients.

No.	Feature domain	Max. absolute CC	Physical meaning
1	Time interval	0.286	sd_IntSys
2	State amplitude	-0.159	sd_Amp_S2Sys
3	Energy	0.345	Standard deviation of $Ratio_state_energy_{dia_cycle}$
4	High-order statistics	0.185	sd_S1_kurtosis
5	Cepstrum	0.216	The seventh cepstrum coefficient of S2 state
6	Frequency spectrum of state	0.417	Spectrum value of 30 Hz of S2 state
7	cyclostationarity	-0.240	Sharpness of the peak of cycle frequency spectral density
8	Entropy	-0.374	Average value of sample entropy of diastolic state
9	Frequency spectrum of records	-0.272	Ratio of spectrum magnitude sum in [90 120] Hz

$$Sp = wn_1 \frac{Nn_1}{Na_1 + Nq_1 + Nn_1} + wn_2 \frac{Nn_2 + Nq_2}{Na_2 + Nq_2 + Nn_2}. \quad (16)$$

The overall score is then the average of these two values:

$$\text{Overall score} = \frac{(Se + Sp)}{2}. \quad (17)$$

3. Results

3.1. Correlation Analysis between the Features and the Target Label. In this paper, a total of 515 features are extracted from a single recording. A question arises about how to evaluate the contribution of a feature for classification. To answer the question, correlation analysis is performed between the features and the target label. The correlation coefficients are plotted in the nine domains in Figure 4. The statistics of the coefficients are listed in Table 10. The top coefficient is 0.417 which is from “frequency spectrum of state” at 30 Hz of S2 state. This feature is called “top feature”. The statistics of top features are listed in Table 11. It is shown that 4 in the top 10 features are from “frequency spectrum of state”. So, this domain is ranked the first. Both “energy” and “entropy” contribute 3 in top 10. But “energy” contributes 12 in top 100 which is greater than “entropy” who contributes 8 in top 100. Therefore, “energy” is ranked the 2nd and “entropy” is ranked the 3rd. Following similar logics, the nine domains are ranked as shown in Table 11. It concludes that the domain “frequency spectrum of state” contributes the most. “Energy” and “entropy” are the second and third place to contribute.

3.2. Signal Quality Estimation by SVM. The SVM model in (9) is used to discriminate signal quality. The reference

labels for clean and noisy PCG recordings are “1” and “0”, respectively. The input to the model is the proposed 515 features. The performance is tested by various input, as shown in Table 12. Firstly, 10% of randomly selected data are used for training and the other 90% of data are used for testing without any overlap. Then the percent of train data increases by 10% and repeats. The performance summary for signal quality estimation is listed in Table 12. The manual reference indicates that there are 2874 clean recordings and 279 noisy recordings. So, the numbers of the two quality groups have great unbalance which has bad effect on network training. It is shown that the performance for good signal quality has excellent sensitivity from 96% to 98% no matter how much the percent of data for training varies from 10% to 90%. However, the performance for bad signal quality is poor. The specificity is around 50%; even the training data varies from 10% to 90%. Fortunately, this performance has little influence on the final classification, shown in the next subsection.

3.3. Classification of Normal/Abnormal by SVM. The classification of normal/abnormal is carried out by the SVM model as given in (9). The 515-feature vector is used as input to the SVM network and the label is used as output. The SVM model is firstly trained by a part of data and then tested by the other. To test the generation ability of the model, it is widely tested in following two cases.

Case 1. All data (3153 recordings) are used to train the model and all data (3153 recordings) are to test the model. So, the training data and the testing data are fully overlapped.

Case 2. 10% of the normal recordings and 10% of the abnormal recordings are randomly selected to train the model, and the other 90% are to test the model. The training data and the testing data are exclusively nonoverlapped. This program

TABLE 11: Rank order of the nine domains based on contribution.

Rank order	Feature domain (Total num. of features)	Num. of top 10 features	Num. of top 100 features	Num. of top 200 features	Num. of top 300 features
1	Frequency spectrum of state (308)	4	39	115	183
2	Energy (47)	3	12	16	24
3	Entropy (16)	3	8	10	11
4	Cepstrum (65)	0	14	28	40
5	Time interval (20)	0	14	17	17
6	Frequency spectrum of records (27)	0	5	5	10
7	High-order statistics (16)	0	4	4	7
8	Cyclostationarity (4)	0	2	3	4
9	State amplitude (12)	0	2	2	4

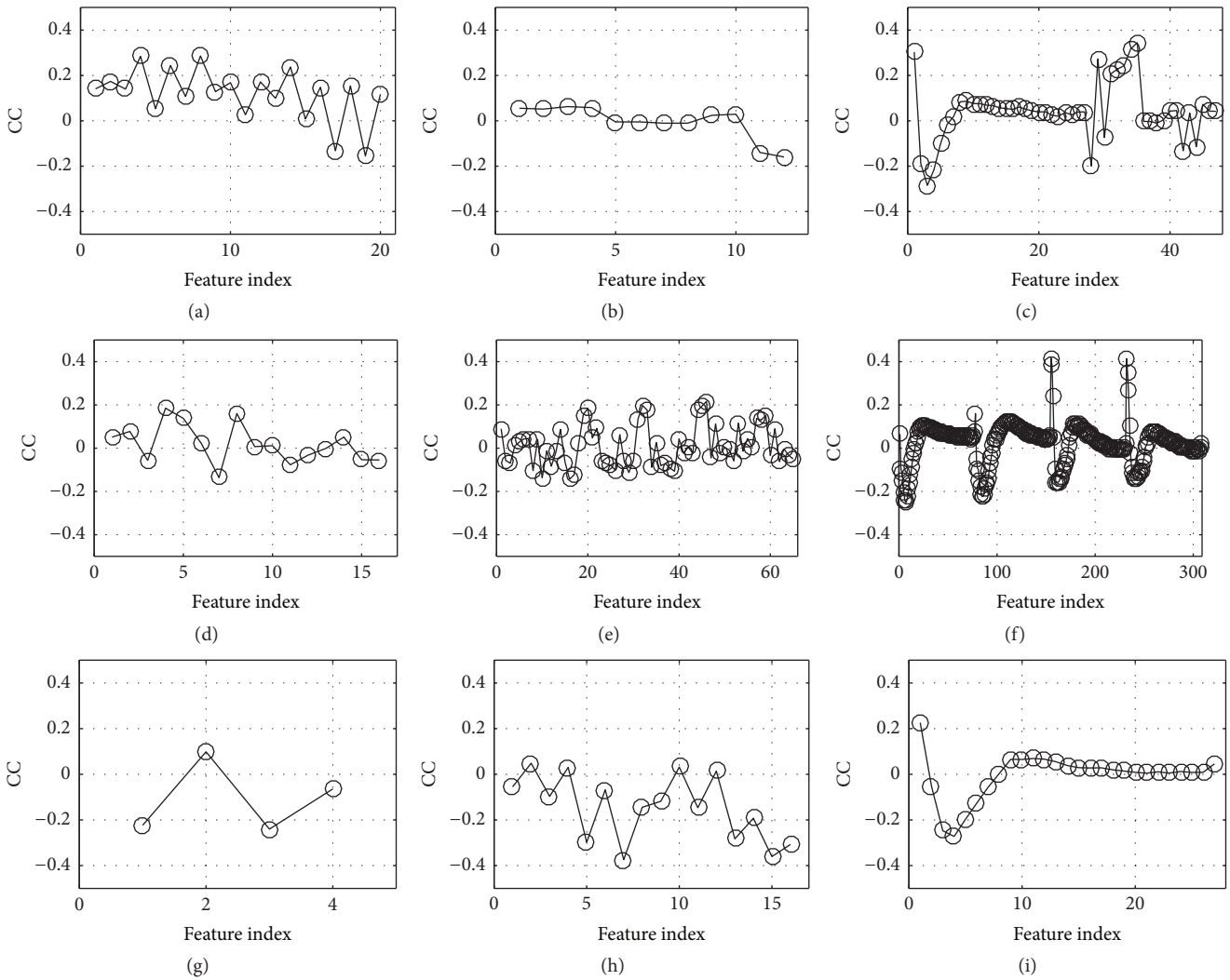


FIGURE 4: Correlation coefficient (CC) between features and the target label. (a) Time interval, (b) state amplitude, (c) energy, (d) high-order statistics, (e) cepstrum, (f) frequency spectrum of state, (g) cyclostationarity, (h) entropy, and (i) frequency spectrum of records.

TABLE 12: Performance of signal quality classification.

Percent data to train	Percent data to test	Estimation results for Reference, clean		Estimation results for Reference, noisy		Sensitivity	Specificity
		Clean	Noisy	Clean	Noisy		
10%	90%	2421	166	104	146	0.96	0.47
20%	80%	2143	154	75	149	0.97	0.49
30%	70%	1869	143	59	136	0.97	0.49
40%	60%	1601	125	44	122	0.97	0.49
50%	50%	1332	104	36	104	0.97	0.50
60%	40%	1064	84	28	84	0.97	0.50
70%	30%	799	63	19	64	0.97	0.50
80%	20%	530	44	12	45	0.98	0.50
90%	10%	265	21	6	22	0.98	0.50

TABLE 13: Performance of the classification.

Case	Percent of data to train	Percent of data to test	Repeat times	Training and test data division	Sensitivity	Specificity	Overall score
Case 1	100%	100%	1	No	0.99	0.91	0.95
	10%	90%	200	Yes	0.68±0.06	0.87±0.03	0.77±0.02
	20%	80%	200	Yes	0.76±0.05	0.86±0.02	0.81±0.02
	30%	70%	200	Yes	0.80±0.04	0.87±0.02	0.83±0.02
	40%	60%	200	Yes	0.82±0.04	0.87±0.01	0.85±0.02
Case 2	50%	50%	200	Yes	0.84±0.03	0.87±0.01	0.85±0.01
	60%	40%	200	Yes	0.85±0.04	0.87±0.01	0.86±0.01
	70%	30%	200	Yes	0.86±0.04	0.87±0.01	0.87±0.02
	80%	20%	200	Yes	0.87±0.04	0.87±0.02	0.87±0.02
	90%	10%	200	Yes	0.88±0.04	0.87±0.02	0.88±0.02

Note: the number is presented as mean±SD.

independently repeats 200 times to evaluate the stability. Sensitivity and specificity are calculated in “mean±SD” to indicate the classification performance. Then, the percent of training data increases by 10%, the percent of test data decreases by 10%, and the evaluation process is repeated until the percent of training data reaches 90%.

The performance of the proposed classification is listed in Table 13. The overall score of Case 1 is up to 0.95. It proves that the proposed features are effective for this classification. In Case 2, it can be found that, with the increasing percent of data for training, both sensitivity and specificity increase. The standard deviation is not greater than 0.02. So, the score variation is very small; even the classifier independently runs 200 times. This simulation proves that the proposed features and the model have excellent generation ability and stability and are effective in discriminating the PCG recordings.

4. Discussions

4.1. Effect of the Number of Top Features. This paper proposes 515 features from multidomains. However, correlation analysis shows that each feature has different degree of

correlation with the target label. The performance will change with the number of selected features. To evaluate the effect of selected features, the authors conduct simulations under condition of varying the top number of features. The mean overall score changing with respect to the number of top features is illustrated in Figure 5, where Figure 5(a) shows the performance with top 1 to top 5 features, Figure 5(b) is with top 10 to top 50 features, and Figure 5(c) is with top 100 to top 515 features. It can be seen that there are two factors to influence SVM classifier’s generalization ability. One is the percent of data for training; the other is the number of top features. An overall look shows that both the two factors have positive effect on the classification performance. Roughly speaking, if any one of them increases, the performance will get improvement. However, it is not always true. A closer look at Figure 5(a) indicates the performance has little change as the percent increases. But the performance gets much improvement as the percent increases, shown in Figures 5(b) and 5(c), where the number of top features is much greater than that in Figure 5(a). A careful look at Figure 5(c) discloses that the case that all the features (515) are involved does not result in the best performance. It can

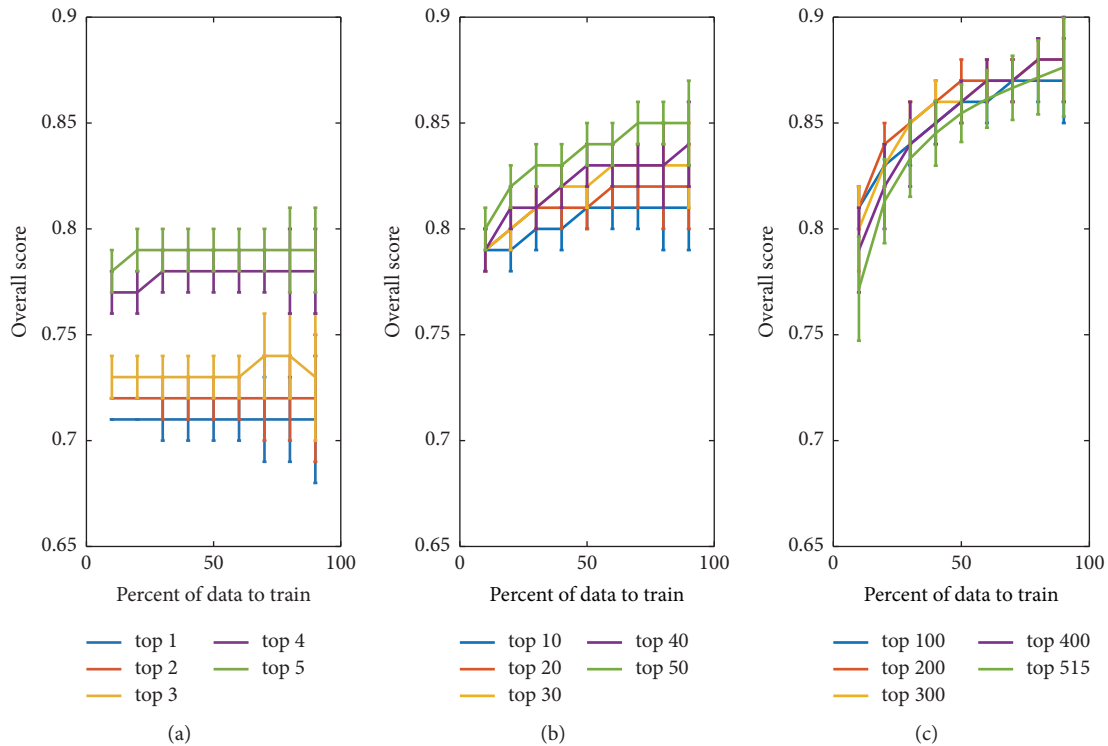


FIGURE 5: Classification performance with respect to the number of top features and percent of data for training. (a) Overall scores obtained by top 1, top 2, top 3, top 4, and top 5 features. (b) Overall scores obtained by top 10, top 20, top 30, top 40, and top 50 features. (c) Overall scores obtained by top 100, top 200, top 300, top 400, and top 515 features.

be found that there is an “edge effect” on the selection of top feature number. That is, much improvement can be obtained via increasing top feature number as the number is small. However, the improvement becomes little when the number is up to some degree. The best performance is obtained with top 400 features in this paper. The performance will get worse if the number continues to increase.

The proposed classification has very good performance even if the number of features is small. For example, it can be noted in Figure 5(a) that the overall score is up to 0.71 as only the top 1 feature is used and the score increases to 0.81 when the top 10 features are used. This is one of attractive advantages of the proposed classification.

Another advantage is that the proposed SVM classifier has very stable output. Even if the SVM classifier is trained independently by randomly selected features, the overall score has very low variations (standard variance is approximately lower than 0.02). That is to say, the proposed features and SVM classifier are adaptable to the classification.

4.2. Classification Performance Based on Features in Specified Domain. Table 13 and Figure 5 show the classification performance based on mixed features from multidomains. It is interesting to test the performance based on features of a separated domain. This test would be evidence to show the power of combined features for classification. So, the SVM classifier and 10-fold validation are used for this purpose. The results are listed in Table 14. It is seen that,

the highest score, 0.85, is produced if only the features in “frequency spectrum of state” are used. Other high scores are obtained based on features in domain of “entropy”, “energy”, and “cepstrum”. It can be found that these results are almost coincident with those of correlation analysis given in Section 3.1 where “frequency spectrum of state”, “energy”, “entropy”, and “cepstrum” are the top domains to contribute effective features. This simulation indicates that it is reasonable to improve classification performance by combining multidomain features.

4.3. Selection of Kernel Function and Influence of the Gaussian Kernel Function. Typically, the kernel functions for a SVM have several selections, such as linear kernel, polynomial kernel, sigmoid kernel, and Gaussian radial basis function. Given an arbitrary dataset, one does not know which kernel may work best. Generally, one can start with the simplest hypothesis space first and then work a way up towards a more complex hypothesis space. The authors followed this lesson summarized by the previous researchers. A bad performance was produced by the SVM classifier using a linear kernel since 515 features in this study were complex and they were not linearly separable, as well as using a polynomial kernel. The authors actually tried out all possible kernels and found that the RBF kernel produced the best performance. The authors believed that the best performance should be attributed to the RBF kernel’s advantages. The first is translation invariance. Let the RBF kernel be $K(\mathbf{x}_i, \mathbf{x}_j) =$

TABLE 14: Classification performance based on features in specified domain.

Rank	Domain (# features)	Mean of overall score	Standard deviation
1	Frequency spectrum of state (308)	0.85	0.021
2	Entropy (16)	0.82	0.028
3	Energy (47)	0.78	0.020
4	Cepstrum (65)	0.75	0.027
5	High-order statistics (16)	0.73	0.029
6	Frequency spectrum of records (27)	0.71	0.025
7	Time interval (20)	0.70	0.025
8	Cyclostationarity (4)	0.65	0.042
9	State amplitude (12)	0.61	0.025

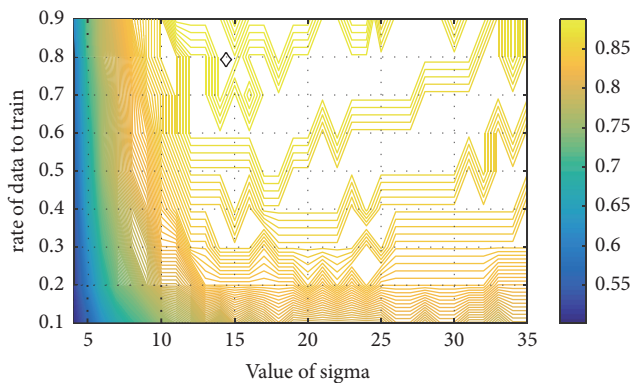


FIGURE 6: The mean overall score with respect to rate of data for training and value of sigma. The diamond shows the peak position.

$\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|/\gamma)$; then $K(\mathbf{x}_i, \mathbf{x}_j) = K(\mathbf{x}_i + \boldsymbol{\nu}, \mathbf{x}_j + \boldsymbol{\nu})$ where $\boldsymbol{\nu}$ is any arbitrary vector. It is known that the kernel is effectively a similarity measure. The invariance is useful to measure the similarity between the proposed features. The second is that the similarity is measured by Euclidean distance. RBF kernel is a function of the Euclidean distance between the features. In this study, the Euclidean distance is a preferred similarity metric. The authors selected the RBF kernel function because the advantages match the classification purpose and features.

One difficulty with the Gaussian RBF kernel function is the parameter sigma governing the kernel width. A general conclusion about sigma has been summarized by previous researchers. A large value of sigma may lead to an over smoothing hyperplane and a washing out of structure that might otherwise be extracted from the feature space. A reducing value of sigma may lead to a noisy hyperplane elsewhere in the feature space where the feature density is smaller. There is a trade-off between sensitivity to noise at small value and over smoothing at large value. To select appropriate value for sigma, the authors did grid search in a specified range, as seen in Figure 6. This figure shows the mean overall score, based on 50 independent runs, with respect to rate of data for training and value of sigma where the value of sigma increases from 4 to 35 by step of 1 and the rate of data for training increases from 0.1 to 0.9 by step of 0.1. It is found that the peak of the overall score occurs at sigma 14 as indicated by the diamond.

5. Conclusions

In this paper, 515 features are extracted from multiple domains, i.e., time interval, state amplitude, energy, high-order statistics, cepstrum, frequency spectrum, cyclostationarity, and entropy. Correlation analysis between the features and the target label shows that the features from frequency spectrum contribute the most to the classification. The features and the SVM classifier jointly show the powerful classification performance. The results show the overall score reaches 0.88 ± 0.02 based on 200 independent simulations, which is competitive to the previous best classification methods. Moreover, the SVM classifier has very good performance with even small number of features for training and has stable output regardless of randomly selected features for training.

Data Availability

The authors state that the data used in this study are available from the website <http://www.physionet.org/challenge/2016/> to support the conclusions.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants no. 61471081 and no. 61601081 and Fundamental Research Funds for the Central Universities under Grant nos. DUT15QY60, DUT16QY13, DC201501056, and DCPY2016008.

References

- [1] A. A. Luisada, C. K. Liu, C. Aravanis, M. Testelli, and J. Morris, "On the mechanism of production of the heart sounds," *American Heart Journal*, vol. 55, no. 3, pp. 383–399, 1958.
- [2] T. Sakamoto, R. Kusukawa, D. M. Maccanon, and A. A. Luisada, "Hemodynamic determinants of the amplitude of the first heart sound," *Circulation Research*, vol. 16, pp. 45–57, 1965.

- [3] T. Sakamoto, R. Kusakawa, D. M. MacCanon, and A. A. Luisada, "First heart sound amplitude in experimentally induced alternans," *CHEST*, vol. 50, no. 5, pp. 470–475, 1966.
- [4] L.-G. Durand and P. Pibarot, "Digital signal processing of the phonocardiogram: Review of the most recent advancements," *Critical Reviews in Biomedical Engineering*, vol. 23, no. 3-4, pp. 163–219, 1995.
- [5] D. S. Gerbarg, A. Taranta, M. Spagnuolo, and J. J. Hoffer, "Computer analysis of phonocardiograms," *Progress in Cardiovascular Diseases*, vol. 5, no. 4, pp. 393–405, 1963.
- [6] Y. Akay, M. Akay, W. Welkowitz, and J. Kostis, "Noninvasive detection of coronary artery disease," *IEEE Engineering in Medicine and Biology Magazine*, vol. 13, no. 5, pp. 761–764, 1994.
- [7] L.-G. Durand, M. Blanchard, G. Cloutier, H. N. Sabbah, and P. D. Stein, "Comparison of Pattern Recognition Methods for Computer-Assisted Classification of Spectra of Heart Sounds in Patients with a Porcine Bioprosthetic Valve Implanted in the Mitral Position," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 12, pp. 1121–1129, 1990.
- [8] H. Uğuz, "A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases," *Journal of Medical Systems*, vol. 36, no. 1, pp. 61–72, 2012.
- [9] T. Ölmez and Z. Dokur, "Classification of heart sounds using an artificial neural network," *Pattern Recognition Letters*, vol. 24, no. 1–3, pp. 617–629, 2003.
- [10] Z. Dokur and T. Ölmez, "Heart sound classification using wavelet transform and incremental self-organizing map," *Digital Signal Processing*, vol. 18, no. 6, pp. 951–959, 2008.
- [11] S. Ari, K. Hembram, and G. Saha, "Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8019–8026, 2010.
- [12] I. Maglogiannis, E. Loukis, E. Zafropoulos, and A. Stasis, "Support Vectors Machine-based identification of heart valve diseases using heart sounds," *Computer Methods and Programs in Biomedicine*, vol. 95, no. 1, pp. 47–61, 2009.
- [13] Y. Zheng, X. Guo, and X. Ding, "A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2710–2721, 2015.
- [14] P. M. Bentley, P. M. Grant, and J. T. E. McDonnell, "Time-frequency and time-scale techniques for the classification of native and bioprosthetic heart valve sounds," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 1, pp. 125–128, 1998.
- [15] L. D. Avendaño-Valencia, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Feature extraction from parametric time-frequency representations for heart murmur detection," *Annals of Biomedical Engineering*, vol. 38, no. 8, pp. 2716–2732, 2010.
- [16] A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Selection of dynamic features based on time-frequency representations for heart murmur detection from phonocardiographic signals," *Annals of Biomedical Engineering*, vol. 38, no. 1, pp. 118–137, 2010.
- [17] T. Yang and H. Hsieh, "Classification of Acoustic Physiological Signals Based on Deep Learning Neural Networks with Augmented Features," in *Proceedings of the 43rd Computing in Cardiology Conference*, vol. 43, Vancouver, Canada, 2016.
- [18] M. Tschannen, T. Kramer, G. Marti, M. Heinzmann, and T. Wiatowski, "Heart sound classification using deep structured features," in *Proceedings of the 43rd Computing in Cardiology Conference*, *CinC 2016*, pp. 565–568, Vancouver, Canada, September 2016.
- [19] T. Nilanon, J. Yao, J. Hao, S. Purushotham, and Y. Liu, "Normal / abnormal heart sound recordings classification using convolutional neural network," in *Proceedings of the 43rd Computing in Cardiology Conference*, *CinC 2016*, pp. 585–588, Vancouver, Canada, September 2016.
- [20] I. Diaz Bobillo, "A Tensor Approach to Heart Sound Classification," in *Proceedings of the 2016 Computing in Cardiology Conference*, vol. 43, Vancouver, Canada, 2016.
- [21] A. Bouril, D. Aleinikava, M. S. Guillem, and G. M. Mirsky, "Automated classification of normal and abnormal heart sounds using support vector machines," in *Proceedings of the 43rd Computing in Cardiology Conference*, *CinC 2016*, pp. 549–552, Vancouver, Canada, September 2016.
- [22] J. J. Gonzalez Ortiz, C. P. Phoo, and J. Wiens, "Heart Sound Classification Based on Temporal Alignment Techniques," in *Proceedings of the 2016 Computing in Cardiology Conference*, vol. 43, Vancouver, Canada, 2016.
- [23] M. Nabhan Homsy, N. Medina, M. Hernandez et al., "Automatic Heart Sound Recording Classification using a Nested Set of Ensemble Algorithms," in *Proceedings of the 2016 Computing in Cardiology Conference*, vol. 43, Vancouver, Canada, 2016.
- [24] P. Langley and A. Murray, "Heart sound classification from unsegmented phonocardiograms," *Physiological Measurement*, vol. 38, no. 8, pp. 1658–1670, 2017.
- [25] M. Nabhan Homsy and P. Warrick, "Ensemble methods with outliers for phonocardiogram classification," *Physiological Measurement*, vol. 38, no. 8, pp. 1631–1644, 2017.
- [26] B. Whitaker and D. Anderson, "Heart Sound Classification via Sparse Coding," in *Proceedings of the 43rd Computing in Cardiology Conference*, vol. 43, Vancouver, Canada, 2016.
- [27] B. M. Whitaker, P. B. Suresha, C. Liu, G. D. Clifford, and D. V. Anderson, "Combining sparse coding and time-domain features for heart sound classification," *Physiological Measurement*, vol. 38, no. 8, pp. 1701–1713, 2017.
- [28] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in *Proceedings of the 43rd Computing in Cardiology Conference*, *CinC 2016*, pp. 621–624, Vancouver, Canada, September 2016.
- [29] M. Zabihi, A. Bahrami Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos, "Heart Sound Anomaly and Quality Detection using Ensemble of Neural Networks without Segmentation," in *43rd Computing in Cardiology Conference*, vol. 43, Vancouver, Canada, 2016.
- [30] E. Kay and A. Agarwal, "DropConnected Neural Network Trained with Diverse Features for Classifying Heart Sounds," in *Proceedings of the 43rd Computing in Cardiology Conference*, vol. 43, Vancouver, Canada, 2016.
- [31] H. Tang, H. Chen, T. Li, and M. Zhong, "Classification of Normal/Abnormal Heart Sound Recordings based on Multi-Domain Features and Back Propagation Neural Network," in *Proceedings of the 43rd Computing in Cardiology Conference*, vol. 43, Vancouver, Canada, 2016.
- [32] <http://www.physionet.org/challenge/2016/>, Nov. 16, 2016.
- [33] C. Liu, D. Springer, Q. Li et al., "An open access database for the evaluation of heart sound algorithms," *Physiological Measurement*, vol. 37, no. 12, pp. 2181–2213, 2016.

- [34] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," *Physiological Measurement*, vol. 31, no. 4, pp. 513–529, 2010.
- [35] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-HSMM-based heart sound segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2016.
- [36] L. E. Boucheron, P. L. De Leon, and S. Sandoval, "Low bit-rate speech coding through quantization of mel-frequency cepstral coefficients," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 2, pp. 610–619, 2012.
- [37] T. Li, H. Tang, T. Qiu, and Y. Park, "Best subsequence selection of heart sound recording based on degree of sound periodicity," *IEEE Electronics Letters*, vol. 47, no. 15, pp. 841–843, 2011.
- [38] C. Liu, K. Li, L. Zhao et al., "Analysis of heart rate variability using fuzzy measure entropy," *Computers in Biology and Medicine*, vol. 43, no. 2, pp. 100–108, 2013.
- [39] L. Zhao, S. Wei, C. Zhang et al., "Determination of sample entropy and fuzzy measure entropy parameters for distinguishing congestive heart failure from normal sinus rhythm subjects," *Entropy*, vol. 17, no. 9, pp. 6270–6288, 2015.