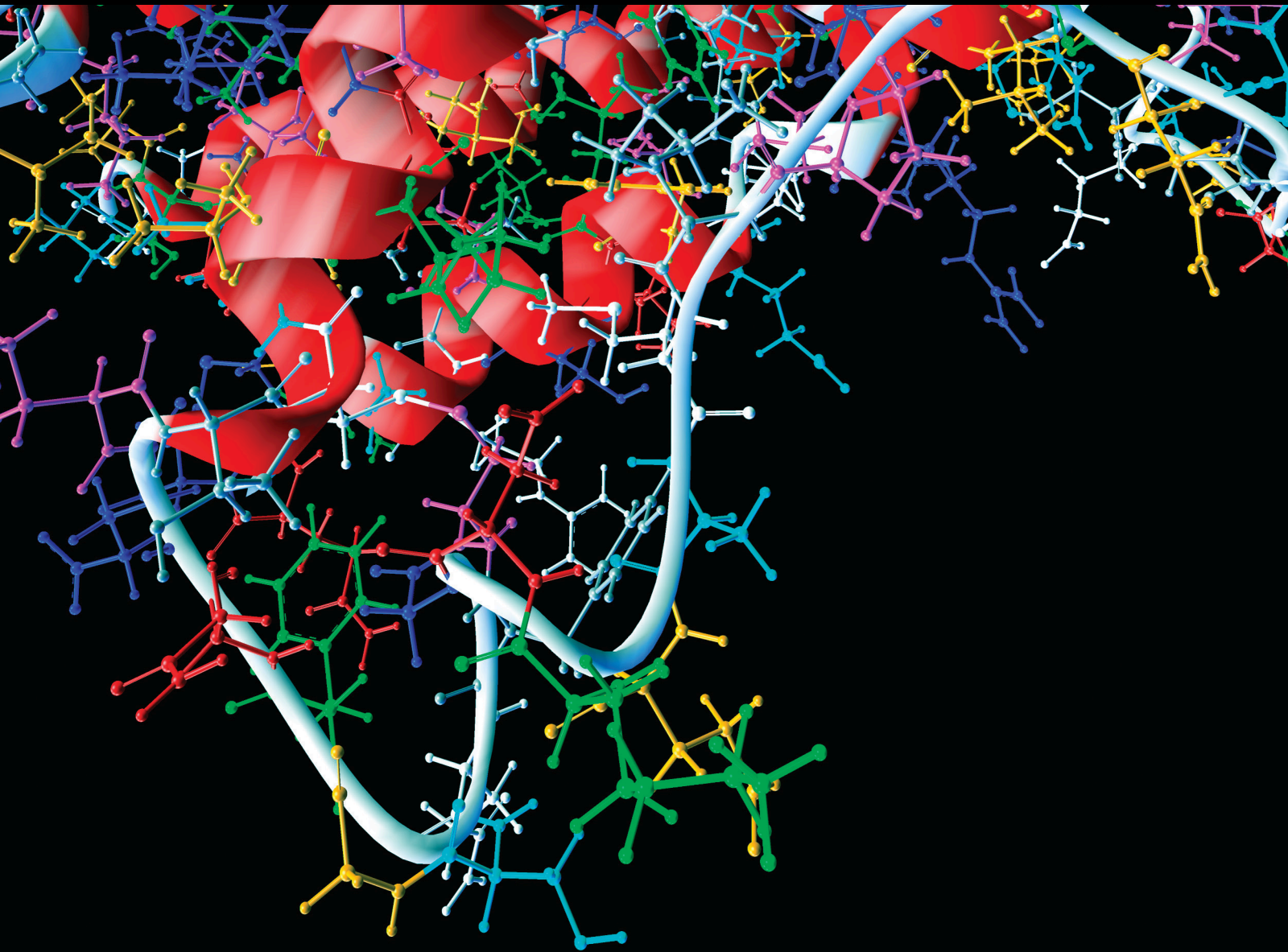


Application and Decision-Making Implications of Novel Optimization Models in Public Health

Lead Guest Editor: Plácido R. Pinheiro

Guest Editors: Miltiadis D. Lytras, Anna Visvizi, and Mirian C. D. Pinheiro





Application and Decision-Making Implications of Novel Optimization Models in Public Health

Computational and Mathematical Methods in Medicine

**Application and Decision-Making
Implications of Novel Optimization
Models in Public Health**

Lead Guest Editor: Plácido R. Pinheiro

Guest Editors: Miltiadis D. Lytras, Anna Visvizi,
and Mirian C. D. Pinheiro



Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in “Computational and Mathematical Methods in Medicine.” 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.

Chief Editor

David A. Winkler, Australia

Editorial Board

Raul Alcaraz, Spain
Emil Alexov, USA
Syed Muhammad Anwar, Pakistan
Konstantin G. Arbeev, USA
Enrique Baca-Garcia, Spain
Enrique Berjano, Spain
Junguo Bian, USA
Elia Biganzoli, Italy
Hans A. Braun, Germany
Zoran Bursac, USA
Guy Carrault, France
Filippo Castiglione, Italy
Prem Chapagain, USA
Huiling Chen, China
Maria N. D.S. Cordeiro, Portugal
Cristiana Corsi, Italy
Qi Dai, China
Chuangyin Dang, Hong Kong
Didier Delignières, France
Jun Deng, USA
Thomas Desaive, Belgium
Dr. Gaurav Dhiman, India
David Diller, USA
Michel Dojat, France
Irina Doytchinova, Bulgaria
Esmail Ebrahimie, Australia
Issam El Naqa, USA
Angelo Facchiano, Italy
Luca Faes, Italy
Maria E. Fantacci, Italy
Giancarlo Ferrigno, Italy
Marc Thilo Figge, Germany
Alfonso T. García-Sosa, Estonia
Igor I. Goryanin, Japan
Marko Gosak, Slovenia
Damien Hall, Australia
Roberto Hornero, Spain
Tingjun Hou, China
Seiya Imoto, Japan
Martti Juhola, Finland
Rafik Karaman, Palestinian Authority
Chandan Karmakar, Australia
Kwang Gi Kim, Republic of Korea
Andrzej Kloczkowski, USA





Andrei Korobeinikov, China
Peng Li, USA
Chung-Min Liao, Taiwan
Ezequiel López-Rubio, Spain
Pinyi Lu, USA
Reinoud Maex, Belgium
Valeri Makarov, Spain
Juan Pablo Martínez, Spain
Richard J. Maude, Thailand
John Mitchell, United Kingdom
Luminita Moraru, Romania
Vishal Nayak, USA
Michele Nichelatti, Italy
Kazuhisa Nishizawa, Japan
Farai Nyabadza, South Africa
Francesco Pappalardo, Italy
Hyuntae Park, Japan
Jovana Paunovic, Serbia
Manuel F. G. Penedo, Spain
Kemal Polat, Turkey
Alberto Policriti, Italy
Giuseppe Pontrelli, Italy
Jesús Poza, Spain
Christopher Pretty, New Zealand
Mihai V. Putz, Romania
Jose Joaquin Rieta, Spain
Jan Rychtar, USA
Murat Sari, Turkey
Simon A. Sherman, USA
Joon Jin Song, USA
Xinyuan Song, Hong Kong
Dong Song, USA
Yung-Shin Sun, Taiwan
Min Tang, China
João M. Tavares, Portugal
Jlenia Toppi, Italy
Anna Tsantili-Kakoulidou, Greece
Markos G. Tsipouras, Greece
Po-Hsiang Tsui, Taiwan
Gabriel Turinici, France
Ruisheng Wang, USA
Liangjiang Wang, USA
Gabriel Wittum, Germany
Yu Xue, China









KI Yanover, Israel
Xiaojun Yao, China
Kaan Yetilmezsoy, Turkey
Hiro Yoshida, USA
Yuhai Zhao, China

Contents





Application and Decision-Making Implications of Novel Optimization Models in Public Health

Plácido R. Pinheiro , Miltiades D. Lytras , Anna Visvizi , and Mirian Caliope D. Pinheiro 
Editorial (2 pages), Article ID 9764895, Volume 2022 (2022)





Application of Data Mining Algorithms for Dementia in People with HIV/AIDS

Luana Ibiapina Cordeiro Caliope Pinheiro , Maria Lúcia Duarte Pereira , Marcial Porto Fernandez ,
Francisco Mardônio Vieira Filho , Wilson Jorge Correia Pinto de Abreu , and Pedro Gabriel Caliope
Dantas Pinheiro 
Research Article (8 pages), Article ID 4602465, Volume 2021 (2021)

A Protocol for the Diagnosis of Autism Spectrum Disorder Structured in Machine Learning and Verbal Decision Analysis

Evandro Andrade, Samuel Portela, Plácido Rogério Pinheiro , Luciano Comin Nunes , Marum Simão
Filho , Wagner Silva Costa, and Mirian Caliope Dantas Pinheiro 
Research Article (14 pages), Article ID 1628959, Volume 2021 (2021)





SARPPIC: Exploiting COVID-19 Contact Tracing Recommendation through Social Awareness

Nana Yaw Asabere , Amevi Acakpovi , Emmanuel Kwaku Ofori , Wisdom Torgby, Marcellinus
Kuuboore , Gare Lawson, and Edward Adjaloko
Research Article (14 pages), Article ID 3460130, Volume 2020 (2020)


A Simple Framework of Smart Geriatric Nursing considering Health Big Data and User Profile

Shijie Li and Yongchuan Tang 
Research Article (9 pages), Article ID 5013249, Volume 2020 (2020)

A Probabilistic Patient Scheduling Model with Time Variable Slots

Danae Carreras-García , David Delgado-Gómez , Enrique Baca-García , and Antonio Artés-
Rodríguez 
Research Article (10 pages), Article ID 9727096, Volume 2020 (2020)



Utility-Based Multicriteria Model for Screening Patients under the COVID-19 Pandemic

Lucia Reis Peixoto Roselli , Eduarda Asfora Frej, Rodrigo José Pires Ferreira, Alexandre Ramalho
Alberti, and Adiel Teixeira de Almeida
Research Article (8 pages), Article ID 9391251, Volume 2020 (2020)

An Intelligent Decision-Making Support System for the Detection and Staging of Prostate Cancer in Developing Countries

Jun Zhang, Zhigang Chen , Jia Wu , and Kanghuai Liu
Research Article (18 pages), Article ID 5363549, Volume 2020 (2020)


Capacity Evaluation of Diagnostic Tests For COVID-19 Using Multicriteria Decision-Making Techniques

Murat Sayan, Figen Sarigul Yildirim, Tamer Sanlidag, Berna Uzun, Dilber Uzun Ozsahin , and Ilker
Ozsahin 
Research Article (8 pages), Article ID 1560250, Volume 2020 (2020)


Applying Queuing Theory and Mixed Integer Programming to Blood Center Nursing Schedules of a Large Hospital in China

Li Luo, Xiaofei Liu , Xinyuan Cui, Yuanjun Cheng , Xinzhu Yu, Yue Li, Li Jiang, and Mingying Tan
Research Article (7 pages), Article ID 9373942, Volume 2020 (2020)

Auxiliary Medical Decision System for Prostate Cancer Based on Ensemble Method

Jia Wu , Qinghe Zhuang, and Yanlin Tan 
Research Article (11 pages), Article ID 6509596, Volume 2020 (2020)

Identification of Upper-Limb Movements Based on Muscle Shape Change Signals for Human-Robot Interaction

Pingao Huang , Hui Wang , Yuan Wang, Zhiyuan Liu, Oluwarotimi Williams Samuel, Mei Yu, Xiangxin Li, Shixiong Chen, and Guanglin Li 
Research Article (14 pages), Article ID 5694265, Volume 2020 (2020)

Using Eye Aspect Ratio to Enhance Fast and Objective Assessment of Facial Paralysis

Jialing Feng , Zhexiao Guo , Jun Wang , and Guo Dan 
Research Article (11 pages), Article ID 1038906, Volume 2020 (2020)

DBT Masses Automatic Segmentation Using U-Net Neural Networks

Xiaobo Lai , Weiji Yang , and Ruipeng Li 
Research Article (10 pages), Article ID 7156165, Volume 2020 (2020)

Editorial

Application and Decision-Making Implications of Novel Optimization Models in Public Health

Plácido R. Pinheiro ¹, **Miltiades D. Lytras** ², **Anna Visvizi** ³,
and Mirian Caliope D. Pinheiro ¹

¹University of Fortaleza, Fortaleza, Brazil

²The American College of Greece, Athens, Greece

³Research and Innovation Institute, Warsaw, Poland

Correspondence should be addressed to Plácido R. Pinheiro; placido@unifor.br

Received 1 February 2022; Accepted 1 February 2022; Published 12 April 2022

Copyright © 2022 Plácido R. Pinheiro 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.

Optimizing the results of public health management is a goal of many managers. Contracting inputs and service providers oriented towards health outcomes, such as models anchored in the population's health, building the capacity to contract, allow countries to make a difference in a pandemic. For example, systems have been increasingly used in various novel optimization models in public health such as disease prevention, detection, diagnosis, innovative education environments, security, and authentication. These systems are suggested in the literature with alternative tools to aid specialists in their decision-making by analyzing and interpreting data.

The main objective of this special issue was to promote a broad discussion on the recent advances in optimization models with novel methods and applications, seeking to identify potential contributions to public health. This special issue of the Computational and Mathematical Methods in Medicine journal contains 13 original works. Therefore, the selected studies address new trends in methods and applied techniques to different public health.

The first article by Xiaobo Lai et al. proposed a digital breast tomosynthesis (DBT) image automatic segmentation algorithm using a U-Net architecture to improve automatic breast mass segmentation accuracy in DBT imaging. The proposed automatic segmentation obtained promising results, which are superior to some classical architectures, and it is expected that they have perspectives of clinical application.

The second article by J. Feng et al. investigated patients with facial paralysis who were enrolled as study objects. The eye aspect ratio (EAR) index was proposed for the eye region. The correlation between EAR and the score facial nerve grading system (FNGS 2.0) was analyzed to verify the ability of EAR to enhance FNGS 2.0 for the rapid and objective assessment of the severity of facial paralysis.

Moreover, *the third article* by Pingao Huang et al. describes a new flexible and extensible nanogold sensor developed to acquire muscle shape change (MSC) signals used to decode various classes of limb motion intentions. The results suggest that the MSC-based method should be feasible in movement identifications for human-robot interaction. At the same time, they provide a frequent reference for using flexible and stretchable sensors in human-robot interaction systems.

Furthermore, *the fourth article* by Jia Wu et al. proposes a framework that uses six relevant tumor markers as the input features. They employed classical machine learning models (support vector machine and artificial neural network) for auxiliary medical decision systems for prostate cancer (PCa). Results revealed the increasing prevalence of PCa and the great negative impact caused by a high-fat diet and genetic inheritance.

In addition, *the fifth article* by Li Lou et al. presented a hybrid model structured on the queuing theory and mixed integer linear programming to develop flexible scheduling for nurses working in blood centers. The number of nurses

needed during a working period was reasonably determined and nursing schedules were organized flexibly, ensuring sufficient rest periods for each nurse.

On the other hand, the *sixth article* by Murat Sayan et al. describes an approach to compare the selection of effective diagnostic methods to make a mutual comparison between existing SARS-CoV-2 diagnostic tests and determine the most effective. The study results with proposed multicriteria decision-making methods (MCDM) indicated that the most effective diagnostic method for COVID-19 was chest CT.

Following, the *seventh article* by Jun Zhang et al. propose a novel intelligent data decision-making system for prostate cancer based on a neural perceptron network. Decisions are mainly made by associating relevant indicators of the disease and combining them with medical images to solve the problems and social contradictions faced by most developing countries.

The *eighth article* by Danae Carreras-Garcia et al. investigates a scheduling system based on the nonattendance probabilities of patients with variable time intervals and a dynamic priority allocation scheme. The system is structured around a mixed integer programming model that is aimed at maximizing the clinic's expected profits, accounting for the first appointments and returns.

The *ninth article* by Lucia Reis Peixoto Rosetti et al. presents a utility-based multicriteria model to support physicians in dealing with an important decision—the screening decision problem—due to the scarcity of resources due to the COVID-19 pandemic. The model is applied to generate the best alternatives for treating patients with suspected COVID-19, namely, an intensive care unit (ICU), a hospital ward, or an isolated home.

The *tenth article* by Shijie Li and Yongchuan Tang proposes an intelligent geriatric nursing framework that consists of three aspects of competent nursing: smart geriatric nursing in physical health using advanced biosensors and devices, intelligent geriatric nursing in mental health based on user profile, and intelligent geriatric nursing for daily life based on big data in health. The implementation of the proposed method relies on Internet of Things (IoT) technologies, user profile systems, big data, and many other advanced information technologies.

The *eleventh article* by Nana Yaw Asabere explored a referral algorithm called socially aware referral of people likely to be infected with COVID-19. Initially, it applies intermediation centrality in a social network to measure the number of target touchpoints (nodes/users) that met an infected touchpoint (patient COVID-19). Then, the clashing forces of the same contact points are also calculated using contact durations and contact frequencies.

The *twelfth article* by Evandro Andrade et al. evaluated a hybrid approach based on machine learning algorithms composition to discover knowledge and concepts associated with the multicriteria method of decision support based on verbal decision analysis to refine the results. The study is aimed at evaluating how the mentioned hybrid methodology proposal can make the protocol derived from ICD-10 more efficient, providing agility to diagnosing autism spectrum disorder by observing a minor symptom.

The *last article of the special issue*, by Luana Ibiapina Cordeiro Caliope Pinheiro et al., describes an approach that is aimed at identifying the best performing algorithm and the most relevant characteristics to categorize individuals with HIV/AIDS at a high risk of dementia from the application of data mining. The principal component analysis (PCA) algorithm was used and tested comparatively between the following machine learning algorithms: logistic regression, decision tree, neural network, KNN, and random forest.

Conflicts of Interest

The guest editors declare that there are no conflicts of interest.







Acknowledgments

The guest editors wish to thank all the authors and reviewers for sharing and helping to improve the works published here, respectively.

Plácido R. Pinheiro
Miltiades D. Lytras
Anna Visvizi
Mirian Caliope D. Pinheiro

Research Article

Application of Data Mining Algorithms for Dementia in People with HIV/AIDS

Luana Ibiapina Cordeiro Calíope Pinheiro ¹, **Maria Lúcia Duarte Pereira** ¹,
Marcial Porto Fernandez ², **Francisco Mardônio Vieira Filho** ²,
Wilson Jorge Correia Pinto de Abreu ³ and **Pedro Gabriel Calíope Dantas Pinheiro** ⁴

¹Graduate Program in Clinical Care in Nursing and Health, State University of Ceará, Fortaleza, Brazil

²Graduate Program in Computer Science, State University of Ceará, Fortaleza, Brazil

³Porto School of Nursing, Porto, Portugal

⁴Graduate Program in Applied Informatics, University of Fortaleza, Fortaleza, Brazil

Correspondence should be addressed to Luana Ibiapina Cordeiro Calíope Pinheiro; luanaibiapina1@hotmail.com

Received 2 June 2020; Accepted 21 June 2021; Published 10 July 2021

Academic Editor: Raul Alcaraz

Copyright © 2021 Luana Ibiapina Cordeiro Calíope Pinheiro 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.

Dementia interferes with the individual's motor, behavioural, and intellectual functions, causing him to be unable to perform instrumental activities of daily living. This study is aimed at identifying the best performing algorithm and the most relevant characteristics to categorise individuals with HIV/AIDS at high risk of dementia from the application of data mining. Principal component analysis (PCA) algorithm was used and tested comparatively between the following machine learning algorithms: logistic regression, decision tree, neural network, KNN, and random forest. The database used for this study was built from the data collection of 270 individuals infected with HIV/AIDS and followed up at the outpatient clinic of a reference hospital for infectious and parasitic diseases in the State of Ceará, Brazil, from January to April 2019. Also, the performance of the algorithms was analysed for the 104 characteristics available in the database; then, with the reduction of dimensionality, there was an improvement in the quality of the machine learning algorithms and identified that during the tests, even losing about 30% of the variation. Besides, when considering only 23 characteristics, the precision of the algorithms was 86% in random forest, 56% logistic regression, 68% decision tree, 60% KNN, and 59% neural network. The random forest algorithm proved to be more effective than the others, obtaining 84% precision and 86% accuracy.

1. Introduction

Data mining (MD) is one of the data exploration processes capable of predicting and extracting consistent patterns by using strategies such as learning algorithms, such as artificial intelligence (AI), or classification in statistics, which can reveal hidden relationships and accurate data [1, 2].

The application of MD is in health information systems, in the public and private spheres, which, through a process of selection, preprocessing, and data transformation, one can discover patterns and generate knowledge through their interpretations. With this method, the health professional will identify,

characterise, and guide the patient based on patterns of health problems and care therapies for different diseases [2].

As one of the most prevalent diseases and considered a public health problem, we have neurocognitive disorders associated with the human immunodeficiency virus (HAND) because of the virus compromising the immune system. Also, it has an affinity for the central nervous system (CNS) by destroying macrophages that cross the blood-brain barrier, which cannot be invaded by highly active antiretroviral therapy (HAART) [3, 4].

Among the cognitive disorders associated with the human immunodeficiency virus (HIV), dementia is consid-

ered its most serious type, as it interferes with the motor, behavioural, and intellectual functions of the infected individual, causing him the inability to perform instrumental activities of daily living, whose symptoms are manifested by irritability, emotional lability, memory loss, dysdiadochokinesia, and spastic gait [5, 6].

The health professional, despite carrying out the clinical history, neurological, and cognitive examination, sometimes feels undecided as to the therapeutic approach to be taken in the face of a possible neurological complication, since the onset of HIV dementia is insidious and initial symptoms may be confused with other psychiatric illnesses, such as depression and anxiety [7].

In this context, the use of technological strategies, such as information and communication technologies (ICT), can be considered essential tools for detecting the early diagnosis of dementia and other diseases. Computer science methodologies with ICT may be able to classify the main characteristics of the disease, in the face of the patient's social and clinical changes, in a practical way and serve as an aid for health professionals in promoting information management and quality safe care [8].

Among the computational techniques, there is the use of algorithms that consists of the understanding and structure of existing data and capable of generating predictive rules, whose results, when combined, can bring benefits in reducing costs and in the effectiveness of interventions, both in treatment and in preventive actions [9].

Thus, the following question arises: can a data mining technique identify the most relevant characteristics given the risk of dementia in people with HIV/AIDS?

Early recognition and categorisation of the characteristics that influence the risk of dementia will allow the ability to reverse the disease to choose an effective HAART and additional treatments through better social and psychological support to improve the person's life expectancy and quality life with HIV/AIDS. It will also assist the health professional by offering more precise and targeted guidance to family members in the face of behavioural and cognitive changes of the seropositive individual.

For some years, predictive models have been developed to predict outcomes of interest to the population's health, like computerised image analysis based on precision medicine. Researchers believed that the machine could interpret the analysis of complex and straightforward testing [10].

Besides, contributing to the agility in the interpretation of results by reducing the time of action for urgent cases and increasing the degree of confidence in the diagnoses is an advantage of the technologies applied to health, together with the presence of the health professional during the verification and validation of the results based on clinical, epidemiological correlations, and therapeutic decision [10].

In this context, more flexible artificial intelligence algorithms have been used in predictive modelling of outcomes of health interest due to the high amount of data available to carry out this research. As a result, the algorithms show a better outcome prediction potential because they capture complex relationships in the data and their ability to deal with many predictors [11, 12].

Regarding sexually transmitted infections to affected people, prognostic information related to the risk of death and other opportunistic diseases has become increasingly important for health professionals. Also, researchers and public policymakers are a tool to assist in decision-making regarding disease tracking targeting preventive programs and offering specialised treatments [10].

In this context, the study seeks to identify the algorithm with the best performance capable of revealing the most relevant characteristics in the high risk of developing dementia in people with HIV/AIDS, based on data mining.

2. Methodology

In this research, the principal component analysis (PCA) algorithm was applied to assess the impact of reduced dimensionality and the quality of the solution between the following machine learning algorithms: logistic regression, decision tree, neural network, KNN, and forest random.

Machine learning algorithms have specific functions, such as logistic regression, whose statistical technique describes the behaviour between a binary dependent variable and independent metric or nonmetric variables. It investigates the effect of the variables by which individuals are exposed to the probability of a specific event occurring [10].

The decision tree is a nonparametric method applied for data classification and regression, in which a set of attributes is used to predict an output value. The model is formed by elements with information called nodes, which are read in a descending way, and which starts at the root node, with a higher hierarchical level until it reaches the leaf node that consists of the output, that is, the predicted result [10–12].

Neural network seeks to simulate the processes of a human neural network through models composed of nonlinear elements that operate in parallel, through the environment and ends when the neural network manages to generalise solutions for a class of problem [13].

The KNN algorithm is a learning based on instances stored. As a new sample is presented to the algorithm, a new classification is generated from a distance closest to neighbouring cases. The decision rule is built directly without estimating the densities conditioned to the classes; relative patterns can be classified in the space of characteristics that probably belong to the same class [13–15].

The random forest consists of a set of decision trees generated within the same object, using a classifier aggregation technique built so that its structure is composed at random. Thus, the method combines several decision trees through a vote that classifies the final class to receive the highest votes [2].

The database used for this study was built by collecting data from 270 HIV/AIDS-infected individuals and followed up at the outpatient clinic of a Reference Hospital for Infectious and Parasitic Diseases, in the State of Ceará, Brazil, from January to April 2019. The questionnaire used for the collection was composed of sociodemographic data, clinical health, and risk behaviours; simplified medication adherence questionnaire (SMAQ), Brazilian version to assess adherence to treatment with antiretrovirals; the International Dementia Scale to detect dementia in the HIV patient; the Barthel Index

TABLE 1: Classes and independent variables of people with HIV/AIDS.

Classes	Variables
Sociodemographic profile	Age, gender, marital status, education
History of infectious disease	Transmission route, virus type, years with HIV, hepatitis B, hepatitis C, comorbidities, previous hospitalisation, opportunistic disease, initial viral load, current viral load, initial TCD4, current TCD4
Drug adherence	Antiretroviral, forgets to use medication, careless use of medication, stops drug therapy
Neurological assessment	Heads up, psychomotor speed, memory, construction
Psychosomatic changes	Anxiety, obsession-compulsion, interpersonal sensitivity, depression, summing, psychoticism, phobic anxiety, hostility, paranoid ideation
Daily activities	Food, wear, bath, hygiene, going to the bathroom, intestinal control, bladder control, climbing ladder, get up, wander

and the Lawton and Brody Scale to assess the degree of functional capacity; and finally the Social Support Scale for people living with HIV/AIDS.

The International HIV-Associated Dementia Scale (IHDS) is a screening tool to identify the high or low risk of dementia in HIV patients who do not need a specialist and are not influenced by the individual's education. Also, the use of the scale in clinical practice has a sensitivity of approximately 80% and specificity around 57%, with a cut-off score less than or equal to 10 for HAD diagnosis. IHDS can be used to identify probable cases of HAD in situations where formal neuropsychological assessment is impossible.

Table 1 shows the classes with the most relevant independent variables used in this research:

The Python programming language is versatile and capable of providing good code performance and readability. This language is viable for working with rapid application development methods, reducing waste, and optimising code quality; development techniques "are not repeated" and do not allow ambiguities, and have fast and straightforward software [15].

The study was carried out from the execution of all the machine learning algorithms already mentioned. It sought to analyse the performance of the algorithms from the 104 available characteristics present in the database.

Before executing the PCA, the data were standardised using the mean and standard deviation, according to the following equation:

$$z = \frac{x - \mu}{\sigma}, \quad (1)$$

where μ is the mean and σ is the standard deviation applied in the execution of machine learning algorithms.

Subsequently, the principal component analysis (PCA) algorithm was performed, whose multivariate statistical technique is widely used to reduce dimensions and extract resources. The processing of this technique can be summarised as a set of orthogonal linear transformations of the original variables so that the transforming variables maintain, as much as possible, the information contained in the actual characteristics [1, 16].

TABLE 2: Analysis of machine learning algorithms against the variation of HIV-associated dementia data.

Total features: 104	Logistic regression	Decision tree	Neural network	KNN	Random forest
No PCA					
Accuracy	0.7805	0.7317	0.5732	0.6098	0.8293
Precision	0.7377	0.7451	0.5732	0.6415	0.7797
Recall	0.9574	0.8085	1.0	0.7234	0.9787

The PCA captured 90% of the variation in the data and selected the main characteristic of each of these components without repetition.

In the third moment, the main components were analysed, which captured 80% of the data variation and selected the main characteristic of each one without repetition.

They sorted the components in descending order by the variation of the data they represented. Thus, the first components reflected the most significant relevance and corresponded to most of the data variation.

For each of these executions, we analysed the performance of the algorithms and the respective most representative characteristics.

3. Results and Discussion

Initially, the logistic regression, decision tree, neural network, KNN, and random forest algorithms were obtained for 104 components, with the following results in Table 2.

The total execution time of the algorithms took a maximum of 03 seconds, which corresponded from reading and processing the database to reading the PCA. It should be noted that each algorithm independently took up to 02 seconds to obtain the results.

Through the PCA, the number of characteristics was reduced from 104 to 47. Generally, there was a reduction in the performance of the algorithms; random forest, for example, started to reach 80.49% accuracy, and therefore, the number of characteristics was reduced from 47 to 33.

TABLE 3: Analysis of machine learning algorithms against the variation of HIV-associated dementia data.

		Logistic regression	Decision tree	Neural network	KNN	Random forest
Total features: 104						
	Accuracy	0.7805	0.7317	0.5732	0.6098	0.8293
No PCA	Precision	0.7377	0.7451	0.5732	0.6415	0.7797
	Recall	0.9574	0.8085	1.0	0.7234	0.9787
Total features: 47 (PCA = 90%)						
	Accuracy	0.7439	0.7073	0.5732	0.5976	0.8049
90% of captured variation	Precision	0.7097	0.7091	0.5732	0.6296	0.7627
	Recall	0.9362	0.8298	1.0	0.7234	0.9574
Total features: 33						
	Accuracy	0.7195	0.7683	0.5610	0.5976	0.8049
80% of captured variation	Precision	0.6935	0.7917	0.7619	0.6296	0.7541
	Recall	0.9149	0.8085	0.3404	0.7234	0.9787
Total features: 23						
	Accuracy	0.5610	0.6829	0.5854	0.5976	0.8659
70% of captured variation	Precision	0.5679	0.6981	0.9333	0.6296	0.8462
	Recall	0.9787	0.7872	0.2979	0.7234	0.9362
Total features: 16						
	Accuracy	0.7317	0.7073	0.6463	0.5366	0.8049
60% of captured variation	Precision	0.7049	0.7447	0.7500	0.5957	0.7818
	Recall	0.9149	0.7447	0.5745	0.5957	0.9149
Total features: 10						
	Accuracy	0.7195	0.6098	0.5610	0.5122	0.7317
50% of captured variation	Precision	0.7000	0.6667	0.5696	0.5714	0.7358
	Recall	0.8936	0.6383	0.9574	0.5957	0.8298
Total features: 6						
	Accuracy	0.6220	0.5610	0.5366	0.5000	0.6829
40% of captured variation	Precision	0.6111	0.6279	0.5570	0.5600	0.7059
	Recall	0.9362	0.5745	0.9362	0.5957	0.7660

Moreover, the performance of the algorithms remained as in the execution already mentioned, except Logistic regression continued to lose performance.

Logistic regression continued to lose performance. The following steps were performed considering the main components that captured 70%, 60%, 50%, and 40% of data variation, respectively, when observed in Table 3.

Figure 1 shows that 70% captured from 23 characteristics were analysed; the accuracy of the different algorithms showed the best performance of random forest:

From the results obtained in the executions carried out, it is possible to analyse approximately 86% in random forest, 56% logistic regression, 68% decision tree, 60% KNN, and 59% neural network.

Dimensionality reduction could improve the quality of machine learning algorithms. It delimits the less relevant components that will only cause noise in the algorithms or increase the complexity of the problem. However, ignoring elements causes some of the data variation to be lost, leading to reduced performance of the prediction algorithms.

It can be seen in Figure 2 that during the tests, even losing about 30% of the data variation, that is, when considering only 23 characteristics, the random forest algorithm obtained better results regarding the precision and accuracy of 84% and 86%, respectively. In different areas of study, random forest has obtained good results and demonstrated that these algorithms had provided superior results in the prediction capacity compared to other machine learning algorithms. Moreover, it can analyse from samples considered small like this one study until analysing complex data structures [17, 18].

It is also noted that when using less than 23 characteristics, the algorithm loses performance, which assumes that essential aspects may be being disregarded. Reducing the dimensionality of PCA has reduced the complexity of the data and filters out the crucial characteristics to classify whether a patient is at risk of developing dementia.

Thus, the main characteristics extracted with this technique are described below, in an increasing order of representativeness of the data: paranoid ideation; walking; adherence to antiretroviral medication; having the impression that

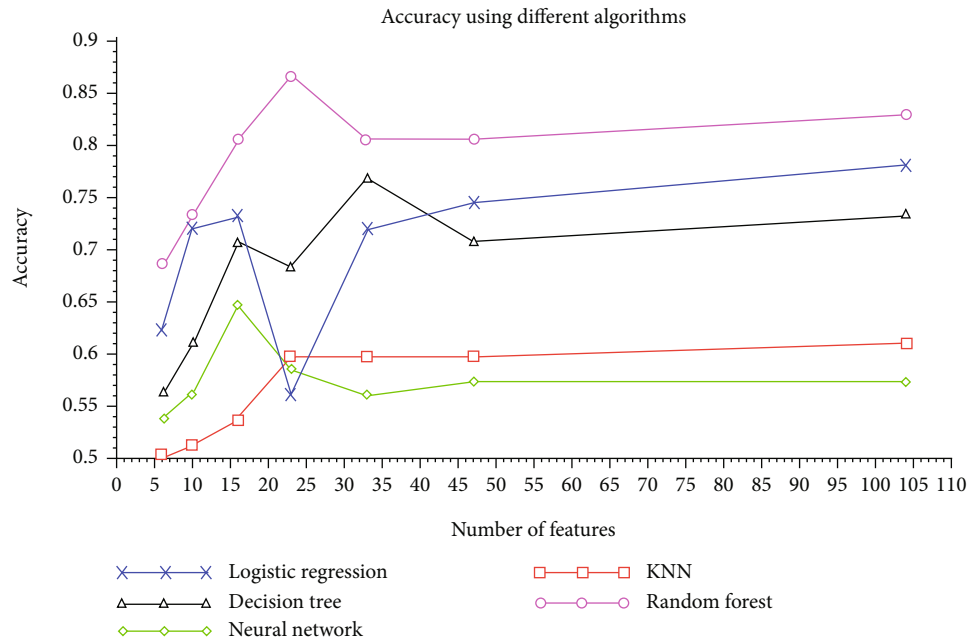


FIGURE 1: Accuracy of machine learning algorithms regarding HIV-associated dementia characteristics.

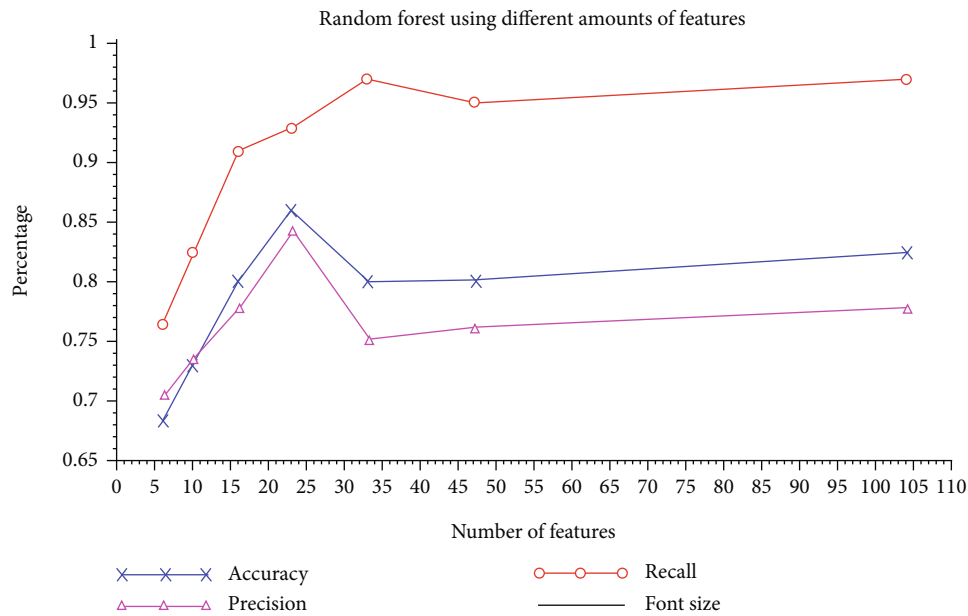


FIGURE 2: Precision ratio, accuracy, and recall of random forest using different amounts of features.

something is wrong in your head or your spirit; having terror or panic attacks; TCD4+lymphocyte values at the first consultation; HIV infection time; type of virus; previous hospitalisation due to HIV infection; anxiety; coinfection with the hepatitis B virus; does not use antiretroviral medication; feeling of sadness and loneliness; interruption of drug treatment; Barthel category; age; use of antiretroviral medication; viral load at the first visit; hopeless about the future; the presence of opportunistic diseases; having the idea that other people are to blame for most of their problems; bladder control; and schooling.

Psychopathological symptoms can be assessed by the Brief Symptom Inventory indicator, whose nine dimensions are elements of psychopathology, such as somatisation that reveals complaints from cardiovascular to gastrointestinal; obsession-compulsion, obsession-compulsion is a symptom of the individual's impulsive behaviour. Interpersonal sensitivity, which turns to the feeling of inferiority and shyness; depression; anxiety, which can be assessed through nervous manifestations and apprehension; hostility; phobic anxiety; paranoid ideation, whose behaviour is considered harmful and disturbing; and psychoticism, in which the individual

TABLE 4: Demographic characteristics of people with HIV.

Gender	Female	91
	Male	179
Age	50 to 55	133
	56 to 60	75
	61 to 70	52
	>70	10
Marital status	Not married	114
	Married	71
	Widowed	43
	Divorced	42
	Incomplete fundamental	89
Education	Completed elementary school	46
	High school	61
	Higher	29
	Illiterate	45

isolates himself and has schizophrenic behaviours, through hallucinations [19, 20].

Phobic conditions such as terror or panic attacks may require the individual to overestimate the danger offered by the object or event in question since the discovery of HIV infection, as well as the infection's relationship with certain mental illnesses, given the neuroinflammatory factors and the stress that contribute to this situation [19–21].

Depression is one of the most common comorbidities in HIV-positive patients, causing worsening of medication adherence and increased mortality. Clinical findings demonstrate a relationship between depression and increased levels of inflammatory cytokines (TNF-alpha, IL-1beta, and IL-6) in both systemic and CSF terms [19–21].

Among the feelings of sadness and loneliness related to a depressive condition, it can be associated with factors other than disease progression. The depressive feeling persists when the subject reveals a lack of hope for the future, thereby affecting adherence to antiretrovirals [21].

The low antiretroviral adhesion directly implies the cognitive damages caused by the high viral load and the degree of immunosuppression of the infected individual, whose determinant most susceptible to HAND corresponds to the nadir values of $LTCD4 + <350 \text{ cells/mm}^3$ or $LTCD4 + \text{current} < 350 \text{ cells/mm}^3$ [22, 23].

Also, it is emphasised about opportunistic neurological diseases of HIV, such as neurotoxoplasmosis, tuberculous meningitis, and neurocryptococcosis, which are essential comorbidities for psychiatric damage. Neurotoxoplasmosis is the most frequent opportunistic infection in the CNS in HIV-positive individuals, causing a lesion with a more common mass effect, especially when the patient has $CD4 + <200 \text{ cells/mm}^3$ [22–26]. Table 4 presents the demographic characteristics of people with HIV according to the research covered.

The predominant sociodemographic characteristics in this research were people with HIV (male) (66.29%), aged

between 50 and 55 years old (49.25%), single (53.33%), and with a low level of education (33%).

Regarding age, scientific evidence reveals that the proportion of older adults living with HIV in the Netherlands in 2030 will reach 73%. In the United States of America (USA) and Brazil, it currently exceeds 50%. Among the statistics, half of HIV infections in older women are acquired at approximately 56 years. Thus, age can be considered significant when associated with HIV, given the high probability of late diagnosis and short-term mortality [27].

In terms of schooling, theoretical studies investigate the relationship between the level of knowledge and its influence on health behaviour in the awareness of self-perceived risk, continued condom use during sexual intercourse, and adherence to antiretroviral drug therapy [28].

Thus, people with a low level of education need social support and continuous monitoring by health professionals regarding the assessment of precession, which is permanently monitored in the evaluation of the information they understand and its impact on the quality of life in the dimensions: social, cultural, psychic, and clinical [28].

In this way, the practical and early diagnosis of these neuropsychological conditions and the correct therapeutic intervention can delay or even reverse dementia associated with HIV and benefit the individual.

4. Conclusion and Future Works

Because of the database used, it was possible to classify and analyse the main characteristics of dementia in people with HIV/AIDS using techniques such as PCA and MD. Through computational tests between the MD algorithms, the random forest algorithm proved to be more effective concerning the others, obtaining higher precision and accuracy.

In addition to dementia, this research could be insightful about other mental disorders in the face of initial symptoms of depression, panic attacks, and feelings of persecution. Thus, it may be able to sensitise health professionals to a more accurate assessment of factors related to the individual's personality, which may be associated with medication adherence and, consequently, with the value of viral load and TCD4 lymphocytes.

Thus, identifying the main characteristics of prediction of dementia due to HIV can assist the health professional for referral and appropriate treatment to these infected people through interpersonal, cognitive-behavioural, and supportive psychotherapy. In addition, orientation contributing to the family about adherence to antiretroviral therapy monitoring, social life, and instrumental activities by the HIV-positive individual.

Since the results of this application are in a local population, it is suggested that these algorithms be applied to more massive databases to confirm and improve their performance. Although the sample is small, its data is independent and distributed equally about those observed in the future and follow the same distribution of the population from which the sample was taken.

Data Availability

The data [database] used to support the conclusions of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

The authors are thankful to the Foundation of Support to the Scientific and Technological Development of the State of Ceara (FUNCAP) for the support received on this project. The sixth author thanks the Coordination for the Improvement of Higher Education Personnel (CAPES). Prof. Marcial Porto Fernandez is funded by the DT, CNPq, Brazil (grant number 313231/2019-2).

References

- [1] G. Blazhenets, Y. Ma, A. Sørensen et al., “Principal components analysis of brain metabolism predicts development of Alzheimer dementia,” *Journal of Nuclear Medicine*, vol. 60, no. 6, pp. 837–843, 2019.
- [2] S. Tabib and D. Larocque, “Non-parametric individual treatment effect estimation for survival data with random forests,” *Bioinformatics*, vol. 36, no. 2, pp. 629–636, 2020.
- [3] P. S. Graciela, “Trastornos neurocognitivos en pacientes con VIH datos preliminares de una cohorte prospectiva uruguaya,” *Revista Medica Del Uruguay*, vol. 35, no. 3, pp. 171–180, 2019.
- [4] M. Airoldi, A. Bandera, D. Trabattoni et al., “Neurocognitive impairment in HIV-infected naïve patients with advanced disease: the role of virus and intrathecal immune activation,” *Clinical & Developmental Immunology*, vol. 2012, article 467154, pp. 1–5, 2012.
- [5] R. E. A. Perdigão, P. d. F. Bonolo, M. R. Silveira, D. I. d. Silva, and M. d. G. B. Ceccato, “Oportunidade de vinculação de pessoas vivendo com HIV em um serviço especializado de saúde, Belo Horizonte (MG),” *Revista Brasileira de Epidemiologia*, vol. 23, article e200020, 2020.
- [6] D. Saylor, A. M. Dickens, N. Sacktor et al., “HIV-associated neurocognitive disorder – pathogenesis and prospects for treatment,” *Nature Reviews Neurology*, vol. 12, no. 4, pp. 234–248, 2016.
- [7] D. M. V. Deus and C. C. Possas, “Atualizações sobre demência cortical em indivíduos infectados pelo HIV,” *Humanae*, vol. 12, no. 1, pp. 1–11, 2018.
- [8] F. R. V. Silveira and L. Y. M. R. Moreira, “Use of machine learning algorithms in the prediction of *Aedes aegypti* transmitted arboviruses,” *Conexoes - Ciencia e Tecnologia*, vol. 14, no. 1, pp. 64–71, 2020.
- [9] I. Tamanini, P. R. Pinheiro, and C. N. dos Santos, “A hybrid approach of verbal decision analysis and machine learning,” *Lecture Notes in Computer Science*, vol. 7413, no. 1, pp. 126–131, 2012.
- [10] K. M. Santos, J. R. Ferreira Júnior, D. T. Wada, A. P. M. Tenório, M. H. N. Barbosa, and P. M. d. A. Marques, “Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine,” *Radiologia Brasileira*, vol. 52, no. 6, pp. 387–396, 2019.
- [11] Z. Obermeyer and E. J. Emanuel, “Predicting the future: big data, machine learning, and clinical medicine,” *The New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016.
- [12] C. M. Maia, J. C. M. Gomes, and L. D. Chagas, “Estudo sobre o uso de árvores de decisão na área da saúde,” *Anais do Encontro de Computação do Oeste Potiguar ECOP/UFERSA*, vol. 1, pp. 23–30, 2017.
- [13] M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, and E. Liu, “Predicting all-cause risk of 30-day hospital readmission using artificial neural networks,” *PLoS One*, vol. 12, no. 7, article e0181173, 2017.
- [14] A. R. Olivera, V. Roesler, C. Iochpe et al., “Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes - ELSA-Brasil: accuracy study,” *Sao Paulo Medical Journal*, vol. 135, no. 3, pp. 234–246, 2017.
- [15] I. R. S. Silva and R. O. Silva, “Python programming language,” *Revista Tecnologias em Projeção*, vol. 10, no. 1, p. 62, 2019.
- [16] M. M. Faria, *Detecção de intrusões em redes de computadores com base nos algoritmos KNN, K-Means++ e J48*, Dissertação (Programa de Mestrado em Ciência da Computação)— Faculdade Campo Limpo Paulista—FACCAMP, São Paulo, 2016.
- [17] A. Géron, *Hands-On Machine Learning with Scikit-Learn and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O’Reilly Media, Inc., 2017.
- [18] S. Raschka and V. Mirjalili, *Python Machine Learning*, Packt Publishing Ltd, 2017.
- [19] L. Derogatis, *The Brief Symptom Inventory (BSI): Administration and Procedures—Manual*, Baltimore, Clinical Psychometric Research, 1982.
- [20] J. P. Sanabria-Mazo, P. A. Hoyos-Hernández, and F. Bravo, “Psychosocial factors associated with HIV testing in Colombian university students,” *Acta Colombiana de Psicología*, vol. 23, no. 1, pp. 158–168, 2020.
- [21] B. A. Ayele, W. Amongne, and L. Gemechu, “HIV-associated neurocognitive disorder and HIV-associated myelopathy in a patient with a preserved CD4, but high viral load—a rarely reported phenomenon: a case report and literature review,” *BMC Infectious Diseases*, vol. 20, no. 1, p. 574, 2020.
- [22] R. K. Heaton, D. R. Franklin, R. Deutsch et al., “Neurocognitive change in the era of HIV combination antiretroviral therapy: the longitudinal charter study,” *Clinical Infectious Diseases*, vol. 60, no. 3, pp. 473–480, 2015.
- [23] A. R. S. Serrano-Villar, F. Gutierrez, and E. S. C. Mirall, “HIV as a chronic disease: evaluating and managing non-AIDS defining,” *Management of HIV-Associated Disease in HIV*, vol. 3, no. 2, pp. 1–15, 2016.
- [24] N. Sacktor, “Changing clinical phenotypes of HIV-associated neurocognitive disorders,” *Journal of Neurovirology*, vol. 24, no. 2, pp. 141–145, 2018.
- [25] L. A. Camargo, C. G. Capitão, and E. M. V. Filipe, “Saúde mental, suporte familiar e adesão ao tratamento: associações no contexto HIV/AIDS,” *Psico-USF*, vol. 19, no. 2, pp. 221–232, 2014.
- [26] Y. Ouyang, L. Liu, Y. Zhang et al., “Discordant patterns of tissue-specific genetic characteristics in the HIV-1 env gene from HIV-associated neurocognitive disorder (HAND) and non-HAND patients,” *Journal of Neurovirology*, vol. 20, no. 4, pp. 332–340, 2014.

- [27] G. E. C. P. Cruz, D. F. B. Cardoso, E. S. da Silva, R. C. P. Silveira, A. E. Silva, and J. L. A. Apostolo, "Late diagnosis of human immunodeficiency virus and acquired immunodeficiency syndrome in the elderly: scoping review protocol," *Enfermería Actual de Costa Rica*, no. 38, pp. 292–299, 2020.
- [28] R. R. de Freitas Magalhães Gomes, M. das Graças Braga Ceccato, L. R. F. S. Kerr, and M. D. C. Guimarães, "Factors associated with low knowledge on HIV/AIDS among men who have sex with men in Brazil," *Cadernos de Saúde Pública*, vol. 33, no. 10, 2017.

Research Article

A Protocol for the Diagnosis of Autism Spectrum Disorder Structured in Machine Learning and Verbal Decision Analysis

Evandro Andrade, Samuel Portela, Plácido Rogério Pinheiro , Luciano Comin Nunes , Marum Simão Filho , Wagner Silva Costa, and Mirian Caliope Dantas Pinheiro 

University of Fortaleza (UNIFOR)-Graduate Program in Applied Informatics, Av. Washington Soares, 1321-BL J Sl 30-60.811-905 Fortaleza, Brazil

Correspondence should be addressed to Plácido Rogério Pinheiro; placido@unifor.br

Received 3 July 2020; Revised 16 February 2021; Accepted 17 March 2021; Published 31 March 2021

Academic Editor: Martti Juhola

Copyright © 2021 Evandro Andrade 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.

Autism Spectrum Disorder is a mental disorder that afflicts millions of people worldwide. It is estimated that one in 160 children has traces of autism, with five times the higher prevalence in boys. The protocols for detecting symptoms are diverse. However, the following are among the most used: the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5), of the American Psychiatric Association; the Revised Autistic Diagnostic Observation Schedule (ADOS-R); the Autistic Diagnostic Interview (ADI); and the International Classification of Diseases, 10th edition (ICD-10), published by the World Health Organization (WHO) and adopted in Brazil by the Unified Health System (SUS). The application of machine learning models helps make the diagnostic process of Autism Spectrum Disorder more precise, reducing, in many cases, the number of criteria necessary for evaluation, denoting a form of attribute engineering (feature engineering) efficiency. This work proposes a hybrid approach based on machine learning algorithms' composition to discover knowledge and concepts associated with the multicriteria method of decision support based on Verbal Decision Analysis to refine the results. Therefore, the study has the general objective of evaluating how the mentioned hybrid methodology proposal can make the protocol derived from ICD-10 more efficient, providing agility to diagnosing Autism Spectrum Disorder by observing a minor symptom. The study database covers thousands of cases of people who, once diagnosed, obtained government assistance in Brazil.

1. Introduction

In a society where competitiveness is a substantial differential in profile and productive capacity, any severe psychological disorder that an individual may present constitutes a factor of their exclusion from the wealth-generating process. In this context, individuals with Autism Spectrum Disorder (ASD), as well as other severe psychological disorders, are observed with reticence from an early age and, often, discriminated against in daycare centers, schools, universities, and companies where they try to get an insert. A probable cause for this discrimination is based on the lack of information about ASD and other psychological disorders, as well as on the lack of conscience of the potential contribution that every human being can give to society, especially when there is an appreciation of the plurality of ideas and points of view.

According to the World Health Organization (WHO), ASD is a mental disorder that affects more than 70 million people worldwide. Worldwide estimates show one child with autism every 160 [1], with a prevalence five times higher among boys. In Brazil, there is a ratio of one autistic child for every 360. Although considered underestimated, Brazil's data show that there is a significant demand for specialized care [2]. Currently, the need for effectiveness and efficiency in diagnosing ASD goes beyond just health since people diagnosed with this disorder also seek to participate in selective processes for the formation of work teams. These diagnoses can help these people in both the educational and professional placement in the job market. On the other hand, the globalized world requires competitiveness and sustainability from companies, which depend heavily on people. In their selection processes, companies seek to select healthy

professionals and, often, have excluded those who suffer from specific problems, such as ASD, even though they have different skills.

As a social contribution, the present study is aimed at applying a machine learning model to make the ASD diagnostic process more precise and concise, reducing the number of criteria necessary for the evaluation. To this end, the study proposes a hybrid approach, based on the juxtaposition of machine learning algorithms, aimed at knowledge discovery, associated with a multicriteria method of decision support based on Verbal Decision Analysis (VDA). The general objective is to evaluate how the hybrid methodology now proposed can make the protocol derived from the International Classification of Diseases, 10th edition (ICD-10) more efficient, enabling the diagnosis of Autism Spectrum Disorder by observing a minor symptom. The study sought to link the ICD-10 with the codes and structure of the DSM-5, aiming at expanding the understanding and diagnosis capacity from different clinical points of view. This objective comes considering that the hybrid model proposal mentioned above is applied to ASD and considering that mental disorders are the main focus of the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5).

Moreover, to validate the hybrid model, the study used a database provided by the Social Security Technology and Information Company of Brazil-DATAPREV. Mentioned data refer to thousands of cases of people who obtained the assistance benefit related to ASD, as provided by Brazil's federal legislation. The granting of these benefits goes through evaluating each case of children with ASD, to whom the Benefit of Continuous Provision (BPC) program applies. These assessments have an extensive questionnaire divided into Social Assessment and Medical-Expert Assessment, which allowed for research to obtain more consistent data on the stages of medical and social evaluation.

This article is divided into five sections. After the introduction, the study presents the second section that outlines the systematization of the fundamental theoretical basis on which the work is based. The third section describes the methodology used, which is subdivided into four subsections—which deal, respectively, with the database, the Random Forest machine learning algorithm, the applied learning framework, and the Verbal Decision Analysis (VDA) method addressed. The fourth section presents the results achieved. Finally, the fifth section concludes the work and proposes actions and perspectives for future work.

2. Theoretical Reference

This section contains some important topics about Autism Spectrum Disorder and highlights proposed information technology solutions used to support the diagnosis and referral of that disorder.

2.1. Highlights about Autism Spectrum Disorder Concepts. Autism Spectrum Disorder (ASD) consists of a neurodevelopmental disorder characterized by patterns of stereotyped and repetitive behaviors and difficulty in communication and social interaction. There are several protocols for diag-

nosing the disease. The most widely adopted are the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition, known as DSM-5, from the American Psychiatric Association; the Revised Autistic Diagnostic Observation Schedule (ADOS-R); the Autistic Diagnostic Interview (ADI); and the International Classification of Diseases and Health-Related Problems (ICD-10), a criterion published by the World Health Organization (WHO) and adopted in Brazil by the Unified Health System (SUS). In Figure 1, there is a highlight of the Autism Spectrum Disorder (ASD) in the context of the Neurodevelopmental Disorders categorization, as discriminates in the DSM-5.

Scientific investigations about Autism Spectrum Disorder have evolved a lot since the beginning of the last century, in both the theoretical and empirical aspects. From this perspective, it is clear that since the term “autism” was coined by the psychiatrist Eugen Bleuler, in 1911, when analyzing symptoms of schizophrenic patients, including the disclosure of other autistic patterns by Hans Asperger, in 1970, research has advanced considerably. With the passage of time and the deepening of knowledge on the subject, the expression Autism Spectrum Disorder (ASD) started to feature a very different set of behavioral changes with early onset, chronic course, and variable impact in multiple areas of development [4].

Furthermore, Martone and Santos-Carvalho [5] performed a bibliographic review of the articles published by the *Journal of Applied Behavior Analysis* (JABA) between the years 2008 and 2012, focusing on the importance of verbal behavior for the detection of the disorder [5]. Cohen et al. [6] present a pioneering study in which it uses a neural network to classify the disorder. The results showed the importance of implementing new technologies in the detection of the disorder and, simultaneously, pointed to the need to better explore the application of neural networks, with all its variations of parameters, to improve the classification of ASD [6].

2.2. Information Technology Solutions Used in the Diagnosis of Autism Spectrum Disorder. Following the technological approach, Yekta et al. [7] proposed an ASD diagnostic system, called the Autism Screening Expert System (ASES), whose main attributes or criteria were chosen from the use of machine learning techniques on a basis obtained by questionnaires. In this case, to select the diagnosis's main attributes, the technology solution “Random Forests and Support Vector Machines” was used. This system seems to be very efficient for detecting the disorder in children aged 2 to 6 years [7]. In a similar approach, Khullar et al. [8] developed a system—the Handheld Expert System (HES)—to diagnose ASD using artificial neural networks. It obtained 100% (one hundred percent) accuracy in detecting the disorder when considering the protocol described in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [8]. Also, in this same line, Nunes et al. [3] present an intelligent solution based on a hybrid model of the expert system associated with the multicriteria method MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) to support the early diagnosis of ASD in children aged up to two years [3].

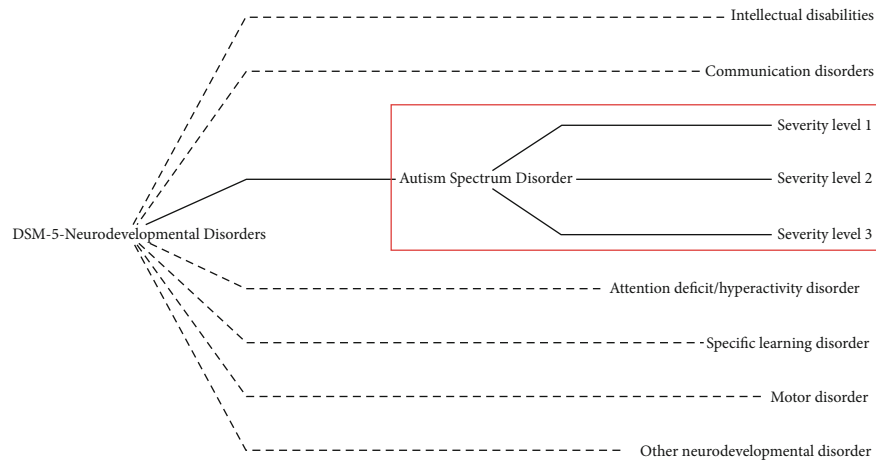


FIGURE 1: Neurodevelopmental Disorders, according to DSM-5. Source: Nunes et al. [3].

In a similar context, Cohen [9] presents a study of a neural network that simulates and shares formal qualitative similarities with selective attention and the generalization of deficits observed in people with autism. In simulations in which the model was taught to discriminate children with autism from children with mental retardation, there were observations that neuropathologies described in the literature are sufficient to explain some of the unique pattern recognition and learning skills, as well as their generalization and acquisition problems [9]. Florio et al. [10] show the importance of a second opinion on the diagnosis of ASD and the use of technology to provide independent clinical evaluation, increasing diagnosis accuracy. Challenges related to using technology to offer independent second opinions include the possibility for individuals and relatives to participate in self-diagnosis activities [10].

Still, from the perspective of intelligent systems, Thabtah [11] points out that ASD has been studied in the area of the behavioral sciences using intelligent methods based on machine learning to speed up the screening time or improve the sensitivity, specificity, or precision of the diagnostic process. In this sense, machine learning considers the problem of ASD diagnosis as a classification task, by which predictive models are built based on historical cases and controls. These models must be connected to a screening tool to achieve one or more of the objectives. This work sought to shed light on recent studies that use machine learning in the ASD classification to discuss its pros and cons. It highlighted a noticeable problem associated with current ASD screening tools: the reliability of these tools using the DSM-IV rather than the DSM-5 manual. As a result, the need to change the current screening tools to reflect the newly imposed criteria for ASD classification in the DSM-5 has been suggested, particularly the diagnostic algorithms incorporated in these methods [11].

In the same sense, Thabtah [12] states that machine learning is a multidisciplinary research topic that employs intelligent techniques to discover relevant hidden patterns used in forecasting to improve decision-making. Thus, some machine learning techniques, such as Random Forests and others, were applied to datasets related to autism to build pre-

dictive models. These models claim to increase clinicians' ability to provide robust ASD diagnostics and prognosis [12].

In a systematic review, Thabtah and Peebles [13] evaluated and critically analyzed 37 different ASD screening tools to identify possible areas that need to be addressed through further development and innovation [13].

In another analysis, Thabtah et al. [14] proposed a new machine learning framework related to screening autism for adults and teenagers containing vital characteristics. The results obtained reveal that the machine learning technology was able to generate classification systems with acceptable performance in terms of sensitivity, specificity, and precision, among others [14].

On the other hand, in a different approach, Neto et al. [15] presented the G-ASD, which consists of a prototype of a game to assist professionals in the field of psychology using the methodology of Applied Behavior Analysis (ABA) in teaching children's learning with Autism Spectrum Disorder (ASD) [15]. The G-ASD shows how important the use of games is in the process of building the knowledge of children with autism, assisting the professional in teaching colors to autistic children. From a social and pedagogical perspective, Lampreia [16] evaluates the official instruments for the diagnosis of autism and comments that, due to the intention of unifying them, it harms some areas such as social interaction, communication, interests, and development activities. Costa et al. [17] address the difficulty of including autistic people in the information technology segment and highlight information for analysis and discussion regarding the inclusion of autistic people in higher education and the respective job market in the area of information technology.

Several approaches are promoted to study neurological disorders. Maenner et al. [18] investigated the Random Forest algorithm in a set of autism data from the Georgia Autism and Developmental Disabilities Monitoring (ADDM) network, using phrases and words obtained in child development assessments. The dataset consists of 5,396 assessments for 1,162 children, of which 601 are on the spectrum. Random Forest classifiers were assessed on an independent test dataset that contained 9,811 evaluations from 1,450 children.

The results reported that the Random Forest reached about 89% of the predictive value [18].

Finally, it is observed that the Random Forest algorithm has been used with an outstanding frequency in decision-making processes in the health area, mainly for ASD diagnosis and treatment. This was one of the main reasons for using the Random Forest in the present study as an algorithm that is part of the hybrid model now proposed.

3. A Methodology for the Autism Spectrum Disorder Diagnostic Protocol

The present study used the quantitative research method, looking for numbers, frequency of characteristics, and quantifiers in the responses to social and medical assessments of children diagnosed with ASD, candidates for the Benefit of Continuous Provision (BPC) of Social Security in Brazil. Also, a qualitative approach was used to interpret the results of the evaluations in the Brazilian medical-social context.

The study database was obtained from the Social Security Technology and Information Company of Brazil-DATAPREV. The said database has 320,302 (three hundred twenty thousand three hundred two) records referring to the responses in the medical evaluations of 3,861 (three thousand eight hundred sixty-one) children from 0 to 5 years old, from all over Brazil, representing 82 (eighty-two) social, environmental, and clinical variables. This base is fed from evaluation forms, whose items can be transformed into labels for the variables that characterize diseases. By mapping the main characteristics, it is possible to assign a response to each question. The variables include information about the person's identity, the socioenvironmental aspects, and the clinical criteria necessary to identify ASD. The process of preparing the model proposal can be summarized in the following steps:

- (i) In the first stage, after collecting the responses in the evaluations, the characteristics were organized with keywords (tag) and related to the respective qualifiers in a data table (Analytical Base Table (ABT))
- (ii) Then, in the second stage, the data were classified using the machine learning algorithm, Random Forest, which combined several decision trees to obtain a prediction with greater accuracy and stability
- (iii) The third stage consisted of the creation of a form with the nine most used characteristics for the application by the decision-makers
- (iv) Finally, in the fourth stage, the ZAPROS-IIIi method, implemented in the ARANAÚ tool [19], performed the Verbal Decision Analysis (VDA), aiming at the order of preference of the main characteristics of the ASD

These characteristics constitute the proposed model.

Figure 2 presents a summary of the process of the steps described above, providing a simplified view of the creation of the proposed hybrid model mentioned.

3.1. Preparation for Processing. There are two preprocessing steps. The first one was developed to appropriate the data for manipulation in the Orange framework, which uses the machine learning model (see step I and step II, in Figure 2). The second one was developed, from the first preliminary result, to adjust the information for the Verbal Decision Analysis (see step III, in Figure 2).

In the first preprocessing stage, the questionnaire, received in a .txt extension, has the questions classified and labeled. After transposition, they are arranged in tabular form, in a panel format. The answers are mapped and assigned to the item, with a specific weight, ranging from 0 to 4. The resulting file has the extension .csv and is consumed by the machine learning framework.

In the second preprocessing step, the result of the model imported will be the shortest path in the decision tree, in which nodes have represented the symptoms of the disorder investigated. Such a way will again be converted into a questionnaire, with each sign being a variable that is the subject of a question, with 3 (three) evaluation criteria, each with 3 (three) possible qualitative answers. The result will serve as input to the Verbal Decision Analysis method, ZAPROS-IIIi, implemented in the ARANAÚ tool [19].

3.2. Random Forest Method Considerations. Random Forests are randomized decision trees capable of predicting expected values of regressions and classifications. The model was published by Breiman [20, 21], and since then, the technique has become an important data analysis tool, quite versatile and with few adjustment parameters. In this sense, despite its simplicity, the method is generally recognized for its precision and ability to handle small sample sizes, large resource spaces, and complex data structures [22]. The versatility of the algorithm contributed to the popularity of Random Forests, in addition to having few parameters to adjust. Many practical issues have successfully involved Random Forests, including forecasting air quality, chemoinformatics, ecology, 3D object recognition, and bioinformatics, to name a few. Many have validly proposed variations of the original algorithm to improve calculation times while maintaining good forecasting accuracy. Breiman's Random Forests were also extended to quantile estimation, survival analysis, and ranking forecast [22].

Regarding the theoretical aspect, the analyses are less conclusive, and, regardless of their extensive use in practical contexts, little is known about the mathematical properties of Random Forests. To date, most studies have focused on isolated parts or simplified versions of the procedure. The most famous theoretical result is that of Breiman [20, 21], which offers an upper limit to the error of generalization of forests in terms of correlation and strength of individual trees. It was followed by a technical note [22], which focuses on a modified version of the original algorithm. A critical step was subsequently followed by Jeon and Lin [23], who established lower limits for nonadaptive forests, independent of the training set [22]. They also highlighted an interesting connection between Random Forests and a specific class of predictors of k -nearest neighbors (KNN) that was further developed [22].

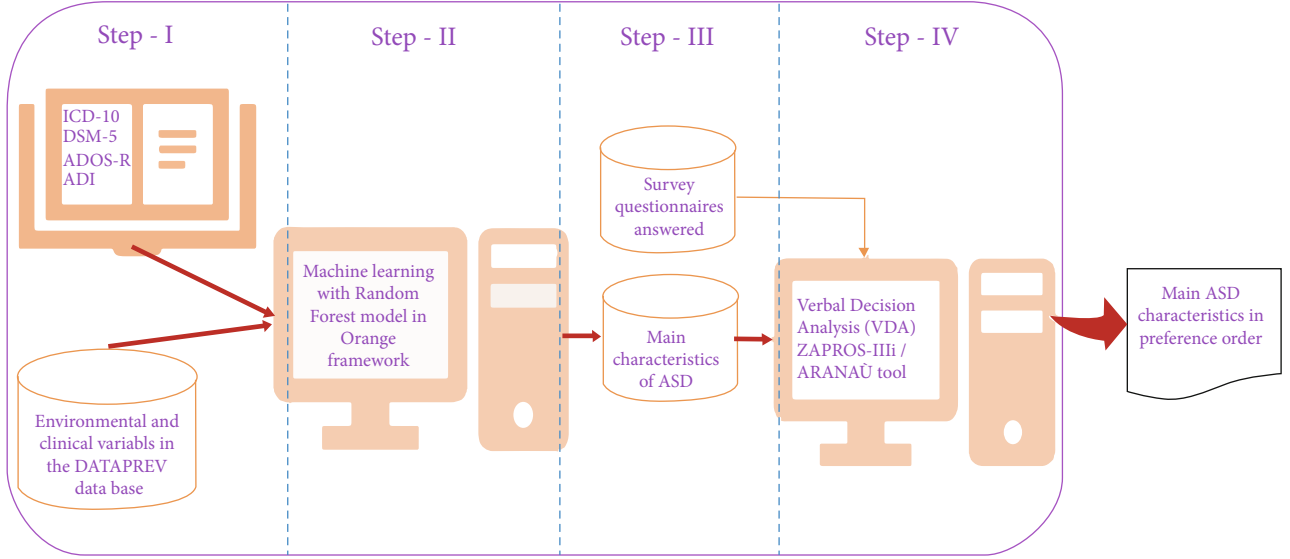


FIGURE 2: The graph with the processing structure of the proposed hybrid model. Source: formatted by the author.

The nature of the procedure can explain the difficulty in adequately analyzing Random Forests, the former being a subtle combination of different components. Among the essential ingredients of the forest, both bagging (Breiman [24, 25]) and classification and regression trees (CART) criterion (Breiman et al. [26, 27]) play a critical role. Bagging, like a bootstrap-aggregating contraction, is a general aggregation scheme that proceeds by generating subsamples of the original dataset, building a predictor for each resampling, and deciding on the mean. It is one of the most useful computationally intensive procedures to improve unstable estimates, especially for large data and high-dimensional datasets, where it is impossible to find a suitable model in one step due to the complexity and scale of the problem [22]. The most influential CART algorithm by Breiman et al. [26, 27] originated the CART-split selection, used in the construction of individual trees to choose the best cuts perpendicular to the axes. In each node of each tree, the best cut is selected, optimizing the CART division criterion, based on the notion of Gini impurity (classification) and quadratic regression error [22].

However, despite the essentiality of the algorithm components mentioned above, both bagging and CART are challenging to analyze, which is why theoretical studies have considered simplified versions of the original procedure. This analysis is usually done by simply skipping the bagging step and replacing the CART partition selection with a more elementary cutting protocol. Also, in Breiman's Random Forests, each leaf—that is, a terminal node—of the individual trees contains a preestablished fixed number of observations (this parameter is usually chosen between 1 and 5). Thus, the authors opt for simplified and independent data procedures, thus creating a gap between theory and practice [22].

Motivated by the discussion, Scornet et al. [22] studied some asymptotic properties of Breiman's algorithm [20, 21] in additive regression models. These researchers proved the consistency of Random Forests, which provides a first basic theoretical guarantee of efficiency for this algorithm. This

finding was the first consistency result for Breiman's [20, 21] original procedure. The approach was based on a detailed analysis of the cell's behavior generated by the selection of the CART-split as the sample size increased. The study also showed that Random Forests could adapt to a sparse structure when the dimension is large, but only a smaller number of coordinates make the information. Figure 3 describes the Random Forest algorithm [22].

The general structure consists of the regression in which a random input vector $X \in [0, 1]^p$ is observed, where " p " is the dimension of the vector. The objective is to predict the integrable random response square $Y \in R$ by estimating the function of regression $m(x) = E[Y | X = x]$. Therefore, it is assumed that the training sample is given by $D_n = (X_1, Y_1), \dots, (X_n, Y_n)$ in $[0, 1]^p \times R$, independently distributed as the independent pair (X, Y) . The objective is to use the dataset D_n to construct an estimate $m_n : [0, 1]^p \rightarrow R$ of the function " m ." Thus, it is said that an estimated regression function m_n is consistent if $E[m_n(X) - m(X)]^2 \rightarrow \infty$, where the expectation E is greater than X and D_n [22].

A Random Forest is a predictor that consists of a collection of M trees of random regression. For the j -th family tree, the predicted value at query point x is indicated by $m_n(x, \Theta_j, D_n)$, where $\Theta_1, \dots, \Theta_M$ are independent random variables, distributed as a generic random variable Φ and independent of D_n . In practice, this variable is used to resample the training set before the growth of individual trees and to select successive directions for separation. The trees are combined to form the finite estimate of the forest [22].

$$m_{M,n}(x, \Theta_1, \dots, \Theta_M, D_n) = \frac{1}{M} \sum_{j=1}^M m_n(x, \Theta_j, D_n). \quad (1)$$

Since one can choose, in practice, M as large as possible, Scornet et al. [22] demonstrated the property; according to which, the infinite forest estimate obtained as the limit of

Algorithm 1: Breiman's random forest predicted value at x

Input: Training set D_n , number of trees $M > 0$, $m_{\text{try}} \in \{1, \dots, p\}$,
 $a_n \in \{1, \dots, n\}$, $t_n \in \{1, \dots, a_n\}$, and $x \in [0, 1]^p$.

Output: Prediction of the random forest at x .

```

1 for  $j = 1, \dots, M$  do
2   Select  $a_n$  points, without replacement, uniformly in  $D_n$ .
3   Set  $P_0 = \{[0, 1]^p\}$  the partition associated with the root of the tree.
4   For all  $1 \leq l \leq a_n$ , set  $P_l = \emptyset$ .
5   Set  $n_{\text{nodes}} = 1$  and level = 0.
6   While  $n_{\text{nodes}} < t_n$  do
7     if  $P_{\text{level}} = \emptyset$  then
8       level = level + 1
9     else
10      Let  $A$  be the first element in  $P_{\text{level}}$ .
11      if  $A$  contains exactly one point then
12         $P_{\text{level}} \leftarrow P_{\text{level}} \setminus \{A\}$ 
13         $P_{\text{level}+1} \leftarrow P_{\text{level}+1} \cup \{A\}$ 
14      else
15        Select uniformly, without replacement, a subset
16         $M_{\text{try}} \subset \{1, \dots, p\}$  of cardinality  $m_{\text{try}}$ .
17        Select the best split in  $A$  by optimizing the CART-split
18        criterion along the coordinates in  $M_{\text{try}}$  (see details below).
19        Cut the cell  $A$  according to the best split. Call  $A_L$  and  $A_R$ 
20        the two resulting cell.
21         $P_{\text{level}} \leftarrow P_{\text{level}} \setminus \{A\}$ 
22         $P_{\text{level}+1} \leftarrow P_{\text{level}+1} \cup \{A_L\} \cup \{A_R\}$ 
23         $n_{\text{nodes}} = n_{\text{nodes}} + 1$ 
24      end
25    end
26  end
27  Compute the predicted value  $m_n(x; \Theta_j, D_n)$  at  $x$  equal to the average of
28  the  $Y_i$ 's falling in the cell of  $x$  in partition  $P_{\text{level}} \cup P_{\text{level}+1}$ .
29 end
30 Compute the random forest estimate  $m_{M,n}(x; \Theta_1, \dots, \Theta_M, D_n)$  at the query
31 point  $x$  according to (1).

```

FIGURE 3: The Random Forest algorithm. Source: Scornet et al. [22].

equation (1) is verified as follows when the number of M trees grows to infinity [22]:

$$m_n(x; D_n) = E_{\Theta} [m_n(x, \Theta, D_n)], \quad (2)$$

where E_{Θ} denotes the expectation in relation to the random parameter Θ , conditional on D_n . The law of large numbers, which states that it almost certainly depends on D_n , justifies this operation [22]:

$$\lim_{M \rightarrow \infty} m_{n,M}(x; \Theta_1, \dots, \Theta_M, D_n) = m_n(x; D_n). \quad (3)$$

See Breiman [20, 21] for more details. From now on, to simplify the notation, $m_n(x)$ will be written instead of $m_n(x; D_n)$ [22].

In the original forests of Breiman [20, 21], each node of a single tree is associated with a hyperrectangular cell. At each stage of the tree's construction, the set of cells forms a partition $[0, 1]^p$. The root of the tree is $[0, 1]^p$ itself, and each tree grows, as explained in Algorithm 1 in Figure 3. This algorithm has three parameters [22]:

- (1) $m_{\text{try}} \in \{1, \dots, p\}$, which is the number of preselected directions for the tree's dividing
- (2) $a_n \in \{1, \dots, n\}$, which is the number of data points sampled in each tree
- (3) $t_n \in \{1, \dots, a_n\}$, which is the number of leaves in each tree

By default, in the original procedure, the parameter m_{inytry} is set to $p/3$, $a!$ is set to n (resampling is done with substitution), and $t_n = a_n$. However, in this approach, resampling is done without replacement, and the parameters a_n and t_n may differ from their default values [22].

The algorithm works by growing M different trees as follows. For each tree, data points are drawn at random without replacing the original dataset; then, in each cell of each tree, a division is chosen maximizing the CART criterion; finally, the construction of each tree is interrupted when the total number of cells in the tree reaches the value t_n . Therefore, each cell contains exactly one point in the case $t_n = a_n$ [22].

3.3. *Framework Orange Canvas*. Demsar et al. [28] designed a framework called Orange, used in the present work. The

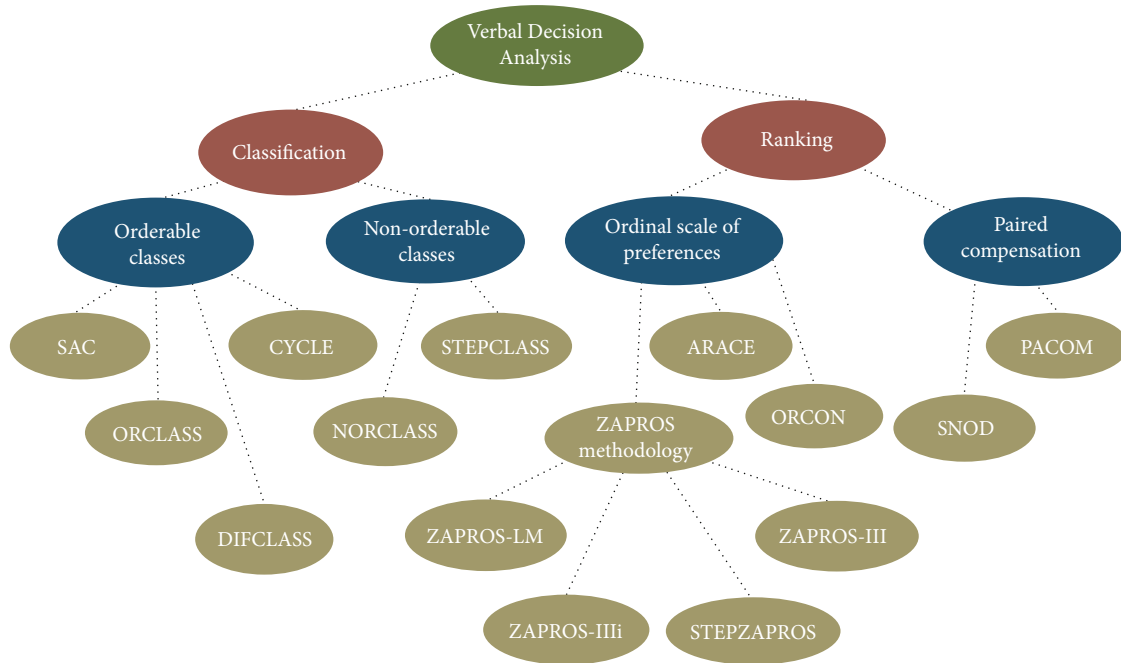


FIGURE 4: Verbal Decision Analysis methodologies. Source: Nunes et al. [3].

framework is a set of machine learning and data mining tools for data analysis through Python scripts and visual programming. Orange is intended for experienced users and programmers, as well as data mining students. It is based on the C++ language; however, it allows developers to work with the Python language.

3.4. *Considerations about the ZAPROS-IIIi Verbal Decision Analysis Method.* Verbal Decision Analysis (VDA) assumes that most decision-making problems can be expressed in natural language and consists of the decision-making process based on the related representation of the problem [29]. It is also noteworthy that this methodology addresses unstructured problems, characterized by the absence of logical and well-defined procedures to be applied to its resolution [29]. These problems are qualitative and are complex to be organized, formalized, and measured numerically. Also, it is not possible to have all the information necessary to resolve them. Thus, the analysis process is also subjective, which requires the collection of information from the decision-maker.

According to Larichev and Moshkovich [29], the methods that make up the structure of Verbal Decision Analysis are ZAPROS-III, ZAPROS-LM, PACOM, and ORCLASS, as well as their characteristics and applications. The analysis of a large amount of data by humans has shown that the correct way of operating is used by the mentioned methods and works as follows [29]:

- (1) Comparison of two assessments on a verbal scale by two criteria
- (2) Assignment of multicriteria alternatives to decision classes

- (3) Comparative verbal evaluation of alternatives according to separate criteria

The mentioned methods are Decision Support Systems (DSS) that help a decision-maker to classify alternatives of multiple attributes. The Verbal Decision Analysis framework’s first three methods are aimed at establishing a ranking of the alternatives in order of preference. The latter presented is the only methodology for classification based on the VDA structure. Figure 4 shows an easy visualization of several Verbal Decision Analysis methodologies according to the types of problems.

For this work, the ZAPROS-IIIi method will be applied. The method fits the characteristics of the problem addressed, which allows the structuring of a decision rule used to compare the alternatives and will not be changed, even if the set of alternatives is modified and applies to problems with a large number of alternatives. The ZAPROS-III method is structured in three well-defined main stages: formulation of problems, elicitation of preferences, and comparison of alternatives. As proposed in the main version of the ZAPROS method, the method is aimed at classifying alternative criteria in scenarios that involve minimal set criteria and criteria values and a large number of alternatives. The relevant criteria and their values for decision-making and the scale of preferences based on the preference of the decision-maker are obtained in the first and second stages. In the last step, the comparison between the alternatives based on the decision-maker’s preferences is performed. These stages are described below.

As part of the Verbal Decision Analysis structure as well, the ZAPROS-III methodology can be considered an evolution of ZAPROS-LM. Similar to the ZAPROS-LM and PACOM methods, this method is aimed at classifying a

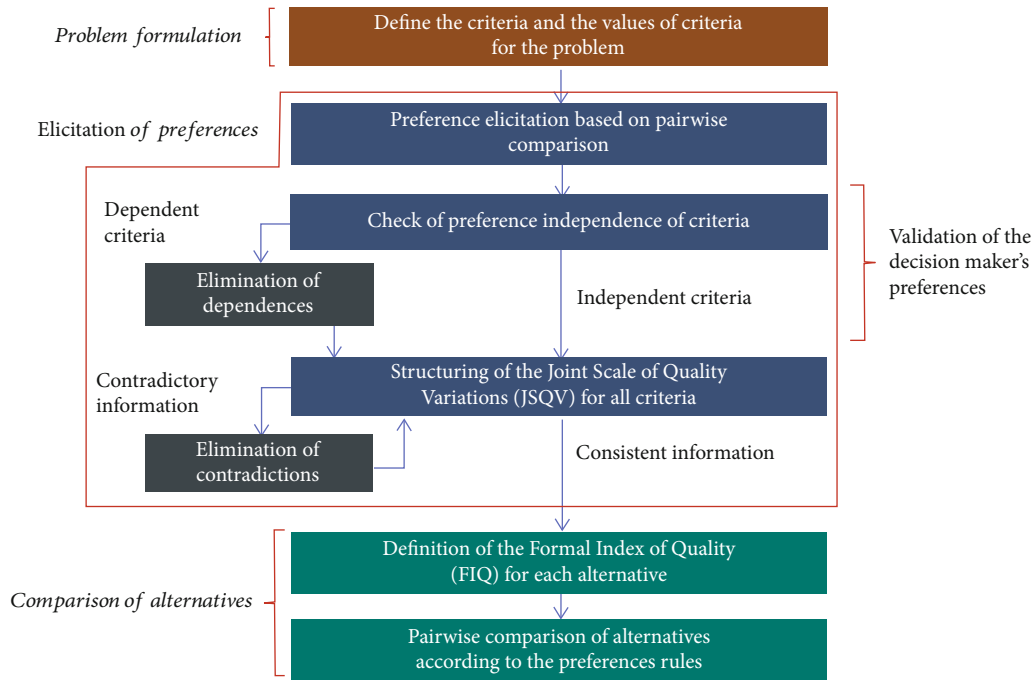


FIGURE 5: Procedure for applying the ZAPROS-IIIi method. Source: Pinheiro et al. [30].

group of alternatives, from the most preferable to the least preferable.

Although ZAPROS-III applies a similar procedure to obtain the preferences of its successor, it implements modifications that make it more efficient and more accurate concerning inconsistencies. The number of incomparable alternatives is substantially less than in the previous ZAPROS.

Figure 5 presents a flowchart with the steps to apply the VDA ZAPROS-IIIi method. According to the scheme described in the procedure, it is possible to divide the application of the method into three stages: formulation of the problem, elicitation of preferences/validation of the decision-maker's preferences, and comparison of alternatives.

Since there is an exponential growth in the alternatives of the problem and the growth of the necessary information in the process of obtaining preferences, a disadvantage of the method is that to control the complexity, the number of criteria and values of the treated criteria is limited. The ZAPROS-IIIi methodology brings an essential difference in relation to the previous models: the division into different stages. The method proposes that the two substations are transformed into one instead of basing the decision-maker's preferences on the first reference situation and then establishing another preference scale using the second reference situation. Therefore, the questions asked considering the first reference situation are the same as those asked considering the second reference situation. Thus, both situations will be considered in answering the question at the same time. The change implies process optimization: dependence on criteria is avoided.

Ultimately, these modifications increased method comparability so that several alternatives defined as incomparable, when applying the ZAPROS method purely, could now

be compared directly or indirectly. Furthermore, these changes in the method process did not change its computational complexity [30].

Considering that it is complicated for a decision-maker, the process of ordering preferences, Tamanini et al. conducted a study with data from a battery of tests for patients with a possible diagnosis of Alzheimer's disease to structure a decision tree based on the characteristics that led to the determination of that disease. In this study, the preference scale was established through the analysis of the resulting tree and, subsequently, the scale was subjected to the ZAPROS method to classify the tests involved in the case studied [31]. This hybrid model shows the potential of integrating VIDA with machine learning solutions [32].

4. A Protocol to Determine the Main Characteristics for Diagnosing Autism Spectrum Disorder

The model was structured using decision trees with the Random Forest to classify the main characteristics of medical evaluations. Then, these characteristics were submitted to the VDA ZAPROS-IIIi method to be placed in order of preference of the decision-maker. The construction of the protocol, structured in machine learning and the ZAPROS-IIIi method of VDA, used the criteria defined based on each qualifier of obstacles experienced by individuals with ASD, namely [33]:

Mild/moderate (from 0% to 49% commitment). Has a slight, regular barrier or difficulty.

Severe (from 50% to 95% commitment). Has high or extreme difficulty.

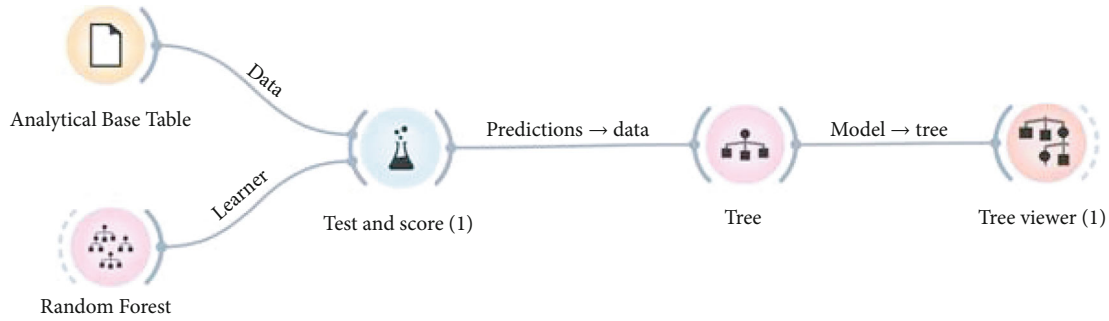


FIGURE 6: Configuration for using Random Forest. Source: formatted by the author.

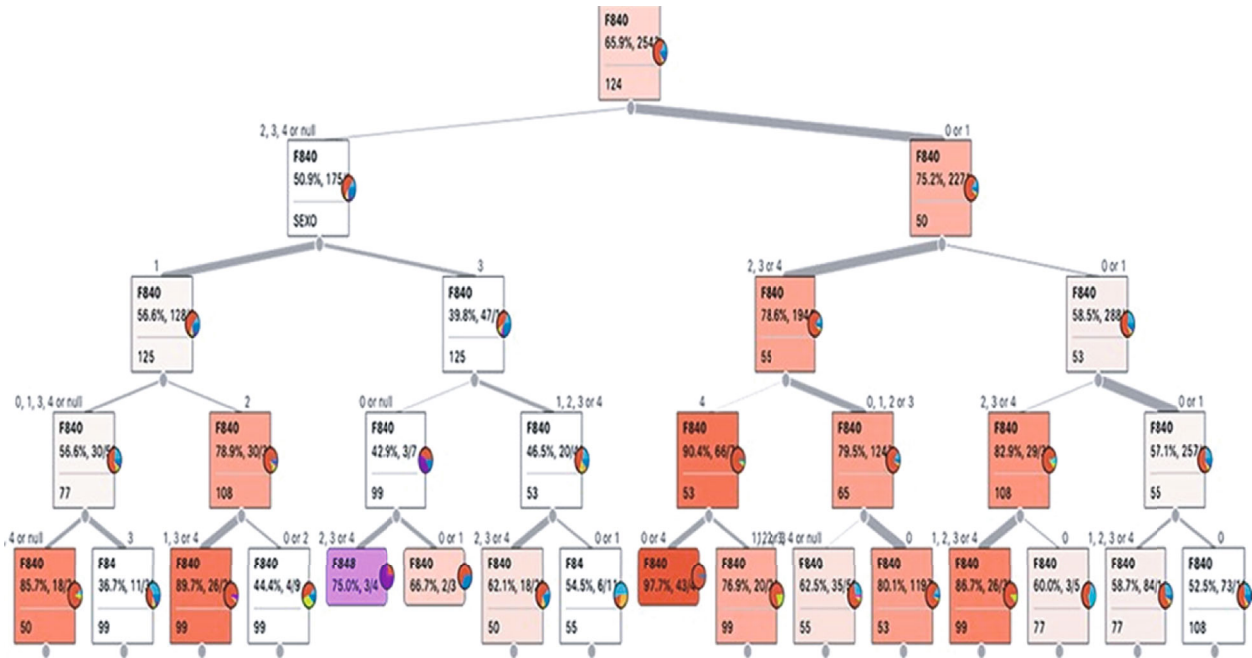


FIGURE 7: The sampling of the main alternatives considered in the evaluation of the resulting tree. Source: formatted by the author.

Complete (96% to 100% commitment). Has a barrier or total difficulty.

The evaluations of 3,861 (three thousand eight hundred sixty-one) children aged 0 to 5 years were selected, indicating the classification “F84” and family of the ICD-10, which characterizes the global developmental disorders—this classification in the ICD-10 corresponds to the code “299.00” in the DSM-5. After data collection, it had the use of keywords for each of the characteristics of the social and medical evaluations, aiming at generating the ABT (Analytical Base Table) and executing the Random Forest, using Orange. Figure 6 shows the configuration for using the Random Forest.

4.1. Identification of the Main Characteristics of the Diagnosis of Autism Spectrum Disorder. The evaluation of the sample of 3,861 (three thousand eight hundred sixty-one) cases, using the decision tree, made it possible to identify the following characteristics [33, 34]:

- (i) Functions of speech fluency and rhythm (changes in fluency, stuttering, verbiage, dyslalia-tachylalia, and bradylalia, among others), in a way compatible with the age group-b330
- (ii) Difficulty to intentionally use the sense of sight (follow objects visually, observe people, watch a sporting event, and observe people or children playing, among others), in a manner compatible with the age group-d110 from 1 year
- (iii) Difficulty in acquiring and executing necessary skills (using cutlery and pencils, among others) and complex skills (games, sports, using tools, and watching, among others), in a way compatible with the age group-d155 from 2 years
- (iv) Sleep functions (start, maintenance, quantity, and quality of sleep), in a way compatible with the age group-b134

TABLE 1: The main characteristics of ASD.

According to your experience and assessment for early diagnosis of Autism Spectrum Disorder (ASD) in children up to 5 years of age, evaluate the following:

Questions	Criteria	Alternatives		
		1. Complete	2. Severe	3. Mild/moderate
1. Regarding the functions of fluency and rhythm of speech (changes in fluency, stuttering, verbiage, dyslalia-tachylalia, and bradylalia, among others), for a definite diagnosis of ASD-b330, you believe that:	A	6	54	31
	B	5	42	44
	C	6	32	53
2. For the difficulty of intentionally using the sense of sight (following an object visually, observing people, and watching a sporting event or children playing, among others), for a definite diagnosis of ASD-d110, you believe that:	A	9	44	38
	B	1	43	47
	C	4	36	51
3. For the difficulty in acquiring and executing necessary skills (using cutlery and pencils, among others) and complicated skills (games, sports, using tools, and watching, among others), in a way compatible with the age group-d155 from 2 years old, do you believe that:	A	5	43	43
	B	4	42	45
	C	10	28	53
4. For sleep functions (start, maintenance, quantity, and quality of sleep), in a way compatible with the age group-b134, do you believe that:	A	7	42	42
	B	10	39	42
	C	6	49	36
5. For memory functions (recent, remote, and amnesic memory disorders), in a way compatible with the age group-b144 from 3 years old, you believe that:	A	12	43	36
	B	7	39	45
	C	7	30	54
6. For global psychosocial functions (interpersonal skills necessary for the establishment of reciprocal social interactions, in terms of meaning and purpose, adaptability, responsiveness, predictability, persistence, and accessibility, and interpersonal communications, among others), in a manner compatible with the age group-b122 and age group-b125 from 2 years, do you believe that:	A	22	41	28
	B	19	46	26
	C	8	36	47
7. For the difficulty of moving using specific equipment or device to facilitate movement (walker, wheelchair, crutches, cane, and others), in a manner compatible with the age group-d465 from 3 years, you believe that:	A	6	50	35
	B	5	54	32
	C	4	39	48
8. For the difficulty in walking (moving on foot, for short or long distances, without the aid of people, equipment, or devices), in a way compatible with the age group-d450 from 2 years old and age group-d465 from 3 years, you think that:	A	10	41	40
	B	6	47	38
	C	23	40	28
9. For vision functions (quality, accuracy, light and color perception, monocular and binocular vision, myopia, hyperopia, astigmatism, hemianopsia, presbyopia, color blindness, tunnel vision, central and peripheral scotoma, diplopia, night blindness, and adaptability to light, among others), in a way compatible with the age group-b210 and age group-d465 from 3 years old, you think that:	A	7	35	49
	B	2	44	45
	C	13	49	29

Source: formatted by the author.

- (v) Memory functions (recent, remote, and amnesic memory disorders), in a way compatible with the age group-b144 from 3 years old
- (vi) Global psychosocial functions (interpersonal skills necessary for the establishment of reciprocal social interactions, in terms of meaning and purpose, adaptability, responsiveness, predictability, persistence, and accessibility, and interpersonal interactions, among others), in a manner compatible with the age group-b122 from 2 years
- (vii) Difficulty moving using specific equipment or device to facilitate movement (walker, wheelchair, crutches, cane, and others), in a manner compatible with the age group-d465 from 3 years old
- (viii) Difficulty walking (moving on foot, for short or long distances, without the aid of people, equipment, or devices), in a manner compatible with the age group-d450 from 2 years
- (ix) Vision functionalities (quality, acuity, light and color perception, monocular and binocular vision, myopia, hyperopia, astigmatism, hemianopsia, presbyopia, color blindness, tunnel vision, central and peripheral scotoma, diplopia, night blindness, and adaptability to light, among others), in a way compatible with the age group-b210

Figure 7 shows the resulting tree of the Orange application, with five levels, and the alternatives considered in the assessment. These main characteristics were applied to the

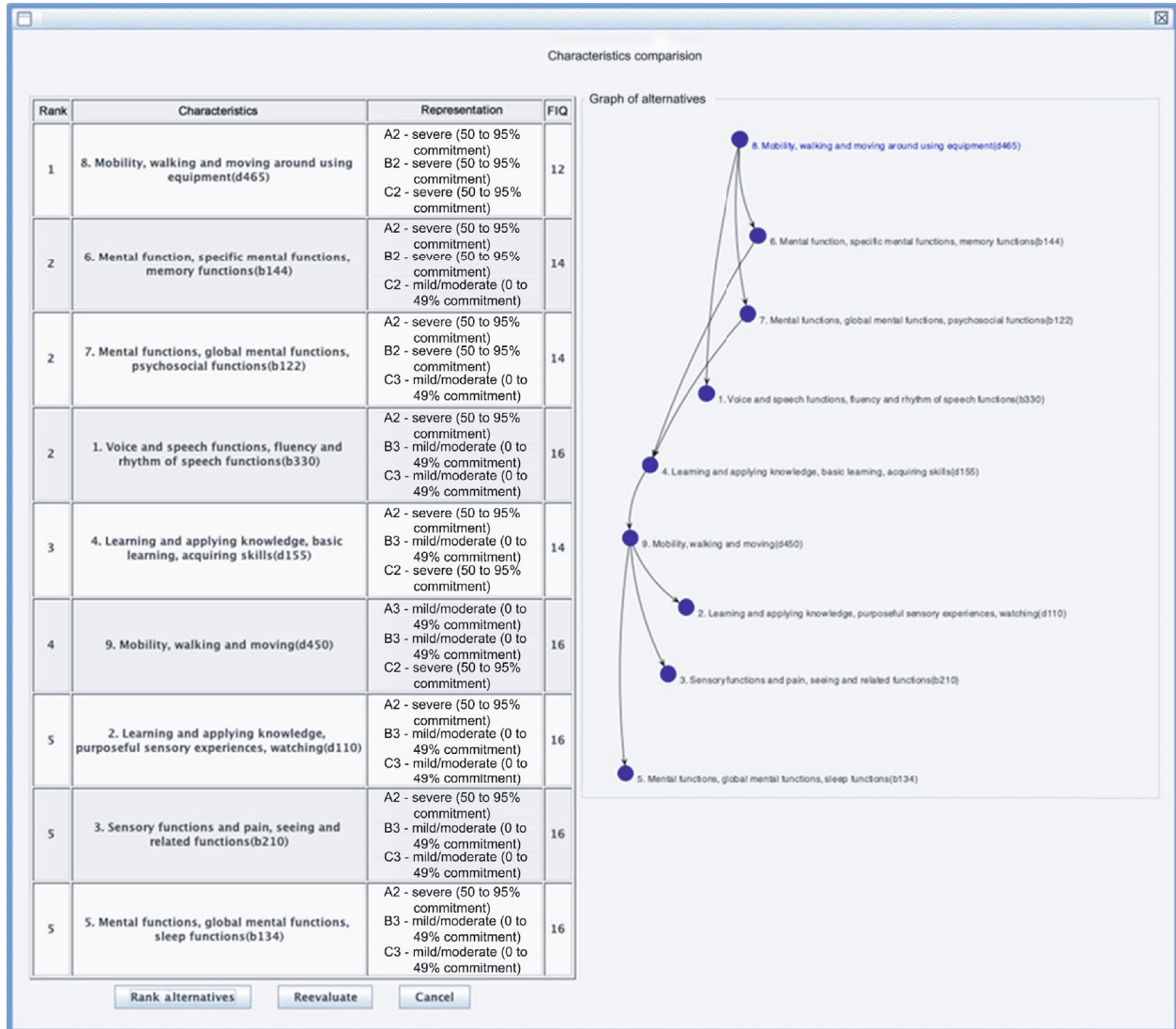


FIGURE 8: Order of preference among the main characteristics of ASD. Source: formatted by the author.

ZAPROS-IIIi method, implemented in the ARANAÚ tool, for an ordering of the most preferable to the least preferred characteristic, in the answers for cases of ASD diagnosis.

Based on the characteristics identified early in Section 4.1, a form was prepared to be filled out by decision-makers in a survey format. Table 1 presents these characteristics in the form of a questionnaire. Each question in this questionnaire had answers under three criteria discriminated in the lines of Table 1, namely:

A. Voice, mental, or vision functions must present a barrier or difficulty.

B. Learning and the application of knowledge must present a barrier or difficulty.

C. Mobility must present a barrier or difficulty.

In the sequence, each of these criteria was evaluated under the following three alternatives discriminated and identified in the columns of Table 1:

Complete (96% to 100% commitment). Has a barrier or total difficulty.

Severe (from 50% to 95% commitment). Has high or extreme difficulty.

Mild/moderate (from 0% to 49% commitment). Has a slight, regular barrier or difficulty.

The mentioned questionnaire served as a basis for conducting research with a sample of 91 (ninety-one) professionals with knowledge and experience in the treatment of people who have ASD:

- (i) 74 (seventy-four) health professionals, twenty doctors, forty physiotherapists, forty nurses, four psychologists, four speech therapists, and two occupational therapists

- (ii) 17 (seventeen) professionals in the educational field

After the elicitation phase, each professional's preferences in the research generated the results shown in Table 1, which were compared with the results obtained through decision trees. Then, the data with values of criteria and alternatives, acquired in the research, were loaded in the

ARANAÚ tool to define the order of preference of the main characteristics and their qualifiers most likely to determine the definitive diagnosis for ASD. The table in Figure 8 shows the order of preference of the ASD's main characteristics, established by the ZAPROS-IIIi method, implemented in the ARANAÚ tool after processing the data obtained in the research. These data enabled the combination of vectors formed by "criteria \times alternatives," that is, "rows \times columns" of Table 1. The graphical representation in Figure 8 makes it possible to observe that the characteristics most preferable to decision-makers have more outgoing arcs than incoming arcs.

5. Conclusion and Future Work

The present study made it possible to analyze the questions of social and medical assessments of the Benefit of Continuous Provision (BPC) applied to children with Autism Spectrum Disorder.

The evaluations have an extensive questionnaire. With the application of decision support methods, it was possible to simplify the choice of specific characteristics for faster and more accurate diagnosis, taking into account the importance of early diagnosis since most of the cases are still detected later. With that, there was a speed gain in identifying the disease, facilitating the physician's decision-making.

This work presented a structured protocol, using decision trees with the Random Forest to classify the main characteristics of the evaluations and the ZAPROS-IIIi method for ordering these characteristics.

The first phase of the study demonstrated that the main symptoms—considered variables or vertices of the tree graph constructed by the Random Forest—are the following, in summary: speech dysfunction, difficulty in intentionally using vision, trouble in acquiring and executing basic and complex skills, sleep dysfunction, memory dysfunction, global social dysfunctions, difficulty moving, difficulty walking, and difficulty in vision. Also, the next step, based on Verbal Decision Analysis, resulted in the following ordering of the criteria selected in the previous step, namely:

- (i) 8. For the difficulty in walking (moving on foot, for short or long distances, without the aid of people, equipment, or devices), in a way compatible with the age group-d450 from 2 years old
- (ii) 6. For global psychosocial functions (interpersonal skills necessary for the establishment of reciprocal social interactions, in terms of meaning and purpose, adaptability, responsiveness, predictability, persistence, and accessibility, and interpersonal communications, among others), in a manner compatible with the age group-b122 from 2 years
- (iii) 7. For the difficulty of moving using specific equipment or device to facilitate movement (walker, wheelchair, crutches, cane, and others), in a manner compatible with the age group-d465 from 3 years

- (iv) 1. Regarding the functions of fluency and rhythm of speech (changes in fluency, stuttering, verbiage, dyslalia-tachylalia, and bradylalia, among others), for a definite diagnosis of ASD-b330
- (v) 4. For sleep functions (start, maintenance, quantity, and quality of sleep), in a way compatible with the age group-b134
- (vi) 9. For vision functions (quality, accuracy, light and color perception, monocular and binocular vision, myopia, hyperopia, astigmatism, hemianopsia, presbyopia, color blindness, tunnel vision, central and peripheral scotoma, diplopia, night blindness, and adaptability to light, among others), in a way compatible with the age group-b210 and age group-d465 from 3 years old
- (vii) 2. For the difficulty of intentionally using the sense of sight (following an object visually, observing people, and watching a sporting event or children playing, among others), for a definite diagnosis of ASD-d110
- (viii) 3. For the difficulty in acquiring and executing necessary skills (using cutlery and pencils, among others) and complicated skills (games, sports, using tools, and watching, among others), in a way compatible with the age group-d155 from 2 years old
- (ix) 5. For memory functions (recent, remote, and amnesic memory disorders), in a way compatible with the age group-b144 from 3 years old

The overall result suggests that of the more than 80 (eighty) variables evaluated, only 9 (nine) would be sufficient to indicate the diagnosis of Autism Spectrum Disorder safely.

The authors suggest, for future work, the application of a broader approach to the protocol, conducting future research in the universe of children and adolescents from 0 to 18 years old, and a more in-depth analysis of the characteristics and their qualifiers.

A comparative study of the present protocol proposal with another model, using algorithms based on Bayesian networks, logistic regression, or other machine learning techniques, will expand and improve the model now proposed for the decision-making process. At this point in the study, comparisons can be made between the mentioned algorithms [34].

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

Acknowledgments

The authors would like to thank INSS and DATAPREV for making available the data used in this study. The authors would like to thank the Edson Queiroz Foundation/University of Fortaleza for all the support. The third author thanks Brazilian National Council for Research and Development (CNPq) via Grant No. 304272/2020-5. The fourth and fifth authors are thankful for the support given by the 7 de Setembro University Center during this project.





References

- [1] WHO Homepage, "Autism spectrum disorders," 2020, <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>.
- [2] C. S. Paula, S. H. Ribeiro, E. Fombonne, and M. T. Mercadante, "Brief report: prevalence of pervasive developmental disorder in Brazil: a pilot study," *Journal of Autism and Developmental Disorders*, vol. 41, no. 12, pp. 1738–1742, 2011.
- [3] L. C. Nunes, P. R. Pinheiro, M. C. D. Pinheiro et al., "A hybrid model to guide the consultation of children with autism spectrum disorder," *Springer Proceedings in Complexity*, A. Visvizi and M. D. Lytras, Eds., , pp. 419–431, Springer International Publishing, 2019.
- [4] APA–American Psychiatric Association, "Diagnostic and Statistical Manual of Mental Disorders (DSM –5)," 2020, <https://www.psychiatry.org/psychiatrists/practice/dsm>.
- [5] M. C. C. Martone and L. H. Z. Santos-Carvalho, "A review of articles published in the Journal of Applied Behavior Analysis (JABA) on verbal behavior and autism between 2008 and 2012," 2012.
- [6] I. L. Cohen, V. Sudhalter, D. Landon-Jimenez, and M. Keogh, "A neural network approach to the classification of autism," 1993.
- [7] M. S. Yekta, M. Mahmoudi, and S. Akbari-Zardkhaneh, "Developing Autism Screening Expert System (ASES)," *AWER Procedia Information Technology & Computer Science*, vol. 4, pp. 1063–1068, 2013, 3rd World Conference on Innovation and Computer Sciences, 2013.
- [8] V. Khullar, H. P. Singh, and M. Bala, "Intelligent Handheld Expert System (HES) for diagnosis of autism spectrum disorder and its severity level," *ASEAN Journal of Psychiatry*, vol. 19, no. 1, 2018.
- [9] I. L. Cohen, "An artificial neural network analogue of learning in autism," *Biological Psychiatry*, vol. 36, no. 1, pp. 5–20, 1994.
- [10] T. Florio, S. Einfeld, B. Tonge, and A. Brereton, "Providing an independent second opinion for the diagnosis of autism using artificial intelligence over the internet," *Psychotherapy, and Health*, vol. 5, (1), The Use of Technology in Mental Health Special Issue, pp. 232–248, 2009.
- [11] F. Thabtah, "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment," in *ICMHI '17: International Conference on Medical and Health Informatics*, Taichung City Taiwan, 2017.
- [12] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: a review and ways forward," *Informatics for Health and Social Care*, vol. 44, no. 3, pp. 278–297, 2019.
- [13] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection," *Health Informatics Journal*, vol. 26, no. 1, pp. 264–286, 2020.
- [14] F. Thabtah, N. Abdelhamid, and D. Peebles, "A machine learning autism classification based on logistic regression analysis," *Health Information Science and Systems*, vol. 7, no. 1, p. 12, 2019.
- [15] O. P. d. S. Neto, V. H. V. d. Sousa, G. B. Batista, F. C. B. G. Santana, and J. M. B. O. G.-A. S. D. Junior, "A tool to aid the learning of children with autism spectrum disorder, based on the ABA methodology," 2013.
- [16] C. Lampreia, "Avaliações quantitativa e qualitativa de um menino autista: uma análise crítica," *Psicologia em Estudo*, vol. 8, no. 1, pp. 57–65, 2003.
- [17] B. S. Costa, E. B. Nakandakare, and E. Paulino, "The insertion of autistics in the academic and professional information technology," 2014, <http://www.revistarefas.com.br/index.php/RevFATECZS/article/view/190>.
- [18] M. J. Maenner, K. A. Shaw, J. Baio et al., "Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring network, 11 sites, United States, 2016," *MMWR. Surveillance Summaries*, vol. 69, no. 4, pp. 1–12, 2020.
- [19] I. Tamanini, "A structured tool in verbal decision analysis applying ZAPROS," ANAIS|PROCEEDINGS-Brazilian Computer Society–SBC, Maringá-Paraná, 2016, <http://cbsoft.org/cbsoft2016/anais-dos-eventos/cbsoft2016wdes.pdf>.
- [20] L. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [21] E. Scornet, G. Biau, and J. P. Vert, "Consistency of random forests," *The Annals of Statistics*, vol. 43, no. 4, pp. 1716–1741, 2015.
- [22] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, pp. 197–227, 2016.
- [23] Y. Jeon and Y. Lin, "An effective method for high-dimensional log-density ANOVA estimation, with application to nonparametric graphical model building," *Statistica Sinica*, vol. 16, pp. 353–374, 2006.
- [24] L. L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [25] O. Larichev and H. M. Moshkovich, *Verbal decision analysis for unstructured problems*, Kluwer Academic Publishers, Boston, MA, USA, 1997.
- [26] L. L. Breiman, J. H. J. H. Friedman, R. A. R. A. Olshen, and C. J. C. J. Stone, *Classification and Regression Trees*, Wadsworth Advanced Books and Software, Belmont, CA, 1984.
- [27] T. C. S. Machado, P. R. Pinheiro, and I. Tamanini, "Project management aided by verbal decision analysis approaches: a case study for the selection of the best SCRUM practices," *International Transactions in Operational Research*, vol. 22, no. 2, pp. 287–312, 2014.
- [28] J. Demsar, T. Curk, A. Erjavec et al., "Orange: data mining toolbox in Python," *Journal of Machine Learning Research*, vol. 14, pp. 2349–2353, 2013.
- [29] O. I. Larichev and H. M. Moshkovich, *Verbal Decision Analysis for Unstructured Problems*, Kluwer Academic Press, Boston, 1997.
- [30] P. R. Pinheiro, I. Tamanini, M. C. Dantas Pinheiro, and V. H. C. de Albuquerque, "Evaluation of the Alzheimer's disease clinical stages under the optics of hybrid approaches in verbal decision analysis," *Telematics and Informatics*, vol. 35, no. 4, pp. 776–789, 2018.
- [31] I. Tamanini, P. R. Pinheiro, D. Santos, and C. Nogueira, "A hybrid approach of verbal decision analysis and machine

- learning,” in *Lecture Notes in Computer Science*, J. T. Yao, Y. Yang, R. Słowiński, S. Greco, H. Li, S. Mitra, and L. Polkowski, Eds., vol. 7413pp. 126–131, Springer, Berlin Heidelberg, 1 edition, 2012.
- [32] M. Simão Filho, P. R. Pinheiro, A. B. Albuquerque, R. P. S. Simão, R. S. N. Azevedo, and L. C. Nunes, “A multicriteria approach to support task allocation in projects of distributed software development,” *Complexity*, vol. 2019, Article ID 3926798, 22 pages, 2019.
- [33] WHO Homepage, “International Classification of Functioning, Disability and Health,” 2020, https://psychiatr.ru/download/1313?view=name=CF_18.pdf.
- [34] E. C. d. Andrade, *Um Modelo Híbrido em Aprendizagem de Máquina e Análise Verbal de Decisão Aplicado ao Diagnóstico de Autismo*, Masters Dissertation, University of Fortaleza, 2020.

Research Article

SARPPIC: Exploiting COVID-19 Contact Tracing Recommendation through Social Awareness

Nana Yaw Asabere ¹, Amevi Acakpovi ², Emmanuel Kwaku Ofori ³, Wisdom Torgby,¹
Marcellinus Kuuboore ⁴, Gare Lawson,¹ and Edward Adjaloko¹

¹Department of Computer Science, Accra Technical University, Ghana

²Department of Electrical/Electronic Engineering, Accra Technical University, Ghana

³Department of Chemical Pathology, University of Ghana, Ghana

⁴Department of Information Technology Studies, University of Professional Studies, Ghana

Correspondence should be addressed to Nana Yaw Asabere; yawasabere2005@yahoo.com

Received 3 June 2020; Revised 3 October 2020; Accepted 16 October 2020; Published 16 November 2020

Academic Editor: Plácido R. Pinheiro

Copyright © 2020 Nana Yaw Asabere 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.

Globally, the current coronavirus disease 2019 (COVID-19) pandemic is resulting in high fatality rates. Consequently, the prevention of further transmission is very vital. Until vaccines are widely available, the only available infection prevention methods include the following: contact tracing, case isolation and quarantine, social (physical) distancing, and hygiene measures (washing of hands with soap and water and using alcohol-based hand sanitizers). Contact tracing, which is key in preventing the spread of COVID-19, refers to the process of finding unreported people who maybe infected by using a verified case to trace back possible infections of contacts. Consequently, the wide and fast spread of COVID-19 requires computational approaches which utilize innovative algorithms that build a memory of proximity contacts of cases that are positive. In this paper, a recommender algorithm called socially aware recommendation of people probably infected with COVID-19 (SARPPIC) is proposed. SARPPIC initially utilizes betweenness centrality in a social network to measure the number of target contact points (nodes/users) who have come into contact with an infected contact point (COVID-19 patient). Then, using contact durations and contact frequencies, tie strengths of the same contact points above are also computed. Finally, the above algorithmic computations are hybridized through profile integration to generate results for effective contact tracing recommendations of possible COVID-19-infected patients who will require testing in a healthcare facility. Benchmarking experimental results in the paper demonstrate that, using two interconnected relevant real-world datasets, SARPPIC outperforms other relevant methods in terms of suitable evaluation metrics such as precision, recall, and F-measure.

1. Introduction

Coronavirus disease (COVID-19), which originated in December 2019 from the city of Wuhan in China, is caused by severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) [1]. Globally, COVID-19 has clearly shown its potential of high fatality rates. A global response pertaining to effective health systems and delivery is extremely imperative and vital. Over one hundred (100) countries worldwide have been hit severely by COVID-19 and are currently awaiting reliable and sustainable vaccines [1–3].

Consequently, until innovative vaccines are made widely available, the only existing infection prevention methods are contact tracing, case isolation and quarantine, social (physical) distancing, and hygiene measures such as washing hands with soap and water and using alcohol-based hand sanitizers [1–3]. Globally, as different states and local governments seek a way out of lockdowns that have brought their economies to a near standstill, “contact tracing” has made its way into everyday conversations as well. COVID-19 contact tracing is similar to detective work. Trained staff conduct interviews with people who have been diagnosed with COVID-19 to

verify who they may have recently been in contact with. Then, once those who have been in contact with the COVID-19 patient is known, the trained staff inform them that they may have been exposed and encourage them to quarantine themselves to prevent spreading of the disease any further.

This process is very laborious. Interviewing COVID-19 patients and reaching out to dozens of contacts takes time. For this reason, contact tracing works best when there are low levels of infection in a community. However, many high-level communities worldwide are currently affected by COVID-19. Consequently, contact tracing methods need to be improved through technology (computerization and automation). The use of a contact tracing application (app) which is designed using computational algorithms for building a memory of proximity contacts would be adequate to stop the epidemic to some extent [1, 2]. Such an application can immediately notify and recommend contacts of positive cases.

Recommender systems have become very popular due to the fact that they can help users to find items of interest (such as movies, books, and music) in order to cope with the information overload (big data) problem [4]. In the past decade, many researchers have worked to develop recommender systems which involve people to people recommendations; some of these include [5–8]. With reference to people-to-people recommendations, the current issue of contact tracing in COVID-19 can also be tackled using recommender systems.

This paper proposes a recommender algorithm called socially aware recommendation of people probably infected with COVID-19 (*SARPPIC*). *SARPPIC* utilizes the computations of between centrality and social ties [9] as contact tracing entities in a social network to generate recommendations regarding possibly infected COVID-19 patients. The major contributions of this paper are summarized as follows:

(i) Betweenness centrality and accurate tie strength computations are applied in a social graph network for predicting people who may be infected with COVID-19.

(ii) A recommendation method which exploits and hybridizes information regarding high betweenness and tie strength to generate people-to-people recommendations is proposed.

(iii) Using appropriate evaluation metrics, benchmarking experiments were conducted on two interconnected relevant real-world datasets, to verify the effectiveness of the proposed recommendation method.

The rest of this paper is structured as follows. Section 2 discusses related studies pertaining to this paper. Section 3 outlines the details of the proposed *SARPPIC* method. Section 4 elaborates on the performance evaluation and further discusses the results achieved. Finally, Section 5 concludes the paper.

2. Related Studies

This section presents related studies and literature pertaining to the study. Related studies in the paper focus on the following: (i) people-to-people recommender systems, (ii) social

recommendations through tie strength and betweenness centrality, and (iii) computational/automation methods in COVID-19 contact tracing.

2.1. People-to-People Recommender Systems. People-to-people prediction and recommendation has recently become an imperative task in many online social networks. In recommender systems, traditional collaborative filtering (CF) approaches are popular for effectively predicting user preferences for items. However, in online social networks, people have a dual role as both “users” and “items,” e.g., both initiating and receiving contacts [8]. The main objective of people-to-people recommender systems is to generate meaningful social suggestions to users [8, 10]. Some related studies regarding people-to-people recommendation include the following: Cai et al. [8] proposed a recommendation model called *SocialCollab* which fully captures the bilateral role of user interactions within a social network and formulates CF methods to enable people-to-people recommendation. In their recommendation model, users can be similar to other users in two ways—either having similar “attractiveness” for the users who contact them or having similar “taste” for the users they contact. Similarly, Tsai and Brusilovsky [10] proposed an innovative method which integrates a global search result using a personalized people-to-people recommender system. Their method employs the user identity as a query keyword and processes the search results through five different customized parsers. In relation to the problem involving CF over-recommending popular items, Krzywick et al. [11] addressed this problem in the context of people-to-people recommendations. Similarly, Bourke et al. [12] examined the practice of leveraging a user’s social graph in order to generate people recommendations. Using various neighbourhood selection strategies, they investigated user satisfaction and the level of perceived trust in the recommendations received. On a large commercial online dating site, Wobcke et al. [7] reported on the successful deployment of a people-to-people recommender system. The deployment was the result of thorough evaluation and an online trial of a number of methods. Xia et al. [5] and Asabere et al. [6] similarly proposed linear hybrid recommender algorithms which employ accurate prediction of tie strengths and personality to generate conference participant (people) recommendations at a smart conference.

2.2. Social Recommendations through Betweenness Centrality and Tie Strength. In a social network, betweenness centrality is a measure of the influence of a vertex over the flow of information between every pair of vertices under the postulation that information primarily flows over the shortest paths between them [13]. Various researchers have applied betweenness centrality in their recommendation approaches. Noteworthy research work regarding such methods includes the following: Samad et al. [14] employed textual and topological similarity measures for citation recommendation of relevant/important research papers and then computed betweenness centrality measures to generate recommendations of important papers for researchers. Similarly, by exploiting betweenness centrality, Sie et al. [15] proposed a

novel tool which employs similar mindedness to recommend potential co-authors. Similar to [14, 15], Christensen and Schiaffino [16] applied a betweenness centrality approach to propose a social recommendation method which involved group modeling.

Mahyar et al. [17] utilized a betweenness centrality concept to propose a novel method which improves recommendation accuracy in accordance to the most central users who are designated as group heads. Pucci et al. [18] utilized betweenness centrality to depict a random-walk-based scoring recommender algorithm, which recommends top-rank items to possibly interested users. Sulieman et al. [19] expounded on a recommendation method between items using a mixed approach of social network analysis and content as well as CF. From a root item, a relationship's graph is generated and used to extract network metrics through betweenness and closeness centrality measures. Souza et al. [20] proposed an algorithmic method which recommends items based on social network centrality and semantic relevance.

Tie strength or social ties usually refer to the social interactions between individuals. A meaningful social relationship such as friendship between two individuals represents the existence or not of ties [9, 21, 22]. Social ties can be categorized into strong ties (e.g., family members or trusted friends) that share relevant information with a huge overlap. In contrast, weak ties (e.g., acquaintances) share more diverse and new information [21]. Quite recently, the application of social ties in recommender systems has been exploited by various researchers. Some of such related studies are presented below.

Xia et al. [23] and Asabere et al. [24] similarly addressed the recommendation of presentation sessions at smart conferences to attendees using their social ties. They proposed a venue recommender algorithm called socially aware recommendation of venues and environments (SARVE). In relation to the research paper recommendation, Asabere et al. [25] and Xia et al. [26] improved the social awareness of attendees in a smart conference by proposing an innovative folksonomy-based paper recommender algorithm, namely, socially aware recommendation of scholarly papers (SARSP).

Using Facebook data for the use case of online news with 193 participants, Oechslein and Hess [21] developed a research model and tested it in an online experiment. Their proposed structural equation model results showed that strong tie relationships have positive impacts on recommendation value. Similarly, using a strong concept of social ties, Jang et al. [27] proposed a system which detects and analyzes the behaviors of group-level socializing to support ex post facto and real-time social applications deployed in real social event situations.

2.3. Computational/Automation Methods in COVID-19 Contact Tracing. In relation to the current and global COVID-19 pandemic and the high proportion of transmissions from presymptomatic individuals, controlling the epidemic through manual contact tracing is infeasible. Quite recently, a number of contact tracing technological solutions have been introduced. For example, from the very early stage

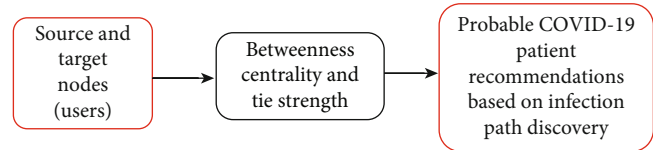


FIGURE 1: Fundamental recommendation procedure of SARPPIC.

of the epidemic, China and South Korea commenced tracing COVID-19 victims and their contacts through facial recognition technologies as well as smartphones [1]. Consequently, the extensive and rapid spread of COVID-19 requires computational approaches which utilize innovative algorithmic and mathematical methods. Very recently, some authors have developed algorithms and models in this regard. Significant among these are as follows:

Ferretti et al. [1] developed an algorithm which involves a mathematical model that encompasses the memory of proximity contacts and immediately notifies contacts of positive cases through a contact tracing app. Abler et al. [28] discussed the implementation of a contact tracing app based on Bluetooth low energy technology between two people to control COVID-19. Similar to [1, 28], Yasaka et al. [29] proposed an anonymized graph of interpersonal interactions to conduct a novel form of contact tracing and further developed a proof-of-concept smartphone app which implements their approach. Additionally, they developed a computer simulation model that validates their proposed method. Drew et al. [30] developed algorithms and modeled a COVID-19 Symptom Tracker mobile application. Their mobile application offers clinical outcomes, herald symptoms, geographical hot spots, and data on risk factors.

Current trends of the computational COVID-19 contact tracing research enumerated above show that there is lack of social properties inclusion regarding people infected with COVID-19 and their respective contacts in a social network. Due to the fact that COVID-19 is a social issue, social properties are very important factors worth considering [9, 22]. The notion of social properties attracts substantial interest initially from the social and behavioral communities, as well data mining communities and network communities [9, 22]. Different from the research work enumerated, this paper proposes a graph-based social recommendation method which utilizes betweenness centrality and social ties as social properties of probable COVID-19 contacts to develop a recommender algorithm (SARPPIC). Benchmarking experimental results below have verified the effectiveness of the proposed method.

3. Proposed Solution—SARPPIC Method

This section presents the framework and proposed solution of the SARPPIC recommendation method. Figure 1 depicts the fundamental recommendation procedure of SARPPIC. In relation to infection path discovery, Figure 2 shows that, through the *betweenness centrality verifier*, SARPPIC initially verifies and computes betweenness centralities of contact points in the social graph network. Additionally, SARPPIC utilizes the *tie strength verifier* to compute tie strength

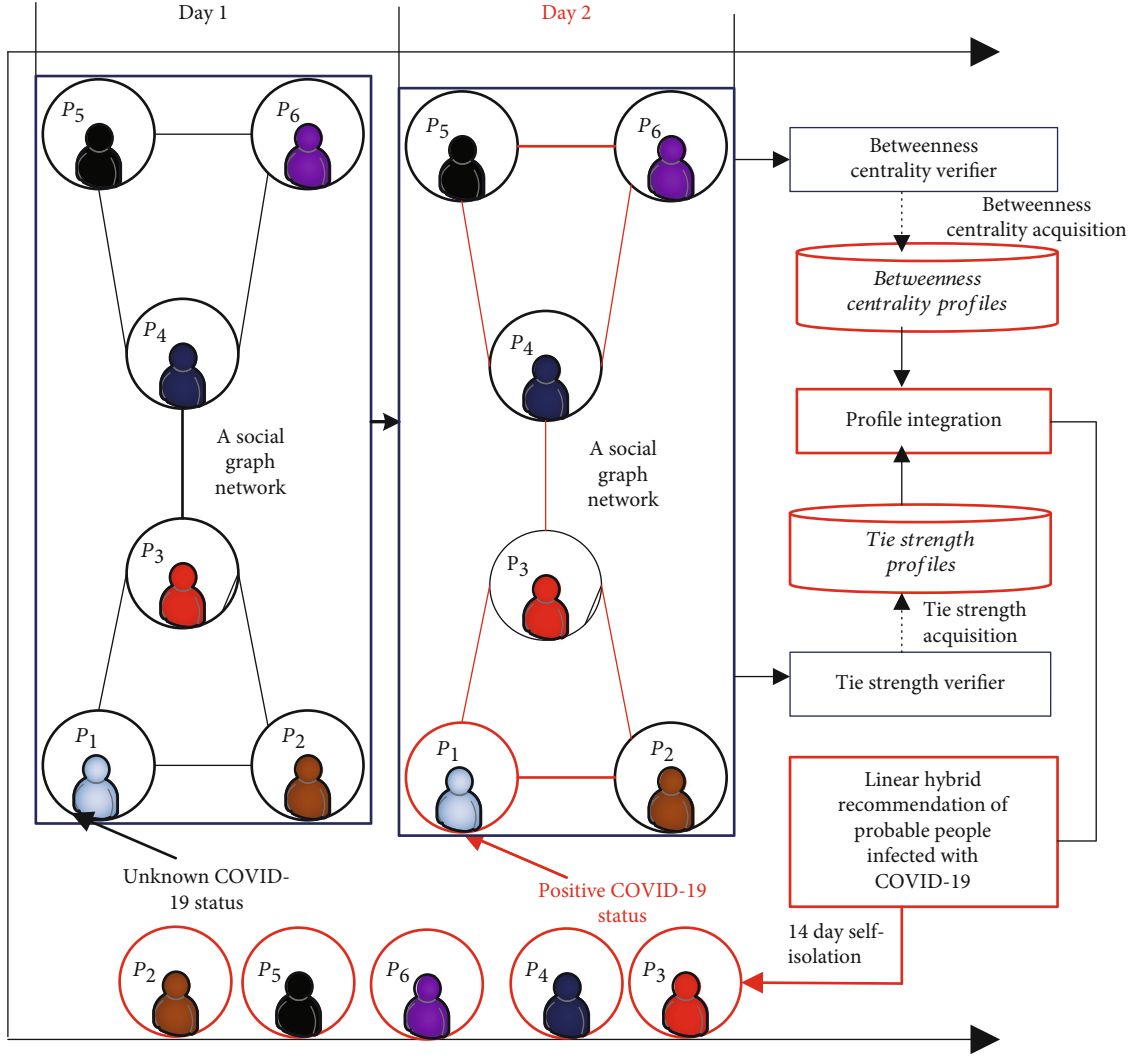


FIGURE 2: SARPPIC recommendation model.

profiles of the contact points through their contact durations and contact frequencies for onward linear hybrid recommendation of probable people infected with COVID-19. Further elaboration on the proposed SARPPIC recommendation model is presented below.

3.1. COVID-19 Infection Path Discovery. Central to the proposed SARPPIC recommendation method is a data structure which is referred to as a social graph. Generally, social networks illustrate the small world ideology that node encounters are adequate to build a connected relationship graph. A social graph is an appropriate tool which represents the relational structure of social networks in a natural manner. In a social graph, vertices (nodes) indicate human individuals, and edges (links) indicate social relationships between individuals [9, 22].

In the proposed method, the social graph consists of directed edges (links), which represent transmission vectors between contact points (nodes). The concept of betweenness centrality is utilized in a defined social graph shown in Figure 2. Betweenness centrality specifies the betweenness

TABLE 1: BC computations of contact points in Figure 2.

Source node	Target node	σ_{ab}	$\sigma_{ab}(v)$	$\frac{\sigma_{ab}(v)}{\sigma_{ab}}$
P_1	P_2	1	0	0
P_1	P_4	1	1	1
P_1	P_5	1	1	1
P_1	P_6	1	1	1
P_2	P_4	1	1	1
P_2	P_5	1	1	1
P_2	P_6	1	1	1
P_4	P_5	1	0	0
P_4	P_6	1	0	0
P_5	P_6	1	0	0

of a vertex in a network, and it indicates the extent to which a vertex lies on the shortest paths between pairs of other vertices. In many real-world situations such as the current COVID-19 pandemic, it has quite a significant role. As

```

Pseudocode for COVID-19 infection path discovery-highest BC in social (graph) network.
Input:  $a, b, c, d, e$ 
Output: Highest BC
1: procedure highest BC computation
2: // Declare and initialize variables
3:  $a, b, c, d,$  and  $e$  // integer variables
4:  $BC[e]$  // floating variables
5: Source Nodes [a]; // array of Source Nodes of size c
6: Target Nodes [b]; // array of Target Nodes of size d
7: for (a=0 to a<c increment a) do
8: for (b=0 to b<d increment b) do
9: Compute BC for source and target nodes using
Equation (1) and store results in  $BC[e]$ 
10: Add results in  $BC[e]$  to generate highest BC
11: end for
12: end for
13: end procedure

```

ALGORITHM 1.

shown in Equation (1), the Betweenness Centrality (BC) of a vertex v , i.e., $BC(v)$ for any graph is defined as follows:

$$BC(v) = \sum_{a,b \in V} \left(\frac{\sigma_{ab}(v)}{\sigma_{ab}} \right), \quad (1)$$

where $\sigma_{ab}(v)$ is the total number of shortest paths between nodes a and b that pass through v , and σ_{ab} is the total number of shortest paths between nodes a and b . The proposed SARPPIC recommendation method computes the BC of node P_3 as tabulated in Table 1.

In the social graph in Figure 2, each contact point denotes a physical interaction between two or more individuals at a specific time and place, during which microorganisms could potentially be transmitted from one individual to others. Each contact point in the social graph can be classified in one of the two states: *positive status* or *unknown status*. A positive status represents a contact point which has been identified as having COVID-19 infection, while an unknown status denotes contact points who are not yet infected by COVID-19.

With reference to Figure 2, using the simple data structure of the social graph, possible transmission paths can be determined for any given target contact point. A possible transmission path is defined as a path from a positive status node which could be carrying microorganisms from a reported point of exposure to a given target node with an unknown status. A demonstration of a simple transmission graph is provided in Figure 2.

According to Figure 2, there are six contact points and they are connected. Figure 2 shows that on day 1, contact point P_1 has an unknown status. However, on day 2, P_1 has a positive status. There is no shortest path of P_3 involving the paths of the following: P_1 to P_2 , P_4 to P_5 , P_4 to P_6 , and P_5 to P_6 ; hence, $\sigma_{ab}(v) = 0$ in all these cases. However, movement between any contact points (e.g., P_1 to P_4) in the net-

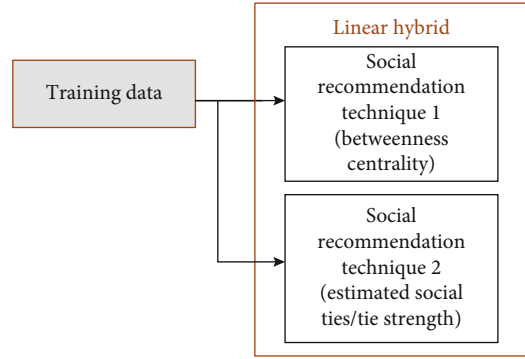


FIGURE 3: Training phase process in SARPPIC.

work as depicted in Table 1 utilizes the shortest path of P_3 ; therefore, $\sigma_{ab}(v) = 1$ in all these cases.

Therefore, as shown in Table 1, BC for node P_3 in Figure 2 is the summation of all values which resulted in 1. BC for node P_3 is therefore equal to 6. Furthermore, due to the fact that nodes P_3 and P_4 are on the same level in the network, the BC for node P_4 will also be 6. Algorithm 1 shows the computation of highest BC. A high betweenness count for an infected contact point indicates that the contact point holds authority over other contact points in the social graph network. Consequently, as illustrated in Figure 2, due to the fact that on day 2, P_1 has a positive status; if P_1 transmits to P_3 , the high centrality of P_3 paves the way for all other contact points (P_4 , P_5 , and P_6) to also be infected with COVID-19.

3.2. Social Ties/Tie Strength of Contact Points. Existing research literature above has shown evidence that in a social network, the relationship, connection, and influence of users enhance reliability, effectiveness, and productivity. Furthermore, recommender systems research has provided evidence that the application of social elements and attributes improves recommendation quality and accuracy by avoiding cold start as well as data sparsity problems [4–6]. In the

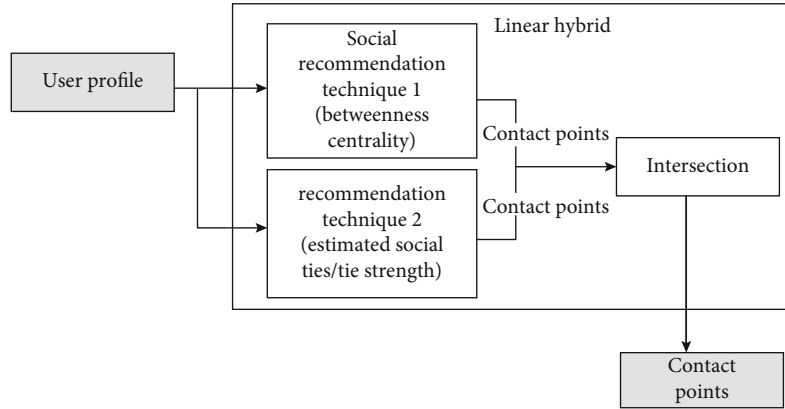


FIGURE 4: Contact point profile modeling in SARPPIC.

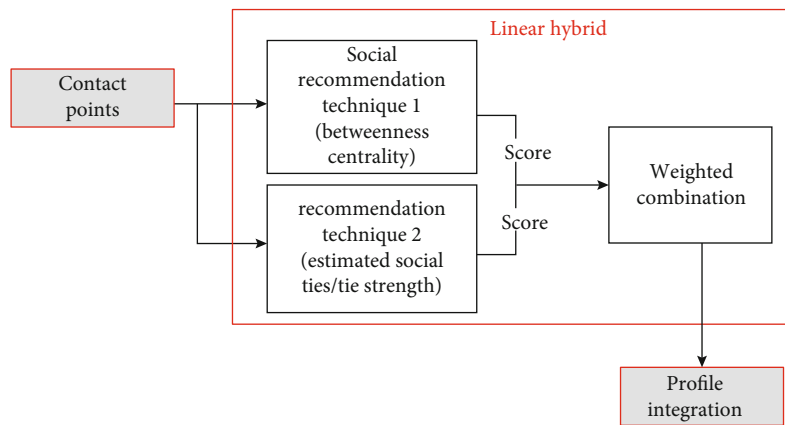


FIGURE 5: Profile integration process in SARPPIC.

SARPPIC pseudocode for social (linear) recommendation of people probably infected with COVID-19.

Input: a, b, c, d, e, f

Output: Hybrid COVID-19 Recommendation

1: **procedure probable people infected with COVID-19 recommendation**

2: // Declare and initialize variables

3: $a, b, c, d, e,$ and f // integer variables

4: $threshold_value, BC[e], Tie_Strength[d]$ and $hybrid_rec[f]$; // floating variables

5: Contact Points $[f]$; // array of contact points of size f

6: **for** ($a=0$ to $a<c$ increment a) **do**

7: **for** ($b=0$ to $b<d$ increment b) **do**

8: Use Algorithm 1 to compute highest betweenness contact points and store results in $BC[e]$

9: Use Equation (2) to compute $Tie_Strength$ highest betweenness contact points and target nodes and store results in $Tie_Strength[d]$

10: Use Equation (3) to merge results of $BC[a][b]$ with $Tie_Strength[a][b]$ and store in $hybrid_rec[f]$

11: **end for**

12: **end for**

13: // Linear Hybrid Recommendation

14: **for** ($i=0$ to $i<e$ increment i)

15: **if** ($hybrid_rec[f] \geq threshold_value$) **then**

Generate Hybrid COVID-19 Social Recommendation

16: **end if**

17: **end for**

18: **end procedure**

ALGORITHM 2.

TABLE 2: Dataset: centrality indices for the HEXACO-60 network.

Nodes	Betweenness
H06, H12, H16	5
H24, H30, H36	31
H42, H48	14
H64, H60	0
E05, E11	6
E17, E23	2
E29, E35, E41	3
E47, E53, E59	17
X58, X52, X46	11
X40, X34, X28	23
X04, X10	7
X16, X22	12
A9, A15	5
A21, A57, A33	5
A27, A51, A03	14
A39, A45	5
C56, C50, C44	7
C02, C08, C14	26
C20	5
C38, C32, C26	19
O01, O55, O25	14
O31, O37, O43	5
O13, O49	10
O19, O07	3

TABLE 3: ATU dataset—contact frequency trends.

Past tie strength data		Present tie strength data	
Contact frequency	Number of contact points	Contact frequency	Number of contact points
1	263	1	438
2	1246	2	1127
3	669	3	805
4	477	4	485
5	291	5	245
6	243	6	134
7	108	7	63

proposed *SARPPIC* recommendation method, another common social property called social ties/tie strength is utilized. Computations of social ties/tie strength are done through contact duration and contact frequency of contact points [5, 6, 9, 21–25].

Generally, the social ties/tie strength of two users in a social network are computed to verify the extent of their relationship and the influence they have on each other. Equation (2) is used to compute the tie strengths of contact points in the social network graph. As shown in Figure 2, these computations are utilized as a strategy to establish the strength of relationship for profile integration with prior betweenness centrality computations, in order to generate effective linear

TABLE 4: ATU dataset—contact duration trends.

Past social tie data		Present social tie data	
Contact duration	Number of contact points	Contact duration	Number of contact points
5	126	5	129
10	370	10	438
15	165	15	154
20	405	20	420
25	299	25	244
30	471	30	432
35	229	35	223
40	301	40	298
45	124	45	119
50	254	50	239
55	61	55	62
60	181	60	182
65	145	65	161
70	104	70	104
75	1	75	1
80	61	80	61

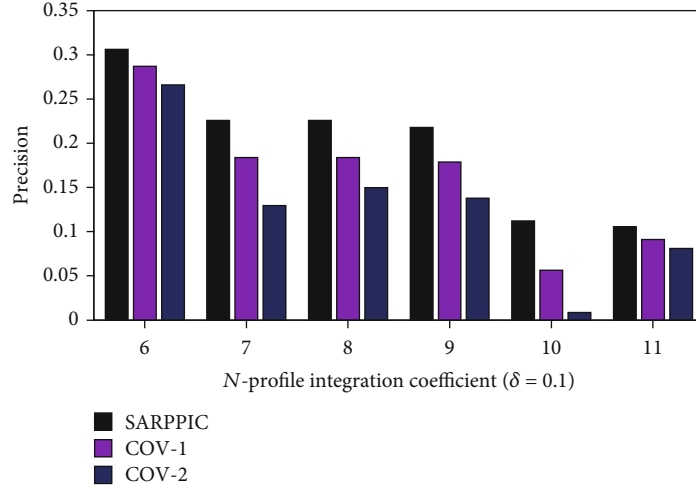
hybrid recommendations of people probably infected with COVID-19.

$$\text{Tie_strength}_{P_3, P_2}(t + \Delta t) = \text{tie_strength}_{P_3, P_2}(t - \Delta t) + (1 - \delta) \times \text{tie_strength}_{P_3, P_2}(t). \quad (2)$$

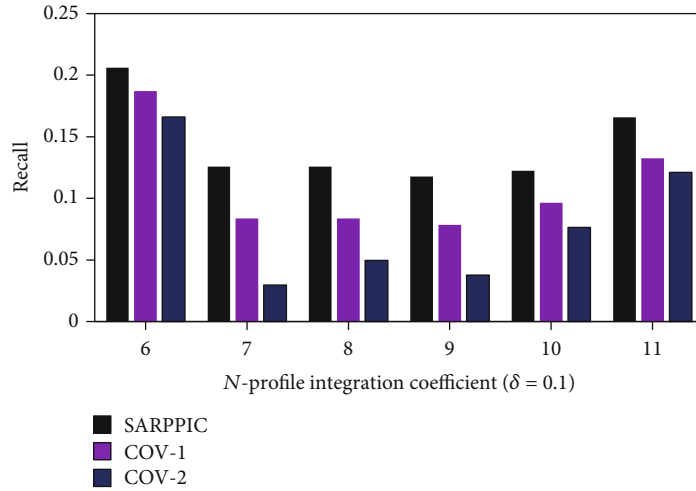
In Equation (2) show above, $\text{tie_strength}_{P_1, P_2}(t - \Delta t)$ and $\text{tie_strength}_{P_1, P_2}(t)$ are the past and present social ties/tie strengths between P_3 and P_2 , where P_3 is the contact point with the highest BC, and P_2 is a target node (contact point). δ is a parameter that resolves the influence proportion of the present and past social ties, and Δt is the time frame used to compute the social ties/tie strength between P_3 and P_2 .

3.3. SARPPIC Algorithm and Linear Hybrid Recommendation. Figure 3 demonstrates the training phase of *SARPPIC*. In Figure 3, each individual social recommendation technique processes the training data. After the training phase, Figure 4 shows how the modeling of user profiles relating to contact points for test users are generated. Therefore, these recommendation techniques jointly propose contact points who have common intersections of user profiles, in terms of the contact point with the highest BC and corresponding tie strengths. Contact point generation is essential in verifying people who will be considered in the linear hybrid recommendation.

As illustrated in Figure 5, the contact points are then sorted out through their combined weighted score, and high-valued profile integrations validate a top linear hybrid recommendation. As explained above, in the experimentation procedure, computations of betweenness centrality and social ties/tie strength of contact points are incorporated



(a)



(b)

FIGURE 6: (a) Precision performance results on datasets ($\delta = 0.1$); (b) recall performance results on datasets ($\delta = 0.1$).

using Equations (1) and (2), respectively. The incorporation of the results for Equations (1) and (2) is linearly hybridized using Equation (3) below. The proposed *SARPPIC* recommendation method therefore improves recommendation accuracy and also enhances the socially aware recommendation for probable people with COVID-19 in a social network graph.

$$PI(P_3, P_2) = \text{tie_strength}_{P_3, P_2}(t + \Delta t) + BC(v). \quad (3)$$

Through profile integration, Equation (3) merges the results of Equations (1) and (2) to finally compute the linear hybridization of P_3 and P_2 , in terms of betweenness centrality and tie strength of contact points. Furthermore, in the experimentation procedure, α is utilized in Equation (4) below to set a threshold for to Equation (3), so that linear hybrid recommendations related to COVID-19 contact points can effectively be determined and generated.

$$PI(P_3, P_2) \geq \alpha. \quad (4)$$

The proposed *SARPPIC* algorithm (Algorithm 2) declares relevant variables in steps 2-4; the computations and hybridizations of BC and tie strength in relation to contact points are shown in steps 6-12. The generation of linear hybrid recommendations of probable people infected with COVID-19 is illustrated in the steps 13-17, which are the final steps of the proposed *SARPPIC* algorithm.

4. Performance Evaluation of SARPPIC

This section presents a sequence of scientific benchmarking experiments to validate the performance of *SARPPIC*. The scientific experimentation procedure compared *SARPPIC* to similar algorithmic methods in [8, 28], respectively, represented as COV-1 and COV-2. The benchmarking experiments were done using computers with the following specifications: Microsoft Windows 64-Bit, 8 GB RAM, 500 HDD, and 3.90 GHz dual intel core processors.

4.1. Datasets and Evaluation Metrics. In order to achieve favourable and reliable experimental results, during the

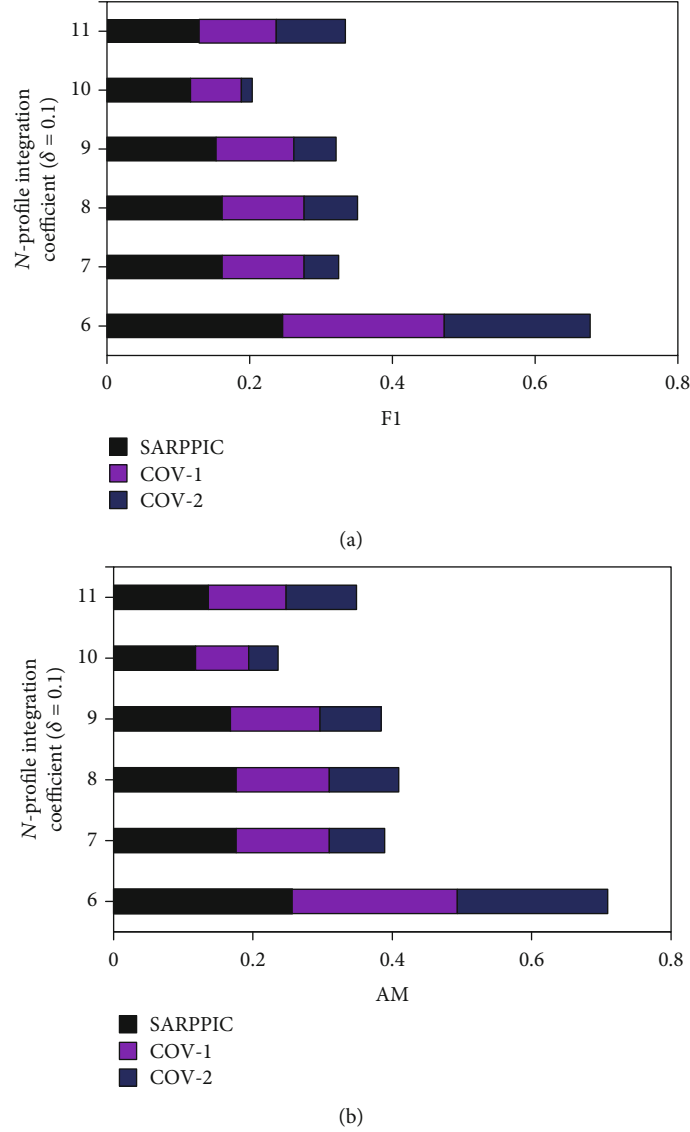


FIGURE 7: (a) F1 performance results on datasets ($\delta = 0.1$); (b) AM performance results on dataset ($\delta = 0.1$).

scientific experimentation process, two real-world datasets were interconnected and utilized, namely, HEXACO-60 dataset which is available in IEEE Data Port (doi:10.21227/phht-pn81) and the ATU dataset in SARVE-2 [31].

As shown in Table 2, the HEXACO-60 dataset contains 60 contact points and a total of 249 betweenness centrality data. In Table 2, the contact points with the highest betweenness centrality initially required for experimentation in accordance to the SARPPIC algorithm are C02, C08, and C14 (26), C38, C32, and C26 (19), H06, H12, and H16 (31), and X40, X34, and X28 (23).

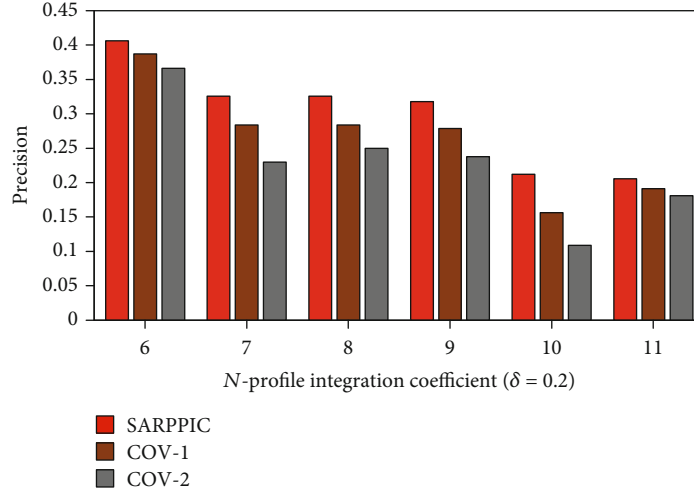
In relation to social tie data, Tables 3 and 4 illustrate the details of past and present tie strength data in [31]. The interconnected datasets were divided into 80% and 20% for the training and test sets, respectively.

During the experimentation process, four commonly used evaluation metrics were employed: (a) precision (P), which is the ratio of the number of probable people (contact point) infected with COVID-19 in the $top-N$ recommenda-

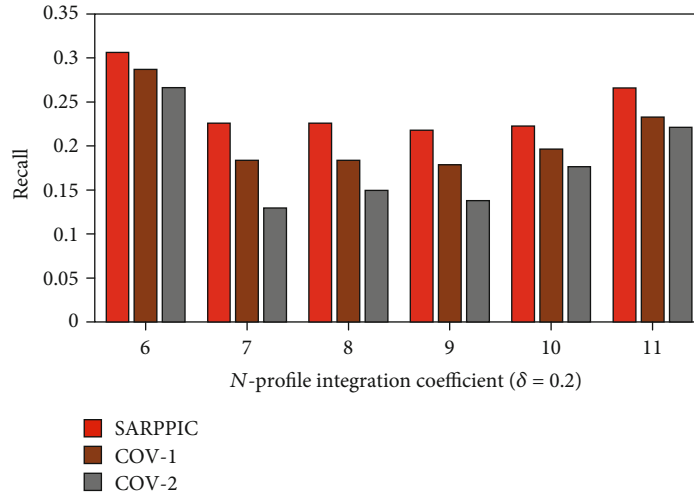
tion list denoted as $\text{Num}(N, d)$ to the length of the same list denoted as $\text{Num}(N)$; (b) recall (R), which is the ratio of the number of probable people (contact points) infected with COVID-19 in the $top-N$ recommendation list represented as $\text{Num}(N, d)$ to the total number of all contact points in the social network graph represented $\text{Num}(d)$; (c) F1 is a harmonic mean of precision and recall; and (d) AM is the arithmetic mean of precision and recall. The equations used for computations of these utilized metrics are illustrated in Equations (5), (6), (7) and (8).

$$P = \frac{\text{Num}(N, d)}{\text{Num}(N)}, \quad (5)$$

$$R = \frac{\text{Num}(N, d)}{\text{Num}(d)}, \quad (6)$$



(a)



(b)

FIGURE 8: (a) Precision performance results on dataset ($\delta = 0.2$); (b) recall performance results on dataset ($\delta = 0.2$).

$$F1 = \frac{2 \times P \times R}{P + R}, \quad (7)$$

$$AM = \frac{1}{2}(P + R). \quad (8)$$

4.2. Baseline Methods and Experimental Parameters. As stated in Section 4 above, the benchmarking experiments involved the performance comparison of *SARPPIC* to the methods in [28, 29] denoted as *COV-1* and *COV-2*, respectively. These methods were selected due to their high relevance and similarity to *SARPPIC* in terms of COVID-19 contact tracing.

The benchmarking experiments are aimed at alleviating data sparsity and cold-start challenges [4–6]. In order to substantiate the experimental results, the following questions required answers:

- (1) In comparison to *COV-1* and *COV-2*, what was the overall performance of *SARPPIC*?

- (2) In comparison to *COV-1* and *COV-2*, how did *SARPPIC* perform in terms of cold-start and data sparsity reduction?

In relation to experimental parameters, the computations of the profile integration coefficients ranged from 6.0 to 11.0. Consequently, profile integration coefficients (N) ranging from 6.0 to 11.0 were utilized for testing and the rest of the computed data for training. During the experimentation process, it was noticed that N results from between 8.0 and 11.0 were more reliable and favourable for effective generation of recommendations involving probable people infected with COVID-19. With reference to Equation (4), the above range (0.8 to 11.0) was therefore used as a threshold to substantiate recommendation quality and accuracy.

4.3. Experimental Results and Analysis. In this section, experimental results and analysis are presented. Figures 6 and 7 show the comparison results of *SARPPIC*, *COV-1*, and *COV-2* (when δ equals 0.1) on the HEXACO-60-ATU

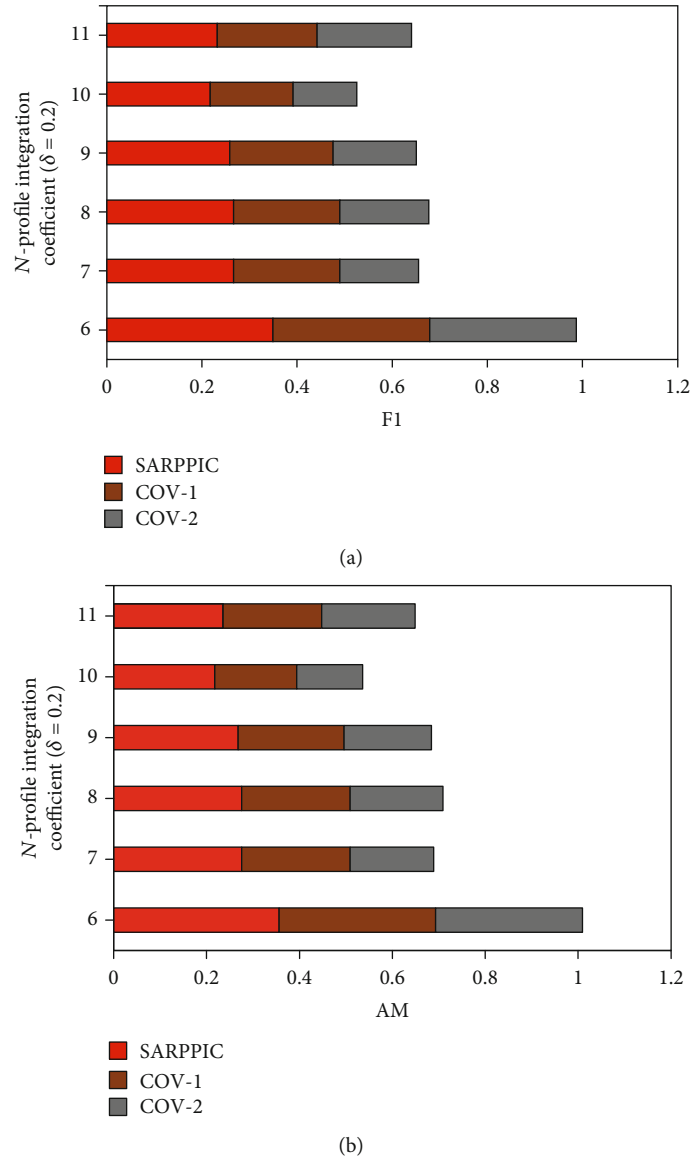


FIGURE 9: (a) F1 performance results on dataset ($\delta = 0.2$); (b) AM performance results on dataset ($\delta = 0.2$).

dataset. As can be verified from this figure, *SARPPIC* achieves much larger values of precision, recall, F1, and AM than *COV-1* and *COV-2* for *different top-N* recommendations in terms of *N*-profile integration coefficients (6.0 to 11.0). In particular, *SARPPIC* significantly achieves high improvement (approximately 10% in precision—Figure 6(a); 16% in recall—Figure 6(b); 49% in F1—Figure 7(a); and 13% in AM—Figure 7(b)), when *N* is 11.

In addition, further comparison of experimental results of *SARPPIC*, *COV-1*, and *COV-2* (when δ equals 0.2) on the HEXACO-60-ATU dataset is shown in Figures 8 and 9. From these figures, it can also see that as the value of *N* is increasing, *SARPPIC* always achieves large values of precision, recall, F1, and AM than *COV-1* and *COV-2* (approximately 20% in precision—Figure 8(a); 26% in recall—Figure 8(b); 79% in F1—Figure 9(a); and 23% in AM—Figure 9(b)), when *N* is 11. As shown in Tables 5 and 6, the experimental process authenticates that the AM results

TABLE 5: *P*, *R*, F1, and AM performance on dataset ($\delta = 0.1$).

Method	Highest <i>N</i>	Precision	Recall	F1	AM
COV-1	11.0	0.09	0.13	0.10	0.11
<i>SARPPIC</i>	11.0	0.10	0.16	0.12	0.13
COV-2	11.0	0.08	0.12	0.09	0.10

TABLE 6: *P*, *R*, F1, and AM performance on dataset ($\delta = 0.2$).

Method	Highest <i>N</i>	Precision	Recall	F1	AM
COV-1	11.0	0.19	0.23	0.20	0.21
<i>SARPPIC</i>	11.0	0.20	0.26	0.23	0.24
COV-2	11.0	0.18	0.22	0.19	0.20

TABLE 7: Proposed algorithm comparison with similar algorithms (advantages and disadvantages).

Criteria	Algorithms	
	<i>SARPPIC</i>	COV-1 and COV-2
Recommendation entities	Profile integration of social properties, i.e., BC and tie strength/social ties as entities for recommendation which is very appropriate for COVID-19 contact tracing.	These algorithms do not utilize social properties, i.e., BC and tie strength as entities for recommendation.
Cold-start and data sparsity challenges	Reduction of cold-start and data sparsity challenges due to its (<i>SARPPIC</i> 's) capability of utilizing social properties, i.e., tie strength/social ties (through contact durations and frequencies) and BC (through shortest paths).	These algorithms utilize traditional collaborative filtering (CF) methods as entities and therefore the effect of cold start and data sparsity is not as minimal as compared to that of <i>SARPPIC</i> due to less social property inclusion.
Algorithm performance in terms of evaluation metrics	In terms of utilized evaluation metrics, namely, precision, recall, F1, and AM (Tables 5 and 6), <i>SARPPIC</i> outperforms COV-1 and COV-2 in relation to effective generation of people-to-people recommendations (COVID-19 patients) due to robustness, suitability, and effective social property inclusion for efficient contact tracing.	In terms of utilized evaluation metrics, namely, precision, recall, F1, and AM (Tables 5 and 6), COV-1 and COV-2 do not perform to the level of <i>SARPPIC</i> in relation to people-to-people recommendations (COVID-19 patients) due to nonutilization of social properties for efficient contact tracing.

achieved for *SARPPIC* are the higher comparison to that of F1 (harmonic mean), which accordingly substantiates that AM should always be higher than F1 regarding the retrieval effectiveness of a recommender system/algorithm [32].

These experimental results on the HEXACO-60-ATU dataset demonstrate that *SARPPIC* significantly outperforms COV-1 and COV-2 in terms of the four evaluation metrics. Furthermore, this indicates that the hybridized profile integration of high BC and tie strength is able to help generate more accurate social recommendations relating to probable people infected with COVID-19.

In summary, Figures 6–9 as well as Tables 5 and 6 show that *SARPPIC* reliably attained more promising results in all the utilized evaluation metrics. Furthermore, Table 7 depicts the comparison of *SARPPIC* to similar algorithms with relevant explanations in each case. These observations corroborate that *SARPPIC* is more suitable, robust, and reduces cold-start and data sparsity challenges due to its capability to utilize profile integration of BC and tie strength. Furthermore, in relation to prediction/recommendation accuracy, the experimental results also depict the importance of social properties in comparison to traditional CF for the generation of effective people-to-people recommendations of probable people with COVID-19.

5. Discussion and Concluding Remarks

Due to the current global COVID-19 pandemic, which is causing high fatality rates, a novel people-to-people (social) recommendation method called *SARPPIC* has been proposed. *SARPPIC* hybridizes two main social properties, namely, betweenness centrality and social ties/tie strength. *SARPPIC* initially applies betweenness centrality to compute the highest betweenness of contact point transmissions in the social graph network, which have transmission authority over other contact points. Secondly, the tie strengths of identified contact points with high betweenness and other target

contact points are computed. The *SARPPIC* recommender algorithm then hybridizes the above computations involving high betweenness and tie strengths to generate recommendations of probable people infected with COVID-19.

Extensive benchmarking experiments on two interconnected real-world datasets have been conducted to validate the effectiveness of *SARPPIC* in terms of people-to-people recommendation for contact points in a social graph network with high betweenness centrality and strong social ties. Experimental results clearly illustrate the effectiveness of *SARPPIC* in comparison to other contemporary methods in terms of appropriate evaluation metrics. The consideration and utilization of other social properties such as closeness centrality, degree centrality, and social neighbours as well as the concept of personality in the current COVID-19 pandemic need to be explored. Future work will focus on these research issues.

Data Availability

The (data type) data used to support the findings of this study are included within the article. Two real-world datasets were interconnected and utilized, namely, HEXACO-60 dataset which is available in IEEE Data Port at doi:10.21227/phht-pn81 and the ATU dataset in SARVE-2 [31] available at doi:10.1109/TETC.2018.2854718.

Conflicts of Interest

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

Supplementary Materials

The supplementary material of this paper elaborates on the dataset which was utilized for the experimentation process. A dataset description of the HEXACO-60 and ATU datasets, which can be accessed through relevant links, is

appropriately provided in the supplementary material. (*Supplementary Materials*)

References

- [1] L. Ferretti, C. Wymant, M. Kendall et al., “Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing,” *Science*, vol. 368, no. 6491, pp. 1–7, 2020.
- [2] A. Remuzzi and G. Remuzzi, “COVID-19 and Italy: what next?,” *The Lancet*, vol. 395, no. 10231, pp. 1225–1228, 2020.
- [3] A. J. Kucharski, T. W. Russell, C. Diamond et al., “Early dynamics of transmission and control of COVID-19: a mathematical modelling study,” *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 553–558, 2020.
- [4] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2013.
- [5] F. Xia, N. Y. Asabere, H. Liu, Z. Chen, and W. Wang, “Socially aware conference participant recommendation with personality traits,” *IEEE Systems Journal*, vol. 11, no. 4, pp. 2255–2266, 2017.
- [6] N. Y. Asabere, A. Acakpovi, and M. B. Michael, “Improving Socially-Aware Recommendation Accuracy Through Personality,” *IEEE Transactions on Affective Computing*, vol. 9, no. 3, pp. 351–361, 2018.
- [7] W. Wobcke, A. Krzywicki, Y. S. Kim et al., “A Deployed People-to-People Recommender System in Online Dating,” *AI Magazine*, vol. 36, no. 3, pp. 5–18, 2015.
- [8] X. Cai, M. Bain, A. Krzywicki et al., “Collaborative filtering for people to people recommendation in social networks,” in *Australasian Joint Conference on Artificial Intelligence*, pp. 476–485, Adelaide, Australia, 2010, Springer, Berlin, Heidelberg.
- [9] N. Vastardis and Kun Yang, “Mobile Social Networks: Architectures, Social Properties, and Key Research Challenges,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1355–1371, 2013.
- [10] C. Tsai, “A Personalized People Recommender System Using Global Search Approach,” in *iConference 2016 Proceedings*, Philadelphia, March 2016.
- [11] A. Krzywicki, W. Wobcke, X. Cai et al., “Using a critic to promote less popular candidates in a people-to-people recommender system,” in *Proceedings of the Twenty-Fourth Conference on Innovative Applications of Artificial Intelligence*, pp. 2305–2310, Toronto, ON, Canada, 2012.
- [12] S. Bourke, K. McCarthy, and B. Smyth, “Power to the people: exploring neighbourhood formations in social recommender system,” in *Proceedings of the Fifth ACM conference on Recommender systems - RecSys '11*, pp. 337–340, Chicago, IL, USA, 2011.
- [13] S. K. R. Unnithan, B. Kannan, and M. Jathavedan, “Betweenness Centrality in Some Classes of Graphs,” *International Journal of Combinatorics*, vol. 2014, 12 pages, 2014.
- [14] A. Samad, M. A. Islam, M. A. Iqbal, and M. Aleem, “Centrality-Based Paper Citation Recommender System,” *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 6, no. 19, p. 159121, 2019.
- [15] R. L. L. Sie, B. J. van Engelen, M. Bitter-Rijpkema, and P. B. Sloep, “COCOON CORE: CO-author REcommendations Based on Betweenness Centrality and Interest Similarity,” in *Recommender Systems for Technology Enhanced Learning*, N. Manouselis, H. Drachler, K. Verbert, and O. Santos, Eds., Springer, New York, NY, 2014.
- [16] I. A. Christensen and S. Schiaffino, “Social influence in group recommender systems,” *Online Information Review*, vol. 38, no. 4, pp. 524–542, 2014.
- [17] H. Mahyar, K. E. Ghalebi, S. M. Morshedi, S. Khalili, R. Grosu, and A. Movaghar, “Centrality-based group formation in group recommender systems,” in *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 1187–1196, Perth Australia, 2017.
- [18] A. Pucci, M. Gori, and M. Maggini, “A Random-Walk Based Scoring Algorithm Applied to Recommender Engines,” in *International Workshop on Knowledge Discovery on the Web*, pp. 127–146, Philadelphia, PA, USA, 2006.
- [19] D. Sulieman, M. Malek, H. Kadima, and D. Laurent, *Combining Social and Semantic Information for Recommendation: Comparative Study*, Hermès Science Publication, Howick Place, London, 2013.
- [20] G. de Souza Junior, C. M. Justel, and J. C. Duarte, “Recommendation system for social networks based on the influence of actors through graph analysis,” in *Proceedings of the 18th Latin-Iberoamerican Conference on Operations Research, CLAIO*, pp. 1–8,antiago, Chile, 2016.
- [21] O. Oechslein and T. Hess, “The value of a recommendation: the role of social ties in social recommender systems,” in *2014 47th Hawaii International Conference on System Sciences*, pp. 1864–1873, Waikoloa, HI, USA, 2014.
- [22] F. Xia, L. Liu, J. Li, J. Ma, and A. V. Vasilakos, “Socially aware networking: a survey,” *IEEE Systems Journal*, vol. 9, no. 3, pp. 904–921, 2015.
- [23] F. Xia, N. Y. Asabere, J. J. P. C. Rodrigues, F. Basso, N. Deonauth, and W. Wang, “Socially-aware venue recommendation for conference participants,” in *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Automatic and Trusted Computing*, pp. 134–141, Vietri sul Mere, Italy, 2013.
- [24] N. Y. Asabere, F. Xia, W. Wang, J. J. P. C. Rodrigues, F. Basso, and J. Ma, “Improving Smart Conference Participation Through Socially Aware Recommendation,” *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 5, pp. 689–700, 2014.
- [25] N. Y. Asabere, F. Xia, Q. Meng, F. Li, and H. Liu, “Scholarly paper recommendation based on social awareness and folksonomy,” *International Journal of Parallel, Emergent and Distributed Systems*, vol. 30, no. 3, pp. 211–232, 2014.
- [26] F. Xia, N. Y. Asabere, H. Liu, N. Deonauth, and F. Li, “Folksonomy based socially-aware recommendation of scholarly papers for conference participants,” in *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*, pp. 781–786, Seoul, South Korea, 2014.
- [27] H. Jang, S. P. Choe, S. N. Gunkel, S. Kang, and J. Song, “A system to analyze group socializing behaviors in social parties,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 6, pp. 801–813, 2016.
- [28] J. Abeler, M. Bäcker, U. Buermeyer, and H. Zillessen, “COVID-19 contact tracing and data protection can go together,” *JMIR mHealth and uHealth*, vol. 8, no. 4, p. e19359, 2020.
- [29] T. M. Yasaka, B. M. Lehrich, and R. Sahyouni, “Peer-to-peer contact tracing: development of a privacy-preserving

- smartphone app,” *JMIR mHealth and uHealth*, vol. 8, no. 4, p. e18936, 2020.
- [30] D. A. Drew, L. H. Nguyen, C. J. Steves et al., “Rapid implementation of mobile technology for real-time epidemiology of COVID-19,” *Science*, vol. 368, no. 6497, pp. 1362–1367, 2020.
- [31] N. Y. Asabere, B. Xu, A. Acakpovi, and N. Deonauth, “SARVE-2: exploiting social venue recommendation in the context of smart conferences,” *IEEE Transactions on Emerging Topics in Computing*, 2018.
- [32] E. Baykan, M. Henzinger, L. Marian, and I. Weber, “A comprehensive study of features and algorithms for URL-based topic classification,” *ACM Transactions on the Web*, vol. 5, no. 3, pp. 1–29, 2011.

Research Article

A Simple Framework of Smart Geriatric Nursing considering Health Big Data and User Profile

Shijie Li¹ and Yongchuan Tang² 

¹Department of Critical Care Medicine, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

²School of Big Data and Software Engineering, Chongqing University, Chongqing 401331, China

Correspondence should be addressed to Yongchuan Tang; tangyongchuan@cqu.edu.cn

Received 22 June 2020; Revised 13 September 2020; Accepted 25 September 2020; Published 23 October 2020

Academic Editor: Mirian C. D. Pinheiro

Copyright © 2020 Shijie Li and Yongchuan Tang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The National Bureau of Statistics of China shows that the population over 65 years old in China exceeds 166 million accounting for 11.93% of the total population by the end of 2018. The importance and severity of taking care of the elderly are becoming increasingly prominent. High-quality and meticulous care for the daily life of the elderly needs helpful and advanced sciences and technologies. Smart geriatric nursing is a must. Basing on the professional knowledge of geriatric nursing, this paper proposes a framework of smart geriatric nursing which consists of three aspects of smart nursing: smart geriatric nursing in physical health using biosensor and advanced devices, smart geriatric nursing in mental health based on user profile, and smart geriatric nursing for daily life based on big data in health. The deployment of the proposed method relies on the technologies of the Internet of Things (IoT), user profile system, big data, and many other advanced information technologies. The framework of methods can provide a useful reference for the systematic technical scheme of smart geriatric nursing in an aging society.

1. Introduction

As a national strategy of China, Healthy China (2019-2030) program is related to the health and happiness of 1.4 billion Chinese people [1]. The program clearly shows that all Chinese people must fully implement the elderly health-promoting activity. The National Bureau of Statistics of China shows that the population over 65 years old in China exceeds 166 million accounting for 11.93% of the total population by the end of 2018. The sense of caring the elderly has become an important consensus of the whole society. High-quality and meticulous care for the daily life of the elderly needs the help of intelligent technologies [2]. This paper proposes a simple framework of smart geriatric nursing which consists of three aspects of smart nursing: smart physical health care, smart mental health care, and smart daily life care for the elderly.

It has been a widely accepted consensus that some advanced information technologies can be applied in health care. Advanced technologies in treatment and nursing field [3–6] include the Internet of Things (IoT) technologies [2],

biosensor technologies, big data [7], and cloud computing [8–12]. With advanced information technologies, the medical efficiency and healthy living standard of the public can be improved significantly [13, 14]. Smart geriatric nursing is a must since China has become an aging society. This paper focuses on exploring and proposing a comprehensive and systematic smart geriatric nursing programs for the elderly's physical health, mental health, and daily life by adopting some of the latest concepts and technologies.

Smart geriatric nursing is based on the IoT technologies, sensors, big data, artificial intelligence, and many other advanced technologies in information society. A systematic framework of smart nursing for the elderly consists of three aspects: smart physical health care for the elderly, smart mental health care for the elderly, and smart daily life care for the elderly. The composition of the smart geriatric nursing method subsystem is shown in Figure 1. Smart geriatric nursing consists of the following three aspects. (1) Smart geriatric nursing in physical health using IoT technology [15, 16] consists of the elderly remote monitoring scheme using biosensor, the elderly medication safety scheme using

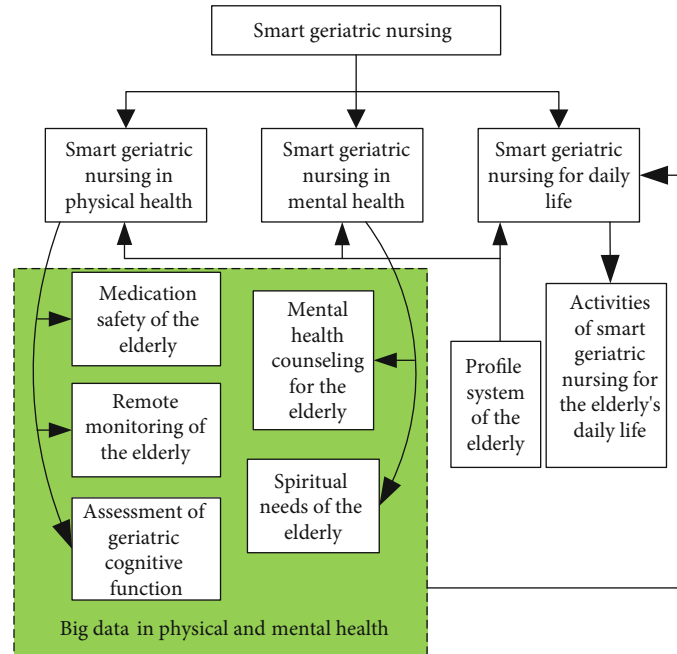


FIGURE 1: The framework of smart geriatric nursing.

radio frequency identification (RFID) technology [17], online (offline) elderly cognitive assessment, and many other intelligent nursing schemes related to physical health. (2) Smart geriatric nursing in mental health is based on user profile system technology [18]. It consists of the elderly profile system, online (offline) elderly psychological consultation (including mental state, psychological needs, and mental state analysis), and other smart nursing methods related to mental health. Many factors affect the mental health of the elderly [19, 20]. (3) Smart geriatric nursing for daily life is based on the big health data [21]. Big health data is derived from the elder's physical health, mental health, and user profile system.

The rest of this paper is organized as follows. The materials and methods in the proposed model are summarized and introduced in Section 2. The smart geriatric nursing in physical health is proposed in Section 3. In Section 4, the framework of smart geriatric nursing in mental health based on user profile is designed. Section 5 is the smart geriatric nursing for daily life based on big data in health. Some limitations of the current work are discussed in Section 6. Section 7 is the conclusion of this paper.

2. Materials and Methods

2.1. Technologies and Tools. The technologies and methods for smart geriatric nursing in physical health include biosensor-based devices and tools, smart medication systems, and smart nursing methods for geriatric common and chronic disease. Tools for mental health mainly include assessment scales, cultural activities, tourism projects, recreational activities, and cognitive training projects. For the elderly's daily life, technology in smart home is helpful.

2.2. Assessment Scales. Some assessment scales are necessary for monitoring the mental health state of the elderly. The research recommends some scales like the Geriatric Depression Scale (GDS) [22, 23], Roy Coping and Adaptation Processing Scale (CAPS-15) [24, 25], and Eysenck Personality Questionnaire (EPQ) [26]. The reliability and validity of the three scales were tested in the community residents in previous studies [23–27]. Since the test results showed a good reliability and validity for the three scales, they could effectively reflect some aspects of the mental health status for the elderly.

2.3. Inclusion and Exclusion Criteria of Areas. To implement the smart geriatric nursing, some advanced technologies and infrastructures should be deployed in the area. As we have known, China is a developing country, and this may be a challenge for some areas. The included cities and areas for smart geriatric nursing should meet the following criteria. (1) The 4G network covers the selected community. Of course, 5G network is much better. (2) The designated hospitals in this area should have mature technologies and devices to get the big data of the elderly for smart geriatric nursing. (3) The health big data is available. Records in the big data include the residents' physical condition, psychological condition, cognitive function status, living condition, and family economic status. (4) There are enough professional medical staffs. Many other criteria are also needed in practical application in a specific area.

3. Smart Geriatric Nursing in Physical Health

The smart geriatric nursing in physical health is aimed at helping the older people keep physical health by curing the disease and repairing the damaged organs. According to

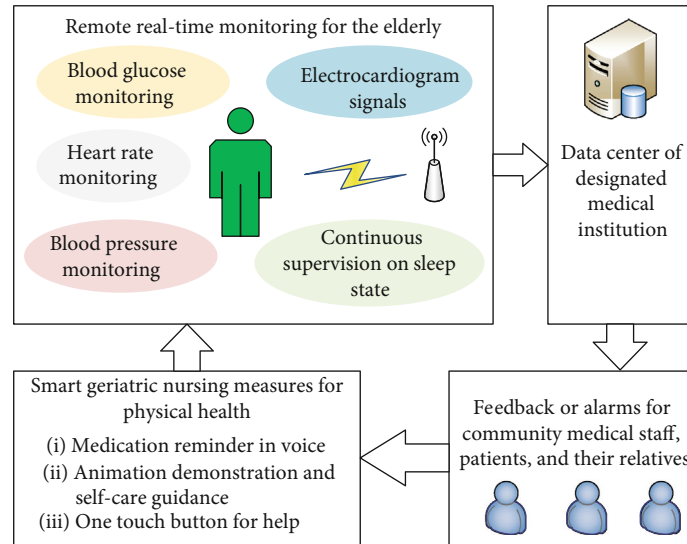


FIGURE 2: Remote real-time monitoring for physical health of the elderly.

smart geriatric nursing in physical health, older people can get recovery from unhealthy condition, get back to normal life, keep healthy, and improve the quality of daily life. Smart geriatric nursing for physical health should be based on IoT technologies. In detail, the smart nursing strategies include but should be not limited to the following:

- (1) Biosensor-based real-time monitoring and remote management of geriatric physical activities and vital signs [28–31]
- (2) Medication safety and management with smart technologies including RFID, sensor, electronic tag, two-dimension code, and global positioning system (GPS) [32–36]
- (3) Smart assessment of geriatric cognitive function with online and offline methods [37–40], e.g., estimating the influence of healthy lifestyle on episodic memory among adults who have subjective memory complaints [40]
- (4) Smart nursing methods for geriatric common and chronic diseases [41–43]

In this section, we construct the framework of smart geriatric nursing in physical health by demonstrating two scenarios. The first scenario is biosensor-based real-time remote monitoring and management of geriatric physical activities and vital signs. The second scenario is medication safety and management with smart technologies.

3.1. Remote Real-Time Monitoring for Physical Health of the Elderly. Remote monitoring based on biosensor can be a real-time technology for smart geriatric nursing in physical health. As shown in Figure 2, the physical activities and vital signs under remote and real-time monitoring include but not limited to regular measurement of blood glucose, blood pressure measurement, cardiac function monitoring and early

warning, and continuous supervision of sleep state. Smart homes can be helpful for smart geriatric nursing with the help of advanced technologies. However, it may also have a negative effect on physical functioning and cause the depression in older adults with chronic conditions [44].

An example of remote real-time monitoring for physical health of the elderly is the regular monitoring of blood glucose. The geriatric diabetic patient should follow the doctor's advice of measuring blood glucose regularly personally or with the help of a nurse according to intelligent devices. The results should be uploaded to the individual electronic medical record for smart monitoring and management. The data centre of designated medical institution records each piece of health evidence for the elderly and transfers the abnormal health data to the local community health centre where the family doctor (team), nursing staff, the elderly themselves or their relatives, and other relevant parties who have the authority to obtain the health information will be responsible for the health condition of the elderly. Smart geriatric nursing measures, such as medication reminder in voice, animation demonstration for self-care guidance, and one touch button for help, will be available whenever needed to help the elderly return to normal living condition and keep physical health by curing the disease and repairing the damaged organs. Above all, an automatic, dynamic, remote, real-time, and closed-loop monitoring and reacting system of smart geriatric nursing for physical health should be constructed to protect the health of the elderly.

3.2. Medication Safety and Management with Smart Technologies. Medication safety and management are a critical problem for the elderly patients [45]. The elderly is more likely to suffer from chronic diseases than the younger people. Meanwhile, medication errors are a risk factor among the elderly because of memory decay and other problems related to aging. Consequently, medication safety and management with smart technologies for the elderly are a must. A smart pill container, which is based on intelligent

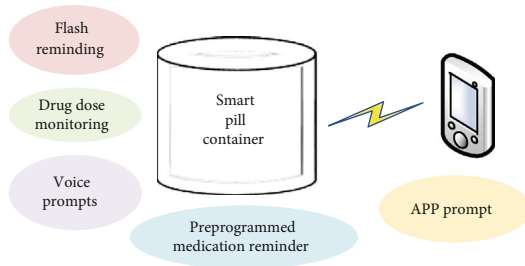


FIGURE 3: Smart nursing for medication safety using the smart pill container.

technologies such as RFID, sensor, wireless communication, and location, can contribute to medication safety for the elderly in cooperation with remote medication guidance whenever and wherever needed.

Smart medication systems are helpful for improving the level of geriatric nursing [46]. Firstly, it can detect drug errors quickly and easily in comparison with manual medication distribution. Secondly, it can improve safety and efficacy by ensuring that older people receive the right dose at the right time. Finally, it can reduce the burden of nursing work no matter in the hospital or at home. As shown in Figure 3, with the help of the IoT technology, especially for the sensor-related technologies, the smart pill container has the functions of flash reminding, drug dose monitoring, voice prompts, preprogrammed medication reminder, and so on. The application (APP) in a smart mobile phone is helpful for realizing the aforementioned smart geriatric nursing measure. More details on smart medicine container can be found in [47–50]. In addition, with the help of intelligent devices such as the smart pill container and mobile phone APPs, we can also popularize the knowledge of (1) pharmacokinetics and pharmacodynamics, (2) principles for the use of medication, (3) smart nursing measures of commonly adverse drug reactions and events, and (4) medication guidance for home and outdoor among the elderly.

4. Smart Geriatric Nursing in Mental Health Based on User Profile

Smart geriatric nursing for mental health means considering the common psychological states, needs, and problems among the elderly regarding geriatric mental health standard to help the elderly to understand, identify, and deal with common mental problems and mental disorders correctly. Consequently, family members, doctors, and nurses can find emotional sustenance methods to contribute a high quality of the elderly's life. Smart mental health care for the elderly based on the user profile system technology uses a geriatric user profile system technology as is shown in Figure 4. The user profile system is designed to digitize and automate information which consists of a model of the basic information, preference information, and dynamic change information of the elderly, by which we can get the geriatric user profile. After modelling the geriatric user profile, we can apply smart mental health care for the elderly. The smart geriatric nursing

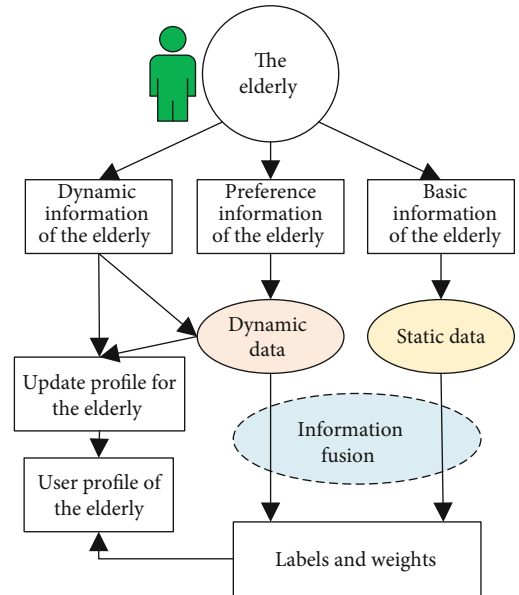


FIGURE 4: User profile system of the elderly (adopted and modified from [18]).

in mental health based on the user profile is shown in Figure 5.

4.1. User Profile System of the Elderly for Smart Nursing in Mental Health. The basic framework of the user profile system for smart geriatric nursing in mental health is described in Figure 4. Through information fusion theory and technology such as Dempster–Shafer evidence theory for uncertain information modelling and processing [51–53], the static data and dynamic data of the elderly care object can be fused and processed. Subsequently, the user label and its weight information are obtained. Finally, the user profile for the elderly is obtained.

The basic information of the elderly including personality characteristics comes from the general information of the elderly. The preference information of the elderly includes hobbies, educational level, and occupational status before and/or after retirement. The dynamic information of the elderly includes changes in their family life, new knowledge from learning in daily life, and the dynamic situation from the surroundings and friends.

4.2. Implementation of Smart Geriatric Nursing in Mental Health. The implementation of smart geriatric nursing in mental health includes two aspects of nursing. On the one hand, the user profile of the elderly using information technology in Figure 4 supplies information that can reflect the psychological status and needs of the elderly. On the other hand, through online and/or offline counseling for the elderly, we can monitor the mental health state of the elderly changing along with the time. The basis of counseling is the well-recognized professional scales. The psychological counseling for the elderly is based on scales such as the Geriatric Depression Scale (GDS) [22, 23], Roy Coping and

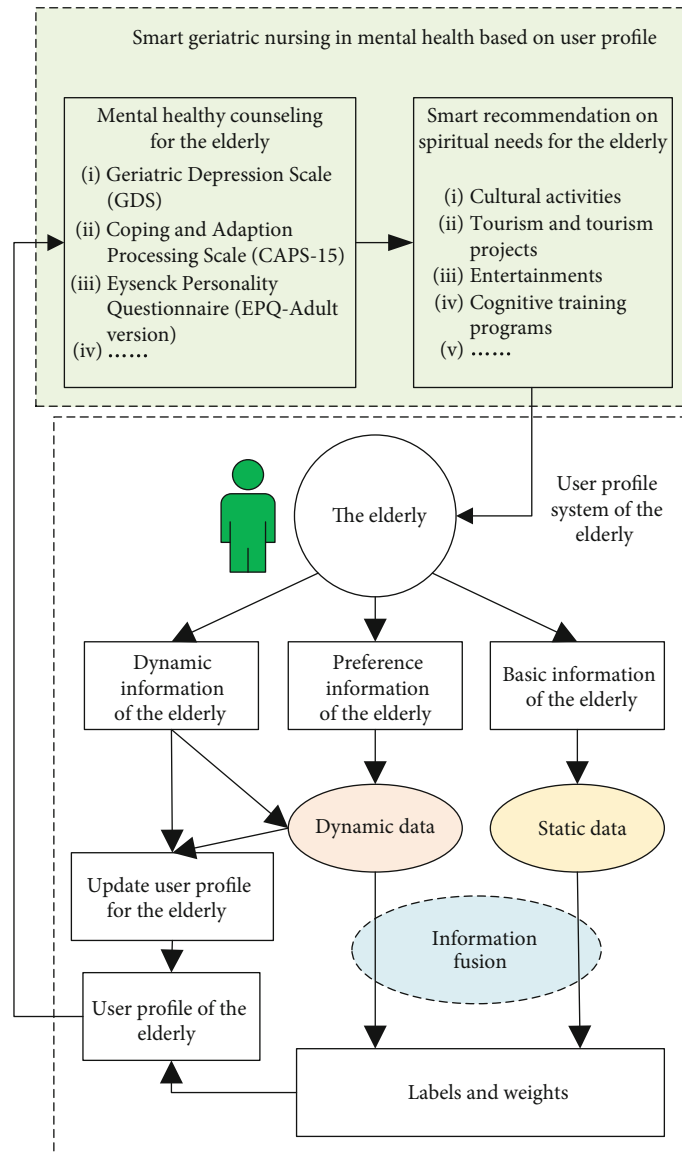


FIGURE 5: Smart geriatric nursing in mental health based on the user profile.

Adaptation Processing Scale (CAPS-15) [24, 25], and Eysenck Personality Questionnaire (EPQ) [26, 27].

As shown in Figure 5, after obtaining the psychological characteristic information of the elderly extracted by the user profile system in Figure 4, combining with the results of psychological counseling for the elderly, personalized mental comfort activities can be recommended intelligently. Recommended activities that can effectively promote the mental health of the elderly include cultural activities, tourism projects, recreational activities, and cognitive training projects.

With the help of smart nursing technology, psychological counseling, and interview activities based on the user profile system of the elderly, smart geriatric nursing in mental health can be implemented and provide more targeted psychological counseling, emotional relief, sadness comfort, daily care, and emotional catharsis for the elderly. Consequently, the risk level and occurrence of depression (emotional handicap), anxiety disorders (seasonal affective disorder) and other

mental disorders, and psychological problems can be reduced. In this way, smart geriatric nursing can help the elderly cope with life in a healthy psychological state and always feel at ease.

5. Smart Geriatric Nursing for Daily Life Based on Big Data in Health

Smart geriatric nursing for daily life means helping the elderly to pursue high quality of life with reasonable arrangement of daily life and helpful assistance. Basing on (1) the user profile information of the elderly provided by Figure 4, (2) healthy big data from remote monitoring of the elderly shown in Figure 2, and (3) mental health big data shown in Figure 5, smart geriatric nursing for daily life can greatly meet the physical and psychological needs of the elderly. Consequently, we can achieve the goal of smart geriatric nursing for daily life based on big data in health.

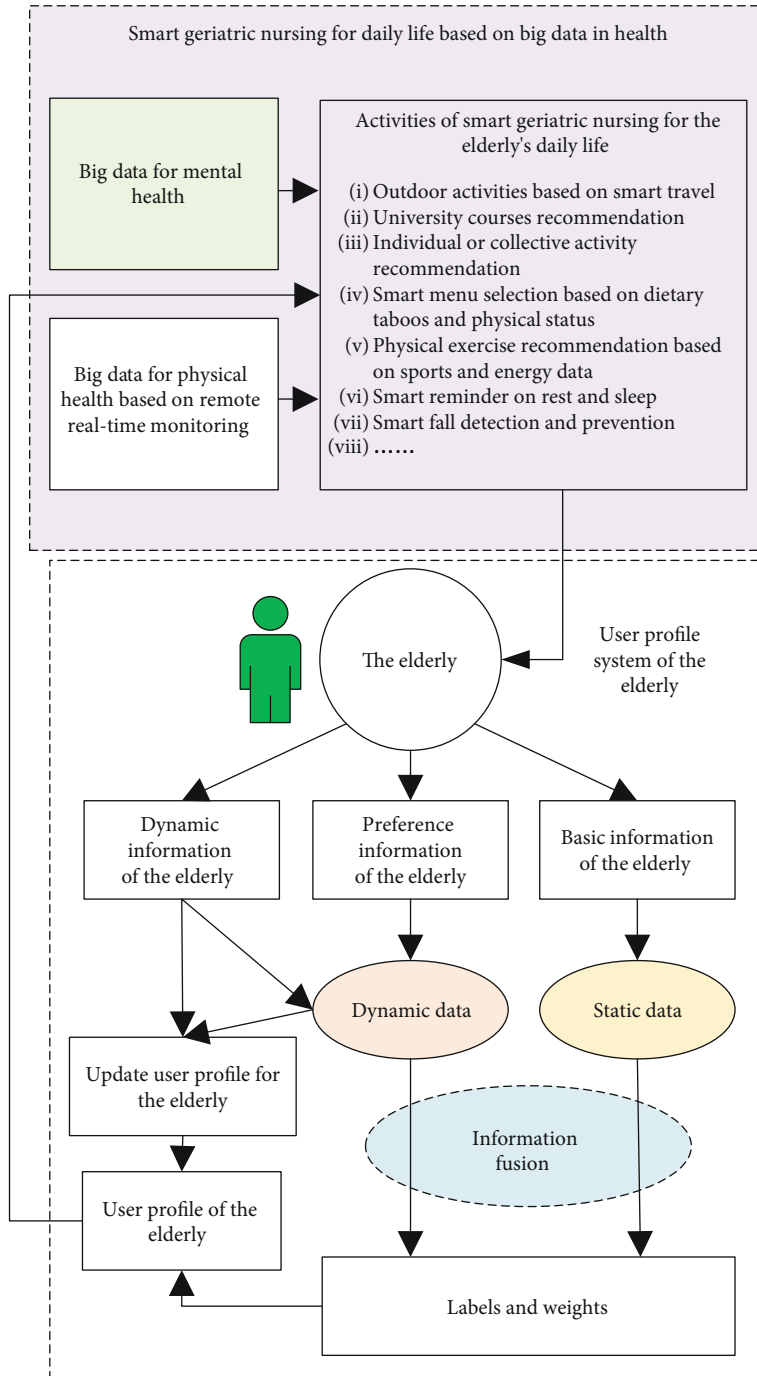


FIGURE 6: Smart geriatric nursing for daily life based on big data for health.

The framework of smart geriatric nursing for daily life based on big data in health is shown in Figure 6. Smart geriatric nursing for daily life includes but is not limited to (1) recommendation of outdoor activities with smart tour for the elderly [54, 55]; (2) research and recommendation of university courses for the elderly or collective activities for the elderly [56]; (3) recommendation of physical status-based dietary habits, food taboos, and perceptions [57, 58]; (4) healthy menu recommendation [59, 60]; (5) data-based physical exercises recommendation [61, 62]; (6) rest and

sleep wisdom reminder; (7) antifall reminder; and (8) cooking safe with smart kitchen [63]. Smart geriatric nursing for daily life should cover the diet and nutrition, environment and life, safety, and activities of the elderly in their daily life.

6. Limitation

There are some limitations in the current work. In general, the framework of smart geriatric nursing designed in this paper is mainly based on an ideal situation with enough

resources in a developed society. Currently, as a developing country, in China, it is hard to fully implement the framework, especially in the rural area of China. But we do believe that the proposed framework is a good idea for smart geriatric nursing.

The big data and user profile of the elderly are the basis of smart geriatric nursing in mental health and for daily life. However, how to manage the privacy during using the big data is a challenge. In fact, protection of peoples' privacy in big data area is a worldwide social problem. Security should be addressed cautiously.

Advanced technologies can be helpful for smart geriatric nursing. However, too many technologies in smart geriatric nursing can be a double-edged sword. On the one hand, technology cannot replace the people to overcome the depression in older adults with chronic conditions. On the other hand, there are limitations in smart geriatric nursing like the economic effectiveness, accessibility, and practicality of some advanced technologies [44].

7. Conclusions

The Healthy China (2019-2030) program puts forward an implementation map for Healthy China strategy made by the Chinese government. Inspired by advanced technologies and the plan of elderly health promotion in the Healthy China program, this paper explores the framework of smart geriatric nursing and contributes a possible technical solution to the aging society. Smart geriatric nursing in physical health can realize a better medical care with smart devices using the technology of IoT. Smart geriatric nursing in mental health based on user profile offers helpful and smart services to the elderly regarding their mental health. In addition, the scales could effectively reflect some aspects of the mental health status for the elderly, which offers helpful information to the medical staff. Smart geriatric nursing for daily life is based on big data in health and serves all needs of the elderly in their everyday life.

The following research of this paper includes but is not limited to (1) more detailed optimization of relevant procedures, methods, and norms of intelligent elderly care, based on cloud computing platform; (2) integration, realization, and collection of related information of big data about geriatric daily nursing by artificial intelligence algorithm instead of the Internet of Things; and (3) continuation of the optimization of related technology and method.

Data Availability

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

Additional Points

Ethical Considerations. Ethical issues (including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy) have been completely observed by the authors.

Conflicts of Interest

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

Acknowledgments

We greatly appreciate the encouragements and supports from Prof. Yanni Yang in the School of Nursing, Army Medical University, China, and Jie Mi in the First Affiliated Hospital of Chongqing Medical University, China. Thanks are due to Mingyu Yang, Shiting Tan, and Xuan Tong in Chongqing University for the valuable discussions on editing the manuscript. This work is partially supported by the Fundamental Research Funds for the Central Universities (No. 2020CDCGRJ037).

References

- [1] W H Organization, *Healthy China: Deepening Health Reform in China: Building High-Quality and Value-Based Service Delivery*, World Bank Publications, 2019.
- [2] V. Tang, K. L. Choy, G. T. S. Ho, H. Y. Lam, and Y. P. Tsang, "An IoMT-based geriatric care management system for achieving smart health in nursing homes," *Industrial Management & Data Systems*, vol. 119, no. 8, pp. 1819–1840, 2019.
- [3] E. Vargiu, J. M. Fernández, M. Gonzales-Gonzales, J. M. Morales-Garzón, K. Prunera-Moreda, and F. Miralles, "A self-management system for complex chronic patients," *International Journal of Integrated Care*, vol. 19, no. 4, 2019.
- [4] S. Mariani, E. Vargiu, M. Mamei, F. Zambonelli, and F. Miralles, "Deliver intelligence to integrate care: the Conne-care way," *International journal of integrated care*, vol. 19, no. 4, 2019.
- [5] E. Baltaxe, C. Embid, E. Aumatell et al., "Integrated care intervention supported by a mobile health tool for patients using noninvasive ventilation at home: randomized controlled trial," *JMIR mHealth and uHealth*, vol. 8, no. 4, 2020.
- [6] S. Montagna, D. Castro Silva, P. Henriques Abreu, M. Ito, M. I. Schumacher, and E. Vargiu, "Autonomous agents and multi-agent systems applied in healthcare," *Artificial Intelligence in Medicine*, vol. 96, pp. 142–144, 2019.
- [7] F. Firouzi, A. M. Rahmani, K. Mankodiya et al., "Internet-of-Things and big data for smarter healthcare: from device to architecture, applications and analytics," *Future generation computer systems-the international journal of escience*, vol. 78, pp. 583–586, 2018.
- [8] H. T. Chen, M. H. Tseng, L. Lu et al., "Cloud computing-based smart home-based rehabilitation nursing system for early intervention," *Advanced Science Letters*, vol. 20, no. 1, pp. 218–221, 2014.
- [9] J. Norouzi, A. Yadollahpour, S. A. Mirbagheri, M. M. Mazdeh, and S. A. Hosseini, "Predicting renal failure progression in chronic kidney disease using integrated intelligent fuzzy expert system," *Computational and Mathematical Methods in Medicine*, vol. 2016, Article ID 6080814, 9 pages, 2016.
- [10] L. Ali, S. U. Khan, N. A. Golilarz et al., "A feature-driven decision support system for heart failure prediction based on χ^2 -Statistical model and Gaussian Naive Bayes," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 6314328, 8 pages, 2019.

- [11] J. Wang, Y. Hu, F. Xiao, X. Deng, and Y. Deng, "A novel method to use fuzzy soft sets in decision making based on ambiguity measure and Dempster-Shafer theory of evidence: an application in medical diagnosis," *Artificial Intelligence in Medicine*, vol. 69, pp. 1–11, 2016.
- [12] E. G. Filho, P. R. Pinheiro, M. C. D. Pinheiro, L. C. Nunes, and L. B. G. Gomes, "Heterogeneous methodology to support the early diagnosis of gestational diabetes," *IEEE ACCESS*, vol. 7, pp. 67190–67199, 2019.
- [13] R. Bootsman, P. Markopoulos, Q. Qi, Q. Wang, and A. A. Timmermans, "Wearable technology for posture monitoring at the workplace," *International Journal of Human-Computer Studies*, vol. 132, pp. 99–111, 2019.
- [14] A. Nourani, H. Ayatollahi, and K. Mirnia, "A smart phone application for the mothers of premature infants," *IRBM*, vol. 40, no. 5, pp. 263–269, 2019.
- [15] D. V. Dimitrov, "Medical internet of things and big data in healthcare," *Healthcare informatics research*, vol. 22, no. 3, pp. 156–163, 2016.
- [16] A. M. Rahmani, T. N. Gia, B. Negash et al., "Exploiting smart e-health gateways at the edge of healthcare internet-of-things: a fog computing approach," *Future Generation Computer Systems*, vol. 78, pp. 641–658, 2018.
- [17] L. Catarinucci, D. De Donno, L. Mainetti et al., "An iot-aware architecture for smart healthcare systems," *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 515–526, 2015.
- [18] R. Saia, L. Boratto, and S. Carta, "A semantic approach to remove incoherent items from a user profile and improve the accuracy of a recommender system," *Journal of Intelligent Information Systems*, vol. 47, no. 1, pp. 111–134, 2016.
- [19] K. C. K. Cheung and K.-L. Chou, "Poverty, deprivation, and depressive symptoms among older adults in Hong Kong," *Aging & Mental Health*, vol. 23, pp. 22–29, 2017.
- [20] L. C. Nunes, P. R. Pinheiro, M. C. Dantas Pinheiro, M. Simao Filho, R. E. Comin Nunes, and P. G. C. Dantas Pinheiro, "Automatic detection and diagnosis of neurologic diseases," *IEEE ACCESS*, vol. 7, pp. 29924–29941, 2019.
- [21] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-cps: healthcare cyber-physical system assisted by cloud and big data," *IEEE Systems Journal*, vol. 11, pp. 88–95, 2015.
- [22] A. Chan, "Clinical validation of the Geriatric Depression Scale (GDS) -Chinese version," *Journal of Aging and Health*, vol. 8, pp. 238–253, 2016.
- [23] W. Haavisto and J. B. Boron, "Examination of discrepancies between subjective and objective memory with latent factors of the geriatric depression scale," *Aging & Mental Health*, vol. 23, pp. 475–484, 2018.
- [24] S. Li, Y. Yang, Y. J. Huang, and H. Li, "Reliability and validity of the Chinese version of the short form Coping and Adaptation Processing Scale on Roy model, Chinese," *Nursing Management*, vol. 18, pp. 766–770, 2018.
- [25] C. E. Song, H. Y. Kim, H. S. So, and H. K. Kim, "Reliability and validity of the Korean version of the Coping and Adaptation Processing Scale-Short-Form in cancer patients," *Journal of Korean Academy of Nursing*, vol. 48, no. 3, pp. 375–388, 2018.
- [26] A. Mackinnon, A. F. Jorm, H. Christensen, L. R. Scott, A. S. Henderson, and A. E. Korten, "A latent trait analysis of the Eysenck Personality Questionnaire in an elderly community sample," *Personality & Individual Differences*, vol. 18, no. 6, pp. 739–747, 1995.
- [27] S. Li, Y. Yang, H. Li, H. Li, and J. Li, "Help-seeking intention for early detection of dementia and its related factors of adults with subjective memory complaints," *Chinese Nursing Management*, vol. 19, pp. 359–363, 2019.
- [28] Z. Wang, Z. Yang, and T. Dong, "A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time," *Sensors*, vol. 17, no. 2, p. 341, 2017.
- [29] D. N. Jamal, S. Rajkumar, and N. Ameen, "Remote elderly health monitoring system using cloud-based WBANs," in *Handbook of Research on Cloud and Fog Computing Infrastructures for Data Science*, pp. 265–288, IGI Global, 2018.
- [30] S. Ghanavati, J. Abawajy, and D. Izadi, "An alternative sensor cloud architecture for vital signs monitoring," in *2016 international joint conference on neural networks (IJCNN)*, pp. 2827–2833, IEEE, 2016.
- [31] L. A. Durán-Vega, P. C. Santana-Mancilla, R. Buenrostro-Mariscal et al., "An IoT system for remote health monitoring in elderly adults through a wearable device and mobile application," *Geriatrics*, vol. 4, no. 2, p. 34, 2019.
- [32] H.-Y. Chien, C.-C. Yang, T.-C. Wu, and C.-F. Lee, "Two RFID-based solutions to enhance inpatient medication safety," *Journal of Medical Systems*, vol. 35, no. 3, pp. 369–375, 2011.
- [33] P. Peris-Lopez, A. Orfila, A. Mitrokotsa, and J. C. A. van der Lubbe, "A comprehensive RFID solution to enhance inpatient medication safety," *International Journal of Medical Informatics*, vol. 80, no. 1, pp. 13–24, 2011.
- [34] Y.-C. Yen, N.-W. Lo, and T.-C. Wu, "Two RFID-based solutions for secure inpatient medication administration," *Journal of Medical Systems*, vol. 36, no. 5, pp. 2769–2778, 2012.
- [35] C. Jin, C. Xu, X. Zhang, and F. Li, "A secure ECC-based RFID mutual authentication protocol to enhance patient medication safety," *Journal of Medical Systems*, vol. 40, no. 1, p. 12, 2016.
- [36] A. Q. Andrade and E. E. Roughead, "Consumer-directed technologies to improve medication management and safety," *Medical Journal of Australia*, vol. 210, pp. S24–S27, 2019.
- [37] D. G. Blazer and R. B. Wallace, "Cognitive aging: what every geriatric psychiatrist should know," *The American Journal of Geriatric Psychiatry*, vol. 24, no. 9, pp. 776–781, 2016.
- [38] C. Opdebeeck, A. Martyr, and L. Clare, "Cognitive reserve and cognitive function in healthy older people: a meta-analysis," *Aging, Neuropsychology, and Cognition*, vol. 23, pp. 40–60, 2015.
- [39] B. Pfistermeister, T. Tümena, K.-G. Gaßmann, R. Maas, and M. F. Fromm, "Anticholinergic burden and cognitive function in a large German cohort of hospitalized geriatric patients," *PLoS One*, vol. 12, no. 2, 2017.
- [40] S. Li and Y. Tang, "Structural model for estimating the influence of healthy lifestyle on episodic memory in adults with subjective memory complaints," *BioMed Research International*, vol. 2020, Article ID 8349819, 2020.
- [41] J.-Y. Kim, N. Liu, H.-X. Tan, and C.-H. Chu, "Unobtrusive monitoring to detect depression for elderly with chronic illnesses," *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5694–5704, 2017.
- [42] M. Moraitou, A. Pateli, and S. Fotiou, "Smart health caring home: a systematic review of smart home care for elders and chronic disease patients," in *GeNeDis 2016*, pp. 255–264, Springer, 2017.
- [43] C. D. Bara, M. Cabrita, H. op den Akker, and H. J. Hermens, "User interaction concepts in smart caring homes for elderly

- with chronic conditions,” in *International Conference on Smart Homes and Health Telematics*, pp. 38–49, Springer, 2015.
- [44] P. Liu, G. Li, S. Jiang et al., “The effect of smart homes on older adults with chronic conditions: a systematic review and meta-analysis,” *Geriatric Nursing*, vol. 40, no. 5, pp. 522–530, 2019.
- [45] E. Metsälä and U. Vaherkoski, “Medication errors in elderly acute care—a systematic review,” *Scandinavian Journal of Caring Sciences*, vol. 28, no. 1, pp. 12–28, 2014.
- [46] R. Turjamaa, S. Kapanen, and M. Kangasniemi, “How smart medication systems are used to support older people’s drug regimens: a systematic literature review,” *Geriatric Nursing*, 2020.
- [47] N. Ratnakar, *Smart medicine container*, vol. 492, US Patent 9, 2016.
- [48] A. Kassem, W. Antoun, M. Hamad, and C. El-Mou Cary, “A comprehensive approach for a smart medication dispense,” *International Journal of Computing and Digital Systems*, vol. 8, no. 2, pp. 131–141, 2019.
- [49] V. Selamneni, P. Barya, N. Deshpande, and P. Sahatiya, “Low-cost, disposable, flexible, and smartphone enabled pressure sensor for monitoring drug dosage in smart medicine applications,” *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11255–11261, 2019.
- [50] W. Antoun, A. Abdo, S. Al-Yaman, A. Kassem, M. Hamad, and C. El-Mou Cary, “Smart medicine dispenser (smd),” in *2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME)*, pp. 20–23, IEEE, 2018.
- [51] D. Wu and Y. Tang, “An improved failure mode and effects analysis method based on uncertainty measure in the evidence theory,” *Quality and Reliability Engineering International*, vol. 36, no. 5, pp. 1786–1807, 2020.
- [52] M. Jing and Y. Tang, “A new base basic probability assignment approach for conflict data fusion in the evidence theory,” *Applied Intelligence*, 2020.
- [53] Z. Liu, Q. Pan, J. Dezert, J.-W. Han, and Y. He, “Classifier fusion with contextual reliability evaluation,” *IEEE transactions on cybernetics*, vol. 48, pp. 1605–1618, 2017.
- [54] T. Batool, Y. Vanrompay, A. Neven, D. Janssens, and G. Wets, “CTASS: an intelligent framework for personalized travel behaviour advice to cardiac patients,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 12, pp. 4693–4705, 2019.
- [55] C. Chuang, “A current travel model: smart tour on mobile guide application services,” *Current Issues in Tourism*, vol. 23, no. 18, pp. 2333–2352, 2020.
- [56] Y. Park, “The social context of collective physical training among Chinese elderly an anthropological case study in a park in Beijing,” *Anthropology & Aging*, vol. 38, no. 1, pp. 30–43, 2017.
- [57] C. Manzo and M. T. Russo, “Dietary calcium intake in a cohort of elderly patients already in drug therapy for osteoporosis. Is it possible and how to modify the eating habits before calcium supplementation?,” *Italian Journal of Medicine*, vol. 10, pp. 42–44, 2016.
- [58] T. C. Mandiwana and V. Masiagwala, “Determination of eating taboos during pregnancy by elderly women at Maungani Village, Vhembe District of Limpopo Province,” *Annals of Nutrition and Metabolism*, vol. 67, p. 104, 2015.
- [59] V. R. Chifu, E. S. Chifu, C. B. Pop, I. Salomie, and A. Niculici, “Optimizing the generation of personalized healthy menus for elderly people using a crab breeding inspired method,” in *2017 13TH IEEE international conference on intelligent computer communication and processing (ICCP)*, pp. 155–162, 2017.
- [60] C. Coenegracht and T. Urgese, “Evaluation of an energy-dense and high-protein menu for the elderly in a geriatric hospital,” *AKTUELLE ERNAHRUNGSMEDIZIN*, vol. 44, pp. 93–98, 2019.
- [61] O. Verschuren, M. D. Peterson, A. C. J. Balemans, and E. A. Hurvitz, “Exercise and physical activity recommendations for people with cerebral palsy,” *Developmental Medicine & Child Neurology*, vol. 58, no. 8, pp. 798–808, 2016.
- [62] N. Feter, R. Alt, M. G. Dias, and A. J. Rombaldi, “How do different physical exercise parameters modulate brain-derived neurotrophic factor in healthy and non-healthy adults? A systematic review, meta-analysis and meta-regression,” *Science & Sports*, vol. 34, no. 5, pp. 293–304, 2019.
- [63] R. Yared and B. Abdulrazak, “Risk analysis and assessment to enhance safety in a smart kitchen,” *Fire Technology*, vol. 54, no. 2, pp. 555–577, 2018.

Research Article

A Probabilistic Patient Scheduling Model with Time Variable Slots

Danae Carreras-García ¹, **David Delgado-Gómez** ¹, **Enrique Baca-García** ^{2,3,4}
and **Antonio Artés-Rodríguez** ⁵

¹Department of Statistics, Universidad Carlos III of Madrid, Universidad Carlos III de Madrid, Leganés, Spain

²Department of Psychiatry, Fundación Jiménez Díaz Hospital, Madrid, Spain

³Madrid Autonomous University, Madrid, Spain

⁴Universidad Católica del Maule, Talca, Chile

⁵Signal Theory and Communications Department, Universidad Carlos III of Madrid, Leganés, Spain

Correspondence should be addressed to Danae Carreras-García; dcarrera@est-econ.uc3m.es

Received 4 July 2020; Revised 31 July 2020; Accepted 10 August 2020; Published 1 September 2020

Academic Editor: Plácido R. Pinheiro

Copyright © 2020 Danae Carreras-García 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.

One of the current challenges faced by health centers is to reduce the number of patients who do not attend their appointments. The existence of these patients causes the underutilization of the center's services, which reduces their income and extends patient's access time. In order to reduce these negative effects, several appointment scheduling systems have been developed. With the recent availability of electronic health records, patient scheduling systems that incorporate the patient's no-show prediction are being developed. However, the benefits of including a personalized individual variable time slot for each patient in those probabilistic systems have not been yet analyzed. In this article, we propose a scheduling system based on patients' no-show probabilities with variable time slots and a dynamic priority allocation scheme. The system is based on the solution of a mixed-integer programming model that aims at maximizing the expected profits of the clinic, accounting for first and follow-up visits. We validate our findings by performing an extensive simulation study based on real data and specific scheduling requirements provided by a Spanish hospital. The results suggest potential benefits with the implementation of the proposed allocation system with variable slot times. In particular, the proposed model increases the annual cumulated profit in more than 50% while decreasing the waiting list and waiting times by 30% and 50%, respectively, with respect to the actual appointment scheduling system.

1. Introduction

One of the current problems faced by health centers is the existence of patients who do not attend their appointments. These patients, commonly known as no-shows, cause damage which includes a drastic reduction in the health center's income and extend patients' access to medical care. No-show rates ranging from 4 to 79.2% [1] and losses reaching 150 million dollars only in the United States [2] have been reported.

In order to reduce these numbers, health centers use reminders and sanctions. However, several studies have shown that these strategies only achieve a slight or moderate reduction in the no-show rates [3, 4]. Moreover, it has been

pointed out that sanctions may limit access to patients with limited income to medical centers [5], and automatic reminder systems might have an important economic impact on the health centers [6].

An alternative to these active strategies is the use of appointment scheduling systems. These systems aim at obtaining an allocation of the patients that reduce given performance measures, such as patient overtime or doctor idle time. The development of these patient scheduling systems is a very active field. Literature review articles covering seminal and more recent approaches include [7–9]. This vast productivity is a consequence of the large variety of peculiarities of the health centers that cause the systems to be practically tailor-made. In general terms, these systems

can be differentiated according to (1) online vs. offline; (2) single vs. multiple servers; (3) appointment rules; (4) performance measures; (5) inclusion of environmental factors; and (6) modeling approach.

Regarding the first classification, systems can consider online (sequential) and offline (simultaneous) scheduling. Primary care centers usually assign the appointment at the time the patient requests it (online scheduling), while specialties assign it days later, after checking the doctors' availability and analyzing the required resources (offline scheduling) [10]. Online systems are more common in practice; however, they are more difficult to model. On the other hand, offline systems are becoming increasingly important as requests can be made automatically, and patients are notified later.

With respect to the second category, queuing theory allows the classification of these systems into single-server or multiple servers according to the number of providers being modelled at a time. The single-server assumption is usually associated with the fact that each doctor has their own set of patients associated. However, it is known that models based on multiple servers are more efficient, and in some cases, such as laboratories or x-rays tests, they are also optimal [7].

In terms of appointment rules, two parameters have to be considered. Firstly, the number of patients scheduled in each time slot (block-size), where the number of patients to be seen in the first slot (initial-block), is usually studied separately. Secondly, the time length of each slot (appointment interval) is being studied also. Each combination of these parameters describes an appointment rule. For example, individual-block/variable-interval describes the rule that assigns only one patient to each slot with variable time length.

Regarding performance measures, these are used to describe the objective pursued when creating a particular scheduling system. Most of the models are built from an optimization problem seeking to achieve the best allocation of patients; although, in some cases, heuristic rules are created that are then validated by simulation. These measures are usually associated with the cost of patient waiting time, or the doctor's idle time, the revenues of attending a patient, or a combination of them.

The inclusion of environmental factors into scheduling systems is still underdeveloped and a large consensus exists on it being one of the most promising lines to explore in future research. Cayirli and Veral [7] point out the importance of including no-shows, walk-in, urgent patient, emergencies, and second consultations in appointments scheduling systems. Although these factors can be addressed separately, through the center's policies, taking them into account can lead to better results. Gupta and Denton [8] add to these late cancellations, which are often classified along with no-shows, and patient preferences. More recently, Ahmadi et al. [9] add another environmental factor such as patient lack of punctuality, physician lateness, service interruption, random service time, and other patient appointment requirements.

As for the mathematical model used for the solution, most scheduling systems make use of stochastic optimization or dynamic stochastic programming scheduling, because these are more robust to random arrivals and random service times. However, the recent availability of Electronic Health

Records (EHR) and advances in data science have made it possible to obtain more accurate predictions about these two factors, random arrivals and service times, making the use of deterministic planning systems possible. Most of the deterministic models are formulated using integer or mixed linear programming models aiming at optimizing some performance measure of the scheduling. This type of model is widely used in specialties, where deterministic service times and near-zero no-show probabilities are assumed [9].

In this paper, we propose a deterministic integer linear programming (ILP) model for offline scheduling of patients in the presence of heterogeneous no-shows and variable times services in a specialty service of a public health center. To the best of our knowledge, this is the first offline scheduling model that considers both heterogeneous patient no-shows and length variable appointment intervals. The system aims to maximize the expected revenues of the clinic considering the different show-rates of each patient during a whole week. The model is designed as a single-server given the fact that each doctor is assumed to have their own list of patients. The appointment rule used is individual-block/variable-interval, with no initial-block, for which only one patient is assigned to each variable time length slot.

In order to validate the model, experiments are carried out to reproduce the routine of the psychiatric department of the Fundación Jiménez Díaz Hospital in Madrid, Spain, for an entire year. In this sense, patient show rates are estimated, and three different appointment intervals with information provided by the center are incorporated into the proposed model. We also take into account other environmental factors such as large waiting lists, major revenues from scheduling new patients, and the dynamic priority assignment scheme. The performance of the model is compared with other scheduling systems, including the one proposed by Ruiz-Hernández et. al [11], which is currently implemented at the psychiatric department of this hospital, showing a considerable improvement.

The rest of the article is structured as follows. In Section 2, we present a literature review in scheduling systems in health centers with heterogeneous no-show probabilities. Next, in Section 2, the probabilistic patient scheduling problem with time-variable slots is introduced. In Section 3, the specific characteristics of scheduling in the health center basis for this study are described. In Section 5, the numerical experiments carried out to evaluate the model are presented and discussed. Finally, the article ends in Section 6 with the conclusions of the results obtained.

2. Literature Review

In this section, we review appointment scheduling systems that take into account variable appointment intervals and heterogeneous no-shows. First, we discuss models that consider different appointment intervals. Then, we move on to models with heterogeneous no-show probabilities. Finally, we will present the contributions of our proposal. A summary with the most recent works compared with the proposed method is presented in Table 1.

TABLE 1: Comparison across recent studies.

Study	Modeling approach	No-shows	Service time	App. interval	Scheduling
Cayirli et al. [12] (2006)	Rules	1-class	Long\short (deterministic)	Variable	Online
Huang and Verduzco [13] (2015)	Rules	1-class	Objective (deterministic)	Variable	Online
Bentayeb et al. [14] (2019)	Rules	1-class	Predicted (deterministic)	Variable	Online
Ratcliffe et al. [15] (2012)	SP (analytical)	2-class (simulated)	Fixed	Fixed	Offline
Muthraman and Lawley [16] (2008)	SP (analytical)	N-class (simulated)	Exponential (stochastic)	Fixed	Online
Zacarias et al. [10] (2014)	SP (analytical)	N-class (simulated)	Log-normal (stochastic)	Fixed	Online
C. Yan et al. [17] (2015)	SP	N-class (simulated)	Exponential (stochastic)	Fixed	Online
Samorani and Harris [18] (2019)	SMIP	N-class (predicted)	Fixed	Fixed	Online
Savelsbergh and Smilowitz [19] (2016)	ILP	N-class	Fixed	Fixed	Offline
Ruiz Hernández et al. [11] (2019)	MIP	Predicted Probabilities	Fixed	Fixed	Offline
<i>Proposed scheduling</i>	<i>ILP</i>	<i>Predicted Probabilities</i>	<i>Variable</i>	<i>Variable</i>	<i>Offline</i>

It is important to point out the difference between the appointment interval and the service time. The former is the scheduled length of an appointment, while the latter is the actual time the patient spends at the appointment. In works considering different appointment intervals, it is usually assumed that the service time is deterministic but unknown, so it can be estimated. Cayirli et al. [12] simulate different sequence and appointment rules on a variety of environmental factors, such as different service times for new and follow-up patients and the presence of homogeneous absences. Huang and Verduzco [13] reclassify patients into different types of visits and determine appointment length by incorporating performance measures such as patient waiting time and physician downtime, in order to converge with the optimal appointment length for each class. Bentayeb et al. [14] developed a new appointment scheduler based on a time-of-service prediction model, which is developed using the data mining method. They use classification and regression trees to predict service times with 84% accuracy. They then simulate different scheduling rules to obtain a better sequence of patients. To the best of our knowledge, there are no research articles addressing variable appointment intervals that use an optimization approach for optimal patient assignment.

On the other hand, systems that take into account heterogeneous no-show probabilities usually follow a stochastic programming approach in terms of the randomness of arrivals and service times. This means that, regardless of the appointment interval (fixed or variable), the service time is assumed to follow a certain probability distribution. These models are computationally intensive, which means that instead of using the probabilities directly, the appointment is normally split according to no-show probabilities. For example, Ratcliffe et al. [15] builds a dynamic stochastic scheduler that maximizes profits by controlling two classes of patients with different show rhythms. They develop analytical bounds and approximations that lead to partially optimal scheduling rules. Muthraman and Lawley [16] create a sequential scheduling model with exponential service times and multiple patient no-show probabilities, yet the appointment interval is constant. Zacarias et al. [10] study the analytical properties of accounting for different class probabilities and different appointment intervals in the scheduling of a full day. For example, they conclude that in the presence of homogeneous probabilities and variable appointment times, the patient should be scheduled according to the rule of the shortest processing time first (SPT). Yan et al. [17] develop a model for scheduling sequential appointments considering patient choice and service fairness simultaneously. They use stochastic programming that uses distinct groups of patients grouped by no-show probabilities and homogeneous appointment intervals. Samorani and Harris [18] determine the impact of the probabilistic classifier in scheduling appointments with no-shows. They try several classifiers to obtain N classes of patients in terms of their probabilities of no-shows. They then use a stochastic mixed-integer scheduler with random arrivals and service time and appointment interval determinants and constants.

An alternative idea to the use of stochastic optimization could be to predict the no-show rates and assume determin-

istic arrivals and service times. The recent availability of Electronic Health Records (EHR) and advances in data science has made it possible to improve this wide variety of scheduling systems. This is because modern predictive techniques applied to EHRs are capable of estimating the probability of patient no-show, which can be used to improve the scheduling system [4]. Regarding deterministic systems, Savelsbergh and Smilowitz [19] are the first to define the probabilities of no-shows for six different categories of patients depending on their preferences (strong or weak) for three different time windows (AM, noon, or PM). These environmental conditions were integrated into an online linear integer program to optimize patient allocation. Later on, Ruiz Hernández et al. [11] proposed a mixed deterministic integer program. The model is probabilistic in the sense that it incorporates the expected income of the center weighted by the probabilities of no-show predicted for each patient. This was the first model to incorporate no-show rates, rather than using an N class approach to obtain different classes of patients in terms of no-show. The present paper proposes an offline scheduling system that extends Ruiz Hernández's work by including the variable appointment interval required for each patient. As it will be seen in the experiments, the inclusion of this information allows to improve considerably the performance of the system.

The contributions of this paper are the following: (1) the inclusion of further heterogeneous probabilities that consider information's about the patient, day and time of the appointment, month, and the indirect waiting time (lead time); (2) the inclusion of the variable appointment interval in a linear binary deterministic problem for an online scheduling system; (3) the development of a model for a weekly scheduling system with dynamic prioritization and differentiation for new patients that maximizes the expected revenue of the center and indirectly minimizes the doctor's idle time; (4) the application of the model to a health center in Madrid that potentially mitigates the effects of patient no-shows.

3. The Probabilistic Patient Scheduling Problem with Time Variable Slots

In this section, the mathematical formulation of the proposed patient scheduling model is presented. As discussed above, it takes into account the patients' no-show probabilities and the consultation times required by each patient. The goal of the model is to maximize the center's expected revenue through the reduction of no-shows. The model distinguishes between two types of patients (first visits and show-up visits). In addition, it takes into account the time the patient has been waiting for an appointment to assign a priority parameter that is updated every week. It also takes into account some policy requirements that set the minimum proportion of first visits that have to be scheduled each week.

Before describing the model, the notation that will be used is presented:

Sets:

I : days of the week;

T : time slots in any given day;

K : set of patients to be scheduled for an appointment during the reference week.

Parameters:

q : proportion of the number of available slots that must be allocated to first visits;

d_k : binary parameter indicating if a patient $k \in K$ has high ($d_k = 0$) or low ($d_k = 1$) priority;

Z_k : binary parameter indicating if a patient $k \in K$ is a first visit ($Z_k = 1$) or follow-up ($Z_k = 0$).

P_{itk} : probability that patient $k \in K$ will show-up to an appointment in $\{i, t\}$, for all $i \in I$ and $t \in T$;

w_z : revenue obtained either from a first visit ($z = 1$) or follow-up ($z = 0$).

t_k : number of time slots required by the patient in their consultation.

h_1 : slack parameter for the minimum number of slots that can be allocated in one day.

h_2 : slack parameter for the maximum number of slots that can be allocated in one day.

Variables:

x_{itk} : binary variable that takes the value 1 if the patient $k \in K$ is assigned to slot $\{i, t\}$, for all $i \in I$ and $t \in T$;

x_k^T : binary variable that takes the value 1 if the patient $k \in K$ is referred back to the waiting list.

With this notation, taking into account that the operator $\lceil \cdot \rceil$ denotes the ceiling function (minimum integer not below), the model is formulated as follows:

$$\max \sum_{i \in I} \sum_{t \in T} \sum_{k \in K} x_{itk} P_{itk} (Z_k w_1 + (1 - Z_k) w_0), \quad (1)$$

$$\text{s.t.} \quad \sum_{k \in K} \sum_{t \in T | \bar{t} \geq t > \bar{t} - t_k} x_{itk} \leq 1, \forall i \in I, t \in T, \quad (2)$$

$$\sum_{i \in I} \sum_{t \in T} x_{itk} + x_k^T = 1, \forall k \in K, \quad (3)$$

$$\sum_{i \in I} \sum_{t \in T} \sum_{k \in K} x_{itk} t_k Z_k \geq b_4, \quad (4)$$

$$\sum_{i \in I} \sum_{t \in T} \sum_{k \in K} x_{itk} t_k (1 - d_k) \geq b_5, \quad (5)$$

$$\sum_{k \in K} \sum_{t \in T | \bar{t} \geq t > \bar{t} - t_k} x_{itk} \geq \sum_{k \in K} \sum_{t \in T | \bar{t} + 1 \geq t > \bar{t} + 1 - t_k} x_{itk}, \quad \forall i \in I, t \in T \setminus |T|, \quad (6)$$

$$\sum_{k \in K} \sum_{t \in T} x_{itk} t_k \geq |T| + h_1, \forall i \in I, \quad (7)$$

$$\sum_{k \in K} \sum_{t \in T} x_{itk} t_k \leq |T| + h_2, \forall i \in I, \quad (8)$$

$$x_{itk}, x_k^T \in \{0, 1\}, \forall i \in I, t \in T, k \in K, \quad (9)$$

where

$$b_4 = \min \left\{ \sum_{k \in K} Z_k t_k, \lceil q |I| |T| \rceil \right\}, \quad (10)$$

$$b_5 = \min \left\{ \sum_{k \in K} (1 - d_k) t_k, |I| |T| - b_4 \right\}$$

The objective function maximizes the clinic's expected revenue. Note that when $w_0 = w_1 = w$, the objective function maximizes the expected show-up rate; that is, it maximizes the weighted show-up rate. The set of constraints (2) ensures that only one patient is seen at a time. The constraints (3) guarantee that if the patient does not schedule in the reference week, they are sent back to the waiting list. As we are working with binary variables, it is also ensured that each patient is not scheduled more than once in a week. Constraint (4) ensures at least the minimum time is used for new patients (first visits). Constraint (5) grants that low priority patients are not scheduled until all high priority patients have been scheduled. Constraints (6) force the next slots of the day to be empty if one slot is. This ensures that either all slots are filled continuously or the rest of the day's slots remain empty. Finally, constraints (7) force the time spent in a day to be within acceptable limits.

It should be noted that this model is an extension of the probabilistic model developed in. That study proposed a model to maximize the expected revenues based on no-show probabilities. It considered the distinction between new (first visits) and old patients and imposed priority for patients with long waiting times. However, the model did not take into account scheduling different appointment times, which can help to attend more patients in a week. Mathematically, this difference is seen in a change from the unit of time of appointment to a 5-minute slot. Moreover, the meaning of assigning a slot to the patient, $k \in K$, changes. In this case, it means that the patient is programmed to enter the appointment in that time slot, while the subsequent $t_k - 1$ slots must all be equal to zero (see equation (2)). Constraints (4) and (5), which are weighted by the time the patient spends in an appointment, also change. Another contribution of our model is that it forces the assigner not to leave empty slots with the set of constraints (6). Finally, as in our proposal, doctor's working time is not restricted to the number of appointments; constraints are added (7) to ensure that the time limits are not exceeded.

4. The Scheduling Process

The scheduling process in the reference health facility used in this study, and for which the model is proposed, works as follows:

- (1) A waiting list is available with the patient's information for the appointment, including the number of weeks on the waiting list (sojourn), whether the patient is a first (new) visit or not, and the patient's consultation time. New patients are added to the list

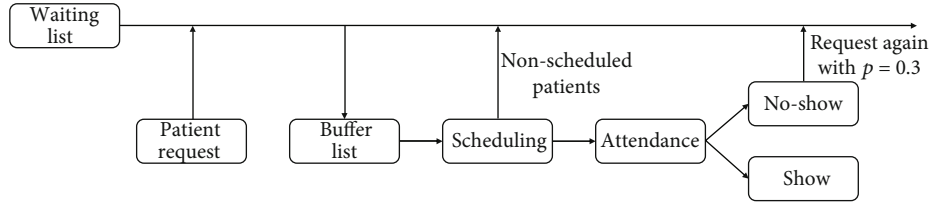


FIGURE 1: Flow chart of the experimental setup.

at the time the appointment is requested and the sojourn is initialized at one

- (2) The list of patients (referred to as a buffer) to be passed to the scheduler each week is constructed as follows:
 - (a) The system first selects the patients with the longest waiting time (sojourn) and assigns them high priority ($d_k = 0$). This group contains both first visits ($Z_k = 1$) and follow-up visits ($Z_k = 0$)
 - (b) If the legal minimum number of slots dedicated to follow-up patients has not been filled ($\lceil (1-q)|I||T| \rceil$), the system sequentially adds patients in decreasing order of sojourn until the previous condition is met (or the waiting list is left empty). In all but the last iteration, patients are assigned high priority ($d_k = 0$). This group contains both first visits ($Z_k = 1$) and follow-up visits ($Z_k = 0$)
 - (c) Finally, if the number of first visits in the buffer is below the legal requirements ($\lceil q|I||T| \rceil$), the system sequentially adds first visits in decreasing order of sojourn until the legal requirements are met or there are no first visits left to include. These patients have low priority ($d_k = 1$) and are first visits ($Z_k = 1$).
- (3) After the buffer is selected, the system passes the lists of candidates to be scheduled to the probabilistic model with variable time (1). Once the appointment schedule is obtained, the patients who did not receive an appointment are sent back to the waiting list with their original sojourn values

5. Numerical Experiments

In order to evaluate the performance of our model, an experiment has been conducted that reproduces the routine of the psychiatric department of the Fundación Jiménez Díaz hospital in Madrid.

5.1. Probabilities Estimation. To estimate the probabilities of no-show for each patient, we used a database with 76,658 appointments belonging to 5261 patients. The average no-show rate on the dataset is 14.05%. Each appointment was described by 97 predictors commonly used to predict no-shows [1]. This set of predictors contained demographic var-

iables, a set of variables that characterize the patient's previous attendance behavior, variables about the patient's condition, and variables related to the appointment. A logistic regression model with L1 regularization was used to obtain the no-show probabilities. This model, commonly known as Lasso Regression, has been previously used to predict no-shows because of its ability to automatically select variables and because of its interpretability [20]. The variables included in the model contain the day of the week, time, and lead time, which allowed to obtain specific and differentiated probabilities for each patient, day of the week, month, time, and sojourn value.

5.2. Experimental Setup. We now describe the procedure used to reproduce the scheduling process during a week in the reference center (presented in Section 3). For an illustration of this process, see Figure 1: flow chart of the experimental setup.

The experiment simulates 48 weeks of a doctor attending patients for six hours from Monday to Friday. These hours are equivalent to 72 slots of 5 minutes each day for a total of 360 slots to be covered throughout the week. Therefore, it is assumed that the doctor does not use extra slots (parameter $h_2 = 0$). Each appointment lasts between 20 and 30 minutes, which means that each patient requires between 4 and 6 slots per consultation. Consequently, each week, the doctor attends between 60 and 90 patients with an average of 75 if the probabilistic model with variable time proposed is adopted.

The experiment assumes that the center has a list of patients where those who have been waiting for the longest have waited 8 weeks. This list is generated at random. The simulation is performed as follows:

- (i) At the beginning of each week, a set of patients (\wp) is generated who ask for an appointment. This is done by generating a random number according to a discrete uniform distribution in $[62, 66]$. This number is used to randomly select patients from the database so that the proportion of first visits is respected. The selected patients are added to the end of the waiting list with a sojourn value of one. These patients are common to all the scheduling approaches. In the case of model with variable time slots, we assume that times follow a discrete uniform distribution in $[4, 6]$.
- (ii) As described above, the buffer of patients to be passed to the scheduler is obtained from the waiting list

TABLE 2: Parameter of the simulation.

Parameters	Values
q	0.3
w_1	70
w_0	50
h_1	3
h_2	0
$ I $	5
$ T $	72
p_r	0.3
t_k	$U[4, 6]$
$ \mathcal{P} $	$U[62, 66]$

- (iii) The probabilistic model with variable time (1) assigns the day and time of the appointment to different patients in the buffer. Those patients not assigned return to the waiting list
- (iv) Attendance is simulated for each patient, based on their estimated show-up probabilities. If a patient attends the appointment, the center obtains the corresponding profit (70 € per new patient, 50 € per follow-up), and the patient is removed from the waiting list. Otherwise, the patient can request an appointment again with a probability of 0.3, denoted by p_r , or leave the waiting list definitively with a probability of 0.7. If the patient reschedules, they are returned to the waiting list with a sojourn value of zero. At the end of the week, the sojourn value of all patients is increased by one

Parameter values are summarized in Table 2. It is important to point out that these values have been provided by the health center basis for this study. In this way, simulation experiments reproduce the expected performance of the center if the proposed system was implemented during an entire year.

5.3. Results. We now present the results obtained in the experiment. The performance of the proposed model is compared with the following: (i) the system currently implemented in the health center which assigns the patient to the first available slot with a fixed duration of 30 minutes (FIFO constant); (ii) the system which would assign each patient to the first available slot but would use the estimate of the number of slots they would need (FIFO variable); and (iii) the model proposed by Ruiz-Hernandez et al. [11] that assigns patients based on their probabilities using patient constant appointment time (Time constant). Our model will be referred to as time variable. All models are coded with CPLEX Solver for mathematical optimization in MATLAB R2020a, and the experiments have been conducted in a PC with an Intel Core I9 (2.6-4.5) GHz and 32 GB RAM processor.

Table 3 shows the average results of the simulation over the 48 weeks, including computing times of the different methods. As can be seen, the proposed model achieves better results in terms of the number of patients in the queue and indirect waiting times. Similarly, the center's profit and the doctor's idle time are increased and reduced, respectively, with this approach. These results are achieved while keeping no-show rates at acceptable levels, just behind those of the time constant scheduling model. With respect to the computing time, our model exceeds the time required for existing methods. Nonetheless, the total computing time for allocating all the patients in a week (~32 seconds) is perfectly assumable for an offline scheduling system.

For a more graphic assessment of the results, the cumulated performance measures over the course of each week are presented. Figure 2(a) shows the number of people on the waiting list for each week. In constant time models, since no more than 60 appointments can be assigned per week, the number of people on the waiting list increases over time. On the other hand, if we compare the variable time models, we can see that the probabilistic model presents better results, by making an assignment that minimizes the effects of the no-shows. It is important to note that the proposed model does not manage to keep the waiting list steady. This is due to the fact that the number of patients is too small in relation to the number that can be assigned. This could be solved either by increasing the number of patients arriving each week or by reducing the center's operating times.

Figure 2(b) shows the sojourn value, i.e., the average number of weeks a patient has to wait before being scheduled for an appointment. As in the previous graph, in constant time systems, the average waiting time for patients increases over the weeks. This is a consequence of not having enough capacity to assign all patients. In contrast, the variable FIFO model remains stable over time but fails to reduce the sojourn value throughout the simulation. Finally, the probabilistic model with variable time not only decreases the time but takes it to the minimum values.

Figure 2(c) shows the cumulated profits. Probabilistic models have higher benefits than the rest. Of these, the model with variable time obtains better revenues, since under this approach, a greater number of patients can be seen and their attendance probability is maximized.

Figure 2(d) shows the cumulated doctor's inactivity time. It should be noted that the constraint added in the probabilistic model ensures that no slots can be left empty throughout the day unless subsequent appointments are not scheduled, see (7). The same applies to the FIFO model. This means that the inactive time of the doctors is highly associated with the number of no-shows. The other factor with direct impact in the doctor inactive time is the difference between the deterministic service time and the scheduled time, which directly affects constant-time models. As in the previous graph, variable time systems offer better results, as they take into account the real patient service times. Among them, the proposed model has a lower cumulative doctor inactive time.

The same pattern can be seen in Figure 2(e), which shows the number of cumulated no-shows. The models with variable time present fewer no-shows, and within them, the model with constant time presents less cumulated no-shows.

TABLE 3: Average weekly simulation results.

	FIFO constant	FIFO variable	Time constant	Time variable
Patient in the queue	688.449	532.918	672.979	356.143
Average sojourn value (weeks)	9.951	7.787	9.824	5.089
Center's profit	2703.673	3199.184	3434.286	4130.612
Doctor inactive time (minutes)	484.695	210.920	417.145	154.695
No-show rate	0.124	0.120	0.081	0.085
Computing time (seconds)	0	0.002	17.554	32.332

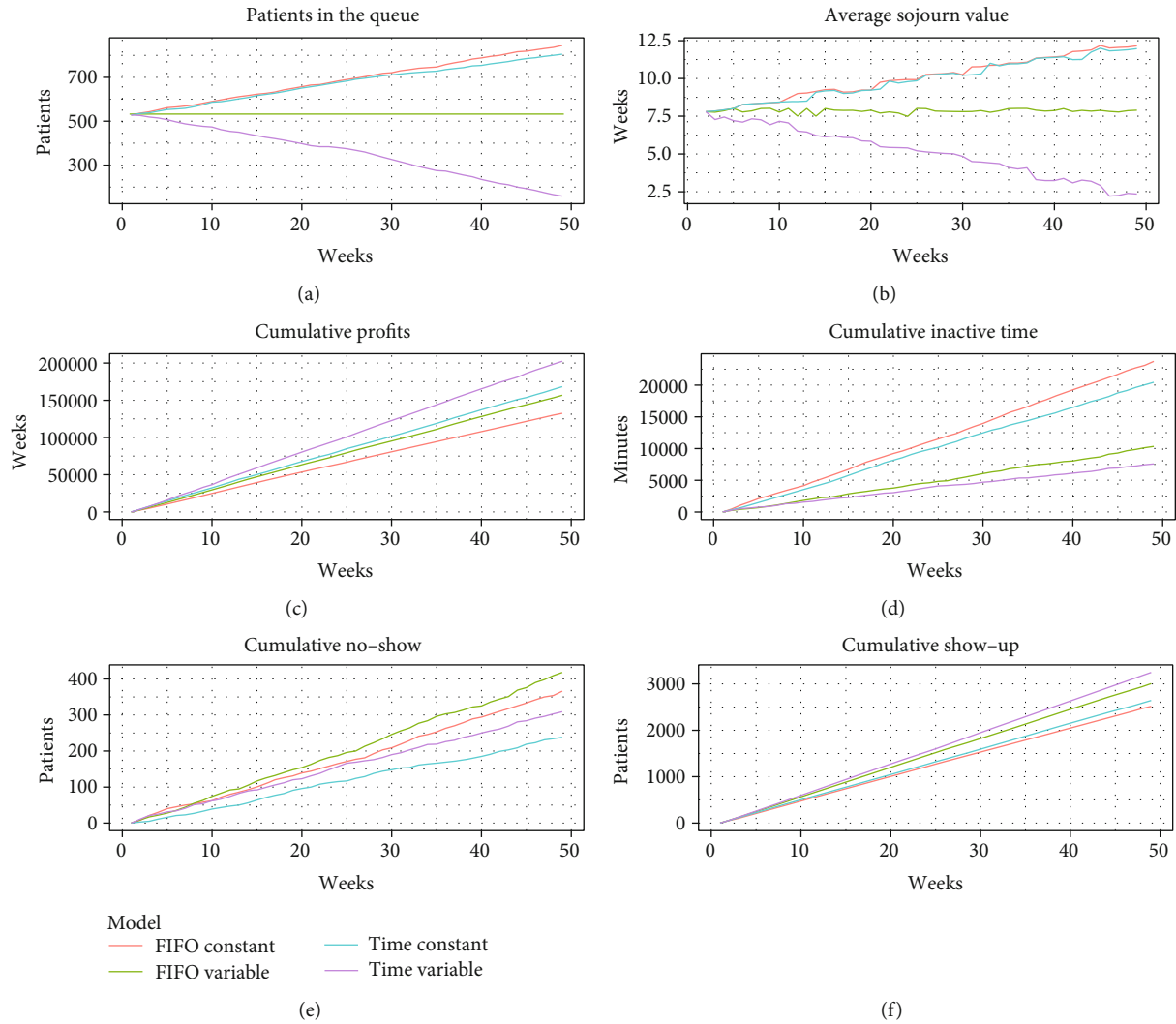


FIGURE 2: Weekly simulation results.

Finally, Figure 2(f) presents the complementary of the previous graphic, that is, the number of assigned patients who showed up. Again, models with variable time present better indicators over time since they can allocate a larger number of patients.

6. Conclusions

In this article, we have addressed the problem of no-shows in health centers. This problem causes significant damage to the

centers, ranging from increased waiting times for patients to severe financial losses. To solve this problem, we have proposed a scheduling system based on a probabilistic scheduling with variable time model together with dynamic priority allocation scheme. The system is based on the solution of a mixed-integer programming model that maximizes the expected profits of the clinic, differentiating between first and follow-up visits. The model minimizes the impact of no-shows on the expected revenues based on the patient's show probabilities and their appointment time.

The model is based on individual estimates of patients' show appointment probabilities. These probabilities have been estimated by using a logistic regression model with L1 regularization (lasso), because of its ability to select variables automatically. In addition, the model can handle different patient appointment times. These values have been simulated based on information provided by the health center from which the data were extracted.

The experiments show that while both the waiting list and the waiting times are increased in the models with constant time, the proposed model is able to reduce the waiting list by 30% and the waiting times by 50% with respect to their values at the beginning of the simulation. The proposed model is also capable of increasing the cumulated earnings by more than 50%, while reducing the cumulated doctor's idle time by more than 40%, with respect to the current system used at the health center.

There are several opportunities for future research. The first is to extend the probabilistic model with variable time developed to allow overbooking. Similarly, the model could be extended to more environmental factors affecting scheduling such as walk-in, early cancelations, and patient preferences. Finally, appointment times could be estimated, just as the attendance probabilities, in order to obtain more realistic results.

To conclude, the proposed model is capable of working in a way that minimizes the probability of a patient missing an appointment, while allowing for more patients to be seen. It has proven to dramatically outperform models with constant time, as well as the variable time extension of the current hospital system.

Data Availability

Answer: No. Comment: The data used to support the findings of this study were supplied by the Jimenez Díaz Hospital under license and so cannot be made freely available. Requests for access to these data should be made to Enrique Baca-García.

Conflicts of Interest

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

Acknowledgments

The authors would like to thank David Garcia-Heredia for his helpful suggestions to the model. This work was partly funded by Carlos III (ISCIII PI16/01852), American Foundation for Suicide Prevention (LSRG-1-005-16), the Madrid Regional Government (B2017/BMD-3740 AGES-CM 2CM; Y2018/TCS-4705 PRACTICO-CM), MINECO/FEDER ('ADVENTURE', id. TEC2015-69868-C2-1-R), MCIU Explora Grant 'aMBITION' (id. TEC2017-92552-EXP), and MICINN ('CLARA', id. RTI2018-099655-B-I00). We would like to thank Angel Blanco, Pablo Jose Ruiz Arcones, Alberto Pardo, Marta Moratilla, Alberto Cid Carrillo, Antonio Herrero Gonzalez, and Juan Jose Serrano Garcia-Ortega

who assisted with the preparation of data. The author is also grateful for the collaboration of the board of directors Juan Antonio Alvaro de la Parra, Raquel Barba Martin, Ana Leal Orozco, Marta Sanchez Menam, and Adolfo Bermudez de Castro.

References

- [1] L. F. Dantas, J. L. Fleck, F. L. Cyrino Oliveira, and S. Hamacher, "No-shows in appointment scheduling—a systematic literature review," *Health Policy*, vol. 122, no. 4, pp. 412–421, 2018.
- [2] J. Gier, "Missed appointments cost the US healthcare system \$150 B each year," *Health Management Technology*, vol. 2, 2017.
- [3] V. Chariatte, A. Berchtold, C. Akre, P.-A. Michaud, and J.-C. Suris, "Missed appointments in an outpatient clinic for adolescents, an approach to predict the risk of missing," *Journal of Adolescent Health*, vol. 43, no. 1, pp. 38–45, 2008.
- [4] D. Carreras-García, D. Delgado-Gómez, F. Llorente-Fernández, and A. Arribas-Gil, "Patient No-Show Prediction: A Systematic Literature Review," *Entropy*, vol. 22, no. 6, p. 675, 2020.
- [5] J. Daggy, M. Lawley, D. Willis et al., "Using no-show modeling to improve clinic performance," *Health Informatics Journal*, vol. 16, no. 4, pp. 246–259, 2011.
- [6] P. E. Hasvold and R. Wootton, "Use of telephone and SMS reminders to improve attendance at hospital appointments: a systematic review," *Journal of Telemedicine and Telecare*, vol. 17, no. 7, pp. 358–364, 2011.
- [7] T. Cayirli and E. Veral, "Outpatient scheduling in health care: a review of literature," *Production and Operations Management*, vol. 12, no. 4, pp. 519–549, 2003.
- [8] D. Gupta and B. Denton, "Appointment scheduling in health care: challenges and opportunities," *IIE Transactions*, vol. 40, no. 9, pp. 800–819, 2008.
- [9] A. Ahmadi-Javid, Z. Jalali, and K. J. Klassen, "Outpatient appointment systems in healthcare: a review of optimization studies," *European Journal of Operational Research*, vol. 258, no. 1, pp. 3–34, 2017.
- [10] C. Zacharias and M. Pinedo, "Appointment scheduling with no-shows and overbooking," *Production and Operations Management*, vol. 23, no. 5, pp. 788–801, 2014.
- [11] D. Ruiz-Hernández, D. García-Heredia, D. Delgado-Gómez, and E. Baca-García, "A probabilistic patient scheduling model for reducing the number of no-shows," *Journal of the Operational Research Society*, vol. 71, no. 7, pp. 1102–1112, 2020.
- [12] T. Cayirli, E. Veral, and H. Rosen, "Designing appointment scheduling systems for ambulatory care services," *Health Care Management Science*, vol. 9, no. 1, pp. 47–58, 2006.
- [13] Y. Huang and S. Verduzco, "Appointment template redesign in a women's health clinic using clinical constraints to improve service quality and efficiency," *Applied clinical informatics*, vol. 2, no. 6, p. 271, 2015.
- [14] D. Bentayeb, N. Lahrichi, and L.-M. Rousseau, "Patient scheduling based on a service-time prediction model: a data-driven study for a radiotherapy center," *Health Care Management Science*, vol. 22, no. 4, pp. 768–782, 2019.
- [15] A. Ratcliffe, W. Gilland, and A. Maruchek, "Revenue management for outpatient appointments: joint capacity control and

- overbooking with class-dependent no-shows,” *Flexible Services and Manufacturing Journal*, vol. 24, no. 4, pp. 516–548, 2012.
- [16] K. Muthuraman and M. Lawley, “A stochastic overbooking model for outpatient clinical scheduling with no-shows,” *IIE Transactions*, vol. 40, no. 9, pp. 820–837, 2008.
- [17] C. Yan, J. Tang, B. Jiang, and R. Y. K. Fung, “Sequential appointment scheduling considering patient choice and service fairness,” *International Journal of Production Research*, vol. 53, no. 24, pp. 7376–7395, 2015.
- [18] M. Samorani and S. Harris, *The Impact of Probabilistic Classifiers on Appointment Scheduling with No-Shows*, Fortieth International Conference on Information Systems, Munich, 2019.
- [19] M. Savelsbergh and K. Smilowitz, “Stratified patient appointment scheduling for mobile community-based chronic disease management programs,” *IIE Transactions on Healthcare Systems Engineering*, vol. 6, no. 2, pp. 65–78, 2016.
- [20] X. Ding, Z. F. Gellad, C. Mather III et al., “Designing risk prediction models for ambulatory no-shows across different specialties and clinics,” *Journal of the American Medical Informatics Association*, vol. 25, no. 8, pp. 924–930, 2018.

Research Article

Utility-Based Multicriteria Model for Screening Patients under the COVID-19 Pandemic

Lucia Reis Peixoto Roselli , **Eduarda Asfora Frej**, **Rodrigo José Pires Ferreira**,
Alexandre Ramalho Alberti, and **Adiel Teixeira de Almeida**

Universidade Federal de Pernambuco, Av. Acadêmico Hélio Ramos, s/n-Cidade Universitária, Recife, PE CEP 50.740-530, Brazil

Correspondence should be addressed to Lucia Reis Peixoto Roselli; lrpr@cidsid.org.br

Received 18 June 2020; Revised 31 July 2020; Accepted 19 August 2020; Published 1 September 2020

Academic Editor: Mirian C. D. Pinheiro

Copyright © 2020 Lucia Reis Peixoto Roselli 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.

In this paper, a utility-based multicriteria model is proposed to support the physicians to deal with an important medical decision—the screening decision problem—given the squeeze put on resources due to the COVID-19 pandemic. Since the COVID-19 emerged, the number of patients with an acute respiratory failure has increased in the health units. This chaotic situation has led to a deficiency in health resources. Thus, this study, using the concepts of the multiattribute utility theory (MAUT), puts forward a mathematical model to aid physicians in the screening decision problem. The model is used to generate which of the three alternatives is the best one for where patients with suspected COVID-19 should be treated, namely, an intensive care unit (ICU), a hospital ward, or at home in isolation. Also, a decision information system, called SIDTriagem, is constructed and illustrated to operate the mathematical model proposed.

1. Introduction

At the start of the current year (2020), the COVID-19 disease, caused by the new coronavirus (the SARS-CoV-2), was deemed to be an epidemic. However, a few weeks later, it was reclassified as a pandemic. Since the COVID-19 has emerged, it has been present in different degrees of illness in the human organism, causing, in severe cases, acute respiratory failure [1–3].

In this context, this disease has resulted in an increasing number of patients requiring hospital treatments, especially in intensive care units (ICUs) with the support of mechanical ventilation equipment. Consequently, in many cities around the world, this increase in the demand for places in hospitals has placed a great strain on medical resources and revealed deficiencies in all forms of provision for a pandemic and therefore adequate treatment has not been available for all the severe cases.

Thus, to deal with this chaotic situation, it is fundamental to have decision-making strategies in place as these are important to ensure that the expectation of the survival of

COVID-19 patients can be maximized. In other words, in such situations, which involves a risk context, it is important to conduct a rational decision-making process in order to be able to save the majority of patients [4].

Therefore, in order to allow the rational conduct of an important medical decision-making problem—the screening decision problem—a utility-based multicriteria model is proposed in this study.

This additive multicriteria decision-making model is based on the multiattribute utility theory (MAUT) approach [5], which considers the concepts of the utility theory [6], to deal with the uncertainty presented in medical diagnostics [7]. Thus, in this study, issues from operational research are considered to support a healthcare decision-making problem.

Also, a decision information system (DIS), called SIDTriagem, is constructed to implement this multicriteria decision-making model. The outcome of this DIS is a recommendation about where a patient with suspected COVID-19 should be directed to, considering the alternatives presented in the screening problem. It is worth mentioning that this

model is a supplement to support physicians to deal with the screening problem. It is left to the physicians to decide whether or not to follow the recommendation made.

This paper is organized as follows. Section 2 details the screening decision problem. Section 3 describes the utility-based model proposed in this study. Section 4 presents the applicability of the proposed approach and discusses these results. Finally, Section 5 draws some conclusions and indicates possible topics for future lines of research.

2. The Screening Decision Problem

The COVID-19 pandemic has provoked several decision situations that physicians are facing in their routines in hospitals and other health units all over the world. One of the main decision problems that they encounter every day is how best to screen patients with suspected COVID-19. A patient arrives at a health unit with symptoms and other complaints. Then, the physician has to decide, depending on the patient's clinical state, whether this patient should be sent for treatment in an intensive care unit (ICU), in a hospital ward (but not in an ICU) or whether he/she should be sent home to isolate instead of being hospitalized. This is the classical screening problem that is addressed in this paper, in which the context of the COVID-19 pandemic has been made more acute. Given that uncertainty is an inevitable factor that is inherently present in medical diagnostics and treatment decisions [6], a decision analysis (DA) model based on the multiattribute utility theory (MAUT) is developed in order to aid physicians when they make such decisions.

DA support for screening patients within the context of other diseases has been widely explored in the literature. Xu et al. [8] used a DA approach based on decision trees to investigate strategies for the triage of patients with symptoms of acute stroke. Outcomes were measured based on workflow times. Probabilities and input parameters were estimated based on guidelines and previously published studies. A practical analysis was conducted using TreeAge Pro software. Jiang et al. [9] designed a DA tree in order to evaluate the best strategy for treating patients after an esophagectomy. The TreeAge Pro software was used to construct the decision tree. Two strategies were compared based on several factors, such as length of stay in the hospital, costs, and possible complications. A sensitivity analysis was performed by using a Monte-Carlo simulation. Felder and Mayrhofer [10] analyze the impact of risk preferences in decisions about medical screening, testing, and treatment. They conclude that a risk averse decision-maker tests and treats patients at lower probabilities of illness, compared to risk neutral and risk vulnerable decision-makers. Cleary et al. (2005) [11] applied DA techniques for comparing three different strategies of screening for herpes, a simple virus, in pregnant women; probability estimations were derived from DA on the literature. Kiberd and Forward [12] developed a DA-based study to investigate the impact of medical screening decisions for West Nile virus in organ transplantation, by considering lives lost and saved.

A cost-effective analysis approach was also widely used by authors when dealing with DA models for screening

patients, and these studies covered a wide range of diseases. Wilson and Howe [13] developed a DA model for screening methods of dysphagia after stroke. Different strategies were compared based on a cost-effective analysis. Medical costs were measured from a societal perspective, and effectiveness was measured in years of quality-adjusted life. Sensitivity analysis using a Monte-Carlo simulation was performed. Donnan et al. [14] conducted a cost-effectiveness analysis for DA of children with acute lymphoblastic leukemia. Probabilities were obtained based on published evidence in the literature, and survival was measured in months of life. Cooper et al. [15] constructed a decision analysis model to handle health outcome states and costs of screening strategies for children in preoperative coagulation tests prior to a tonsillectomy and/or adenoidectomy. Probabilities, costs, and utility data were estimated based on a review of databases. Sensitivity analysis was performed so that parameters were widely varied. Baeten et al. [16] conducted a study to show the potential and impact of three approaches in the use of cost-effective analysis in the scope of breast cancer control: targeting specific groups, by comparing disparities; equity weighting, by valuing high and low health gains differently; and multicriteria decision analysis, giving weights for multiple equity and efficient criteria. Oh et al. [17] used DA based on a cost-effective approach to compare different strategies for screening rheumatoid arthritis and systemic lupus erythematosus patients. Data were obtained from previous studies and from real practical cases. Rulyak et al. [18] applied DA for screening strategies in familial pancreatic cancer kindreds. Life expectancy and lifetime medical care costs were modeled in order to conduct a cost-effective analysis. McGrath et al. (2002) [19] used DA software (TreeAge Pro) for comparing four strategies for screening patients with colorectal cancer, taking into account the cost to find an advanced adenoma. Probabilities, test characteristics, and costs were estimated based on a literature review and local costs.

In this context, this paper is aimed at presenting a multicriteria model for screening patients with suspected COVID-19, based on a DA approach within the multiattribute utility theory. Two main factors are taken into account: the life of the patient being screened and the cost of the alternative indicated for that patient. These criteria are further detailed in this paper. Subjective probabilities are considered for the construction of a decision tree for the screening problem. The next section details the whole structure of the mathematical model proposed for the screening problem.

3. Utility-Based Model for Aiding the Screening of Suspected COVID-19 Patients

3.1. Decision Tree for the Screening Problem. In this section, the decision tree technique is used to illustrate the screening decision problem investigated in this study [9, 10]. According to Cheng et al. [20], this technique can be used to identify the risk factors presented in a decision-making problem, it being possible to consider the outcomes obtained by their combination.

In this context, in the decision tree constructed, the alternatives indicated to conduct the health treatment for the patients with suspected COVID-19 are identified. In the screening problem investigated in this study, three alternatives are considered; these alternatives are ICU stay (ICU), hospital stay (HS), and isolation at home (IH).

Also, for each one of these alternatives, the uncertainty is presented, since according to [6], uncertainty is an inevitable factor that is inherently present in treatment decisions. Thus, in an uncertainty context, the consequences to be obtained depend on the alternatives and the state of nature [6, 21].

In other words, for each alternative and each state of nature, which is represented by the chances to survive and chances to death, a consequence is obtained. For the alternative ICU stay, the patient can survive the ICU stay or die during the ICU stay. For the alternative hospital stay, the patient can survive the hospital stay or die during the hospital stay. Finally, for the alternative isolation at home, the patient can survive isolation at home or die during isolation at home.

It is worth mentioning that in the decision tree technique the squares are the decision nodes, the circles are the chance nodes, and the arrows are used to connect these decision elements [8]. In this context, in the decision tree constructed, four squares and three circles are presented. The decision tree constructed is illustrated in Figure 1.

Based on the decision tree illustrated in Figure 1, the mathematical model used to construct the utility-based multicriteria model is described in the next section. This mathematical model connected the decision elements presented in Figure 1 to obtain the recommendations (outputs) for the screening problem investigated.

3.2. Mathematical Model. In this section, the mathematical model, presented in the utility-based multicriteria model, is described. This mathematical model is based on the multiattribute utility theory (MAUT) [4] and takes into account the concepts of the utility theory [6] and multicriteria approach [5, 22, 23].

The utility theory [6] is a very appropriate way to deal with decision-making under uncertainty. In this context, states of the world (or states of nature) are used to represent the uncertainty presented in the decision scenario. Also, for each state of nature, probabilities are assigned to represent their chance of occurring, and these are obtained by an expert or by the decision-maker (from the subjective expected utility model) [21].

In this context, regarding the screening decision-making process considered in this study, the states of nature are survival or death, and the alternatives are the options to conduct the healthcare treatment with the patients with suspected COVID-19, with three alternatives being considered: ICU stay (ICU), hospital stay (HS) and isolation at home (IH).

Also, two criteria are considered in this complex decision situation, the life of the patient being screened and the cost of the alternative indicated for that patient. The cost of an alternative is subjectively related to the impact on the health system, considering resource constraints. Therefore, alternative “home isolation,” for example, presents no cost for the

health system, since the patient will stay at home and no health resources will be occupied by this patient. For alternative “ICU stay,” however, the cost for the health system might be high, especially when ICU occupation is high and resources are scarce. The alternatives are evaluated in each one of these criteria, considering the multicriteria decision scenario [5, 22, 23].

Thus, for this decision-making problem, the decision-maker’s preferences are assumed to be represented by MAUT [5]. In this context, from the corroboration of the additive independence condition, the additive aggregation analytic form is used to construct the mathematical model. The multiattribute utility function is presented in equation (1), where a is the alternative, k_j is the scaling constant for criterion j , and $u_j(a)$ is the marginal utility function in criterion j :

$$u(a) = \sum_{j=1}^n k_j u_j(a). \quad (1)$$

It is worth mentioning that the scaling constants are obtained by applying an elicitation procedure with a decision-maker. The values of the scaling constants, for both criteria, are equal to 0.5 and their sum is equal to 1, in accordance with MAUT concepts [5]. The values of the scaling constants are presented in equation (2):

$$k_L = k_C = 0.5. \quad (2)$$

Also, the utility functions represent the consequences in each state of nature. For this study, the marginal utility functions are also obtained in the elicitation procedure. In this context, for the criterion patient’s life, the utility functions are equal for the three alternatives, namely, 1 if the state of nature is survival and 0 if the state of nature is death. Equations (3) and (4) illustrate this condition:

$$u(\text{Shi}) = u(\text{Shs}) = u(\text{Sicu}) = 1, \quad (3)$$

$$u(\text{Dhi}) = u(\text{Dhs}) = u(\text{Dicu}) = 0. \quad (4)$$

On the other hand, regarding the criterion cost, the utility function is the same, since it does not depend on the state of nature. In this situation, the resources were consumed in the hope of saving the patient, regardless of whether the patient survives or dies.

In addition, for the criterion cost, the utility function for the alternative ICU stay presented the worst value, since the health treatment in ICU is more expensive. Analogously, for the alternative isolation at home, the utility function presented the highest value, it being the most desirable alternative [24–26].

An important consideration for the ICU stay utility function is the dependence regarding another variable, which is associated to the probability of a “future patient” arriving in the healthcare system, in a severe condition, and requiring to be sent to the ICU, combined with the occupation rate of the ICU. This variable is called F_p , this being an acronym of the probability of a future patient arriving in the healthcare

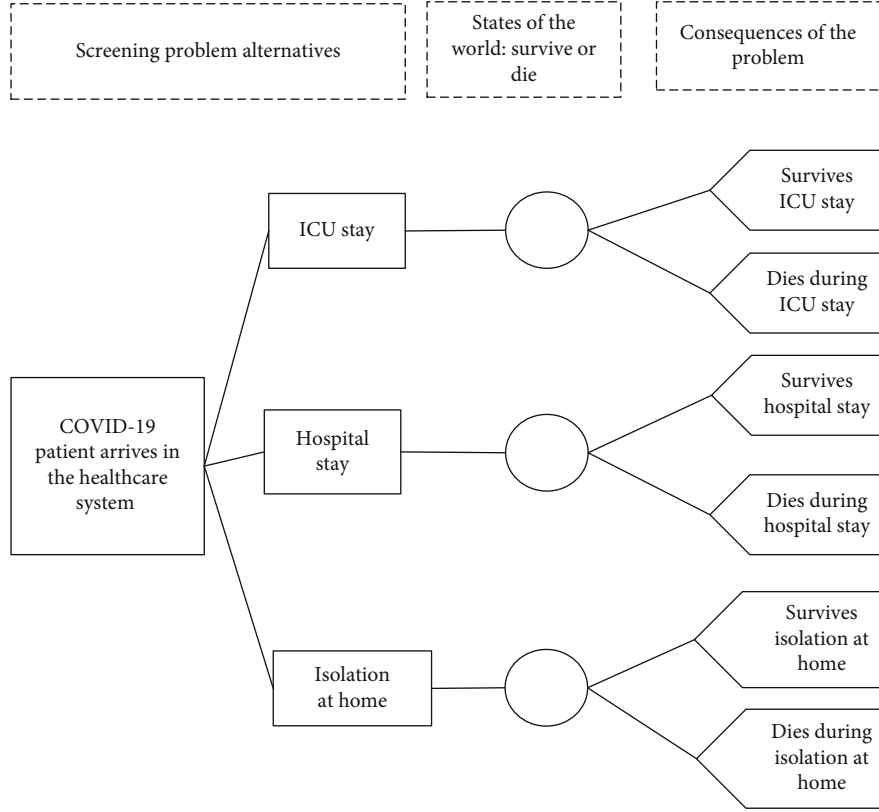


FIGURE 1: Screening problem decision tree.

system. If F_p is equal to 1, this indicates that no beds are available in the ICU. In this context, the utility function for this situation presents the worst value, this being a chaotic situation. In this scenario, it is difficult to accommodate patients in the ICU, with a tendency to recommend the alternatives hospital stay or isolation at home. As to these considerations, the utility function for an ICU stay, for the evaluation in the criterion cost, is defined according to equation (5):

$$u(\text{Cicu}) = 1 - F_p. \quad (5)$$

For the worst case, when the ICU is completely occupied, F_p is equal to 1 (which means that sending the patient to the ICU would lead to a very high cost, thus leading to an utility equal to 0). In the opposite extreme case, F_p would be equal to 0. However, a parameterization for three possible cases between these two extreme situations is considered for the decision information system: high occupation ($F_p = 0.7$), intermediate occupation ($F_p = 0.5$), and low occupation ($F_p = 0.3$). In order to define these values of 0.7, 0.5, and 0.3, three doctors were consulted and simulations were performed in order to verify which values fit best according to doctors' actual attitude. It is worth mentioning, however, that these values depend on the decision-maker judgments, and in the SIDTriagem, these ranges can be adjusted if the user so desires.

The utility function for isolation at home, in the criterion cost, i.e., $(U(\text{Chi}))$, is equal to 1. Also, the utility function for hospital stay $(U(\text{Chs}))$ is between 0 and 1. In this study, the

utility function equal to 0.8 is considered. However, a variation can be applied using the Monte-Carlo simulation, as presented in the next section.

Finally, another important variable to be considered in this mathematical model is the subjective probability assigned for each state of nature in order represent its chance of occurring. Thus, these probabilities are given by the physicians, considering their subjective evaluation about the patient's state of health.

In other words, the physician has to define a probability of surviving (chance of surviving) for the patient considering each one of the alternatives. These probabilities are represented by $\pi(\text{Shi})$, $\pi(\text{Shs})$, and $\pi(\text{Sicu})$; their sum being equal to 1. Also, the chance to dying is the complementary probability of the chance to survive.

Therefore, based on these considerations, the utility-based multicriteria model constructed for the screening decision problem is described by equations (6)–(8).

$$u(\text{ih}) = \pi(\text{Shi})[k_L u(\text{Shi}) + k_c u(\text{Chi})] + (1 - \pi(\text{Shi}))[k_L u(\text{Dhi}) + k_c U(\text{Chi})], \quad (6)$$

$$u(\text{hs}) = \pi(\text{Shs})[k_L u(\text{Shs}) + k_c u(\text{Chs})] + (1 - \pi(\text{Shs}))[k_L u(\text{Dhs}) + k_c U(\text{Chs})], \quad (7)$$

$$u(\text{icu}) = \pi(\text{Sicu})[k_L u(\text{Sicu}) + k_c u(\text{Cicu})] + (1 - \pi(\text{Sicu}))[k_L u(\text{Dicu}) + k_c U(\text{Cicu})]. \quad (8)$$

As to equations (6)–(8), the physicians receive a recommendation about which alternative is the best one for accommodating the COVID-19 patient. This is the one that presents the highest multiattribute utility function.

In the next section, a practical application of this utility-based multicriteria model is presented in order to illustrate how this mathematical model is used to support the decision-making problem about screening. Also, the decision information system, called SIDTriagem, is presented.

4. Practical Applicability and Results

To apply the proposed model for aiding the screening of patients with suspected COVID-19, the physician should first input information about the patient’s chances of survival in three scenarios: isolation at home, a hospital stay, or in an intensive care unit (ICU). This information should be given based on the patient’s symptoms and clinical state. These chances of survival, however, are not precisely established and involve subjective factors that may not be quantified. Therefore, during the development of the system, three physicians were consulted by an analyst in order to find out what would make them feel more comfortable about providing such information. As a result of this consultation process, it was verified that the physicians preferred to give information about chances of survival on a verbal scale, instead of providing numbers. Thus, a 5-point Likert scale was developed for establishing such probabilities: very low, low, medium, high, and very high. Each of these levels is associated to a probability range of 20% width, which was also calibrated with the physicians. The reason of using probability ranges instead of exact values of probabilities is related to the inherent imprecision and subjectivity of this information. Table 1 illustrates the association of each level of the scale with the probability ranges.

According to those probability ranges, a Monte-Carlo simulation is performed in order to obtain a recommendation of conduct for the physician. At each simulation instance, a random number between the lower and upper limits of Table 1 is generated for each probability of survival, according to a uniform distribution and taking into account the levels of the verbal scale provided by the user. The recommendation given by the model is based on a robustness index that is computed for each alternative. The robustness index of an alternative is related to the percentage of simulation instances in which the expected overall utility of that alternative is greater than the expected utility of the other alternatives, in accordance with Equations (5)–(7). The model therefore recommends that the user follow the alternative with the largest robustness index.

The proposed approach for aiding the screening of patients with suspected COVID-19 is operated by means of a DIS, called SIDTriagem, which is available for users at <http://insid.org.br/sidtriagem/app/>. Physicians log on to the system and then he/she enters the patient’s name, age, and gender (optional data). Figure 2 shows the interface of the system, with a practical hypothetical example.

In Figure 2, a 67 year-old woman is considered to have been evaluated by the user of the system (a physician) at a

TABLE 1: Ranges of probabilities of survival.

Verbal scale	Lower limit	Upper limit
Very low	0	20%
Low	20%	40%
Medium	40%	60%
High	60%	80%
Very high	80%	100%

healthcare unit. By examining this patient and analyzing all her symptoms and her clinical state, the physician enters information about the chances of survival of this woman in three scenarios: in an ICU, in a hospital ward, and during isolation at home. Let us assume that the physician evaluates her as either *very high* or *medium* or *very low*, respectively. Then, the physician should estimate the ICU occupancy rate at that time, also on a verbal scale: low, intermediate, and high. As previously explained in Section 3.2, this ICU occupancy rate is for calibration of the F_p parameter, which influences the utility of the cost of sending the patient to the ICU. Let us consider that there is intermediate occupancy rate at that moment. Finally, the physician may optimally state how confident he/she is about the information provided: very unconfident, unconfident, neutral, confident, very confident, or even N/A. This information is not used by the mathematical model, but it may be further used in future studies to evaluate the behavior of physicians in such situations.

After entering all the input data, the user clicks on the “Calculate” button and the recommendations obtained based on the simulations are shown to the user. In this case, the recommendation of the system is to send the patient to the ICU, with 87% of robustness. This means that, in 87% of the simulations instances performed, this alternative had the highest expected utility, compared to hospital stay and isolation at home. Hospital stay had a robustness index of 13%, which means that in 13% of the simulation instances this alternative had the greatest expected utility value. The robustness index for isolation at home was 0, indicating that this alternative never wins against the others in terms of expected utility.

The system also provides user with an alternative way of visualizing the results. By clicking on the “Switch to graphical visualization” button at the bottom of Figure 2, a bar graphic appears as an alternative possible visualization, as Figure 3 shows. These two ways of visualization were included in the system due to the feedback of the physicians; some of whom prefer to visualize them in a table with numbers, and others prefer graphics. Therefore, the system provides both numbers and graphics.

Finally, the user may choose to state whether or not he/she intends to follow the recommendation. Also, there is a space for recording the feedback of the users of the system, by clicking on the “Conclude” button. This feedback helps to make further improvements to the model and to the system itself.

It should be highlighted here that this system should be used as a support tool for aiding screening decisions, based on a structured mathematical model. There are no normative

System Information and Decision for triage of patients with suspected COVID-19 in healthcare units with resource constraints

SCREENING

Patient: Age: Gender: Male Female

Chance of survival:

ICU: * Hospital stay: * Home isolation: *

ICU occupancy rate: *

How confident are you about the information provided?

Recommendation	Robustness index
ICU	87%
Hospital stay	13%
Home isolation	0%

Help us to improve the model with your feedback.

Do you intend to follow this recommendation?

Yes No

inct INSTITUTO NACIONAL DE CIÊNCIA E TECNOLOGIA INSID INSTITUTO NACIONAL DE SISTEMAS DE INFORMAÇÃO E DECISÃO LIKA Laboratório de Imunogenética Raíza Assis

FIGURE 2: Interface of the SIDTriagem system.

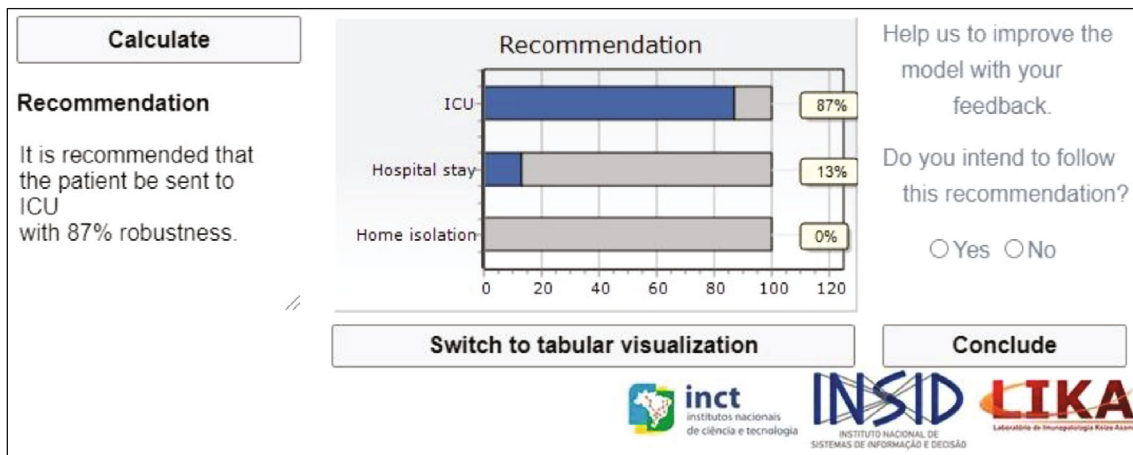


FIGURE 3: Graphical visualization of the results.

purposes, however, with the use of this system. A recommendation is given, but the final decision always rests with the physician, who should take into account all subjective factors involved in each specific situation.

5. Conclusions

In this paper, a utility-based multicriteria model for aiding screening decision situations of patients with suspected COVID-19 was proposed. The screening problem is critical due to the scarcity of treatment resources in hospitals, such as ICU beds, for instance. Therefore, a structured mathematical modeling of this problem is important for aiding physicians to decide if a suspected COVID-19 patient should go to an ICU, a hospital ward, or stay at home in isolation.

The mathematical model was built based on the decision analysis concepts and multiattribute utility theory (MAUT), considering the inherent stochastic nature of this decision-

making problem. Considering the inherent imprecision associated to estimating the patient's chances of survival, the proposed model works with probability ranges that serve as an input for a Monte-Carlo simulation model. Moreover, considering the difficulty that physicians have in providing this information due to the subjectivity of the factors involved, a verbal scale is used for estimating patients' chances of survival.

The proposed approach is operated by means of a decision information system, which has a user-friendly interface and can be easily used by physicians in healthcare units worldwide. The information obtained from the occurrences registered in the system is stored in a database. Finally, as suggestions for future research, the occurrences registered in the system would be extremely useful for conducting several kinds of analyses, including a comparative analysis of what the model proposes and what doctors actually do, in practice. Also, behavioral studies based on the data

obtained from the physicians' records may be useful for improving the design of decision information systems and their functionalities.

Data Availability

The database generated from the occurrences registered in the SIDTriagem used to support the findings of this study are restricted by the Ethical Committee in Research of the Federal University of Pernambuco with CAAE ("Certificado de Apresentação e Apreciação Ética" - Certificate of Presentation and Ethical Appreciation) number 31065820.5.0000.5208 in order to protect patient privacy. Data are available from the corresponding author upon request according with the criteria for access to confidential data.

Conflicts of Interest

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

Acknowledgments

The authors would like to acknowledge the Academic Qualifications of Higher Education Personnel–Brazil (CAPES), the Brazilian Research Council (CNPq) (308531/2015-9), and the Foundation of Support in Science and Technology of the State of Pernambuco (FACEPE) (APQ-0484-3.08/17), for the partial financial support for this research.



References

- [1] Z. Car, S. Baressi Šegota, N. Anđelić, I. Lorencin, and V. Mrzljak, "Modeling the spread of COVID-19 infection using a multilayer perceptron," *Computational and Mathematical Methods in Medicine*, vol. 2020, 10 pages, 2020.
- [2] W. J. Guan, Z. Y. Ni, Y. Hu et al., "Clinical characteristics of coronavirus disease 2019 in China," *The New England Journal of Medicine*, vol. 382, no. 18, pp. 1708–1720, 2020.
- [3] F. Zhou, T. Yu, R. du et al., "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study," *The Lancet*, vol. 395, no. 10229, pp. 1054–1062, 2020.
- [4] G. Vanagas, T. Krilavičius, and K. L. Man, "Mathematical modeling and models for optimal decision-making in health care," *Computational and Mathematical Methods in Medicine*, vol. 2019, 4 pages, 2019.
- [5] R. L. Keeney and H. Raiffa, *Decision Analysis with Multiple Conflicting Objectives*, Wiley & Sons, New York, 1976.
- [6] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton Univ. Press, Princeton, NJ, 1944.
- [7] M. Weatherall, "Information provided by diagnostic and screening tests: improving probabilities," *Postgraduate Medical Journal*, vol. 94, no. 1110, pp. 230–235, 2018.
- [8] Y. Xu, N. S. Parikh, B. Jiao, J. Z. Willey, A. K. Boehme, and M. S. V. Elkind, "Decision analysis model for prehospital triage of patients with acute stroke," *Stroke*, vol. 50, no. 4, pp. 970–977, 2019.
- [9] B. Jiang, V. P. Ho, J. Ginsberg et al., "Decision analysis supports the use of drain amylase-based enhanced recovery method after esophagectomy," *Diseases of the Esophagus*, vol. 31, no. 10, 2018.
- [10] S. Felder and T. Mayrhofer, "Risk preferences: consequences for test and treatment thresholds and optimal cutoffs," *Medical Decision Making*, vol. 34, no. 1, pp. 33–41, 2013.
- [11] K. L. Cleary, E. Paré, D. Stamilio, and G. A. Macones, "Type-specific screening for asymptomatic herpes infection in pregnancy: a decision analysis," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 112, no. 6, pp. 731–736, 2005.
- [12] B. A. Kiberd and K. Forward, "Screening for West Nile virus in organ transplantation: a medical decision analysis," *American Journal of Transplantation*, vol. 4, no. 8, pp. 1296–1301, 2004.
- [13] R. D. Wilson and E. C. Howe, "A cost-effectiveness analysis of screening methods for dysphagia after stroke," *PM & R*, vol. 4, no. 4, pp. 273–282, 2012.
- [14] J. R. Donnan, W. J. Ungar, M. Mathews, R. L. Hancock-Howard, and P. Rahman, "A cost effectiveness analysis of thiopurine methyltransferase testing for guiding 6-mercaptopurine dosing in children with acute lymphoblastic leukemia," *Pediatric Blood & Cancer*, vol. 57, no. 2, pp. 231–239, 2011.
- [15] J. D. Cooper, K. J. Smith, and A. K. Ritchey, "A cost-effectiveness analysis of coagulation testing prior to tonsillectomy and adenoidectomy in children," *Pediatric Blood & Cancer*, vol. 55, no. 6, pp. 1153–1159, 2010.
- [16] S. A. Baeten, R. M. P. M. Baltussen, C. A. Uyl-de Groot, J. Bridges, and L. W. Niessen, "Incorporating equity–efficiency interactions in cost-effectiveness analysis—three approaches applied to breast cancer control," *Value in Health*, vol. 13, no. 5, pp. 573–579, 2010.
- [17] K. T. Oh, A. H. Anis, and S. C. Bae, "Pharmacoeconomic analysis of thiopurine methyltransferase polymorphism screening by polymerase chain reaction for treatment with azathioprine in Korea," *Rheumatology*, vol. 43, no. 2, pp. 156–163, 2004.
- [18] S. J. Rulyak, M. B. Kimmey, D. L. Veenstra, and T. A. Brentnall, "Cost-effectiveness of pancreatic cancer screening in familial pancreatic cancer kindreds," *Gastrointestinal Endoscopy*, vol. 57, no. 1, pp. 23–29, 2003.
- [19] J. S. McGrath, T. P. Ponich, and J. C. Gregor, "Screening for colorectal cancer: the cost to find an advanced adenoma," *The American Journal of Gastroenterology*, vol. 97, no. 11, pp. 2902–2907, 2002.
- [20] G. Cheng, R. Cheng, Y. Pei, and L. Xu, "Probability of roadside accidents for curved sections on highways," *Mathematical Problems in Engineering*, vol. 2020, 18 pages, 2020.
- [21] W. Edwards, M. J. R. RF, and D. von Winterfeldt, *Advances in Decision Analysis: from Foundations to Applications*, Cambridge University Press, 2007.
- [22] V. Belton and T. Stewart, *Multiple Criteria Decision Analysis: An Integrated Approach*, Springer Science & Business Media, Berlin, 2002.
- [23] A. T. de Almeida, C. A. V. Cavalcante, M. H. Alencar, R. J. P. Ferreira, A. T. de Almeida-Filho, and T. V. Garcez, *Multicriteria and Multi-objective Models for Risk, Reliability and Maintenance Decision Analysis. International Series in Operations Research & Management Science. Vol 231*, Springer, New York, 2015.
- [24] A. Giannini and D. Consonni, "Physicians' perceptions and attitudes regarding inappropriate admissions and resource allocation in the intensive care setting," *British Journal of Anaesthesia*, vol. 96, no. 1, pp. 57–62, 2006.

- [25] P. Pronovost and D. C. Angus, "Economics of end-of-life care in the intensive care unit," *Critical Care Medicine*, vol. 29, pp. N46–N51, 2001.
- [26] C. P. M. Gyldmark, "A review of cost studies of intensive care units," *Critical Care Medicine*, vol. 23, no. 5, pp. 964–972, 1995.

Research Article

An Intelligent Decision-Making Support System for the Detection and Staging of Prostate Cancer in Developing Countries

Jun Zhang,^{1,2} Zhigang Chen ^{1,2}, Jia Wu ^{1,2} and Kanghuai Liu^{1,2}

¹School of Computer Science and Engineering, Central South University, Changsha 410083, China

²Central South University, Changsha 410075, China

Correspondence should be addressed to Zhigang Chen; czg@csu.edu.cn and Jia Wu; jiawu5110@163.com

Received 3 May 2020; Accepted 11 June 2020; Published 17 August 2020

Guest Editor: Plácido R. Pinheiro

Copyright © 2020 Jun Zhang 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.

Most developing countries face huge challenges in the medical field; scarce medical resources and inadequate medical personnel will affect the development and stability of the society. Therefore, for most developing countries, the development of intelligent medical systems can greatly alleviate the social contradictions arising from this problem. In this study, a new data decision-making intelligent system for prostate cancer based on perceptron neural network is proposed, which mainly makes decisions by associating some relevant disease indicators and combining them with medical images. Through data collection, analysis and integration of medical data, as well as the disease detection and decision-making process, patients are given an auxiliary diagnosis and treatment, so as to solve the problems and social contradictions faced by most developing countries. Through the study of hospitalization information of more than 8,000 prostate patients in three hospitals, about 2,156,528 data items were collected and compiled for experiment purposes. Experimental data shows that when the patient base increases from 200 to 8,000, the accuracy of the machine-assisted diagnostic system will increase from 61% to 87%, and the doctor's diagnosis rate will be reduced to 81%. From the study, it is concluded that when the patient base reaches a certain number, the diagnostic accuracy of the machine-assisted diagnosis system will exceed the doctor's expertise. Therefore, intelligent systems can help doctors and medical experts treat patients more effectively.

1. Introduction

Prostate cancer (PCA) is a human disease that occurs in malignant tumors of the prostate epithelium [1–6]. The age of its onset is mainly after 55 years of age, and the incidence rate also increases with the increase of age. In Europe and the United States, the incidence of male cancer patients ranked first, [7, 8] while the death rate ranked second. In America, the incidence of prostate cancer has risen to become the first place; the mortality rate is second only to lung cancer. Prostate cancer has become one of the most common cancers in the world and one of the most common malignancies among men in Europe and the United States [9].

In developing countries such as [10] China, the incidence rate is lower than that in many European and American countries. However, due to the large population base in developing countries, the number of cases is not to be underestimated [11].

Before the 1980s, the incidence of prostate cancer in China was less than 1/100,000, and the incidence was extremely low, but after decades of change, the incidence of prostate cancer has now exceeded 5/100,000, and it is skyrocketing, nearly fivefold to the original. Specifically, the incidence of prostate cancer registered in China in 2012 was 9.92/100,000; the sixth highest incidence of malignant neoplasms in men, and in 2018, prostate cancer patients in Asia accounted for 18,100,001 and half of the world's new patients [12, 13].

Nowadays, many developing countries are facing big dilemmas:

- (i) Medical equipment and treatment level are limited, and with the increase of cancer incidence, it is very easy to cause high misdiagnosis rate if the medical level cannot keep up with the rate of growth of morbidity. Therefore, a large number of social contradictions may appear

- (ii) The large population base in developing countries and the relatively small number of medical institutions, as well as medical personnel, have led to an imbalance in the number of doctors and patients

The capital of China, Beijing, has a large population and has nearly 100,000 old people and children. But on the contrary, there are only 3,000 medical staff serving people in this metropolis. As a result, there is an extreme imbalance between the size of the population and the level of medical. In developing countries, even the big cities still have so many problems, let alone small cities in remote areas. It is usually more acute that the ratio of urban to rural population is about 4:6 in 2017 [14], but the city's medical level is much higher than that in the rural.

In developing countries, doctors in hospitals have to bear a huge burden because of heavy work and mental stress. In China, due to the huge population base, each doctor is allocated to 5,000 patients on average, while a lot of work pressure often leads to high rate of misdiagnosis. Moreover, [15, 16] in the treatment of prostate cancer, 640 images are generated for each patient by PET-CT scan. If the screening test is performed only by manual means, it will be very inefficient and with mental fatigue, working time is too long, and the rate of misdiagnosis will also increase.

Therefore, hospitals in developing countries have similar dilemmas:

- (i) A large number of patients and relatively few medical staff form sharp contradictions, which leads to a large workload of medical staff, easy to cause physical fatigue and mental pressure of doctors, and eventually may increase the rate of misdiagnosis
- (ii) For example, screening PET-CT [17] image is a task with a large number of images and a large amount of repetitive work. If handled manually, it will only be inefficient, waste a lot of time, but also will fatigue the doctor easily, and, last, resulting in a high rate of misdiagnosis
- (iii) There is a "generation gap" between doctors and patients. Patients do not understand the physiological indicators of prostate cancer, which makes it difficult for doctors to communicate with patients

Faced with these problems, the article uses an intelligent medical expert system to improve the above medical conditions. The doctor's final choice can be combined with his own experience through system-assisted diagnosis to make the most complete decision-making treatment. In this system, a neural network-based prostate cancer big data intelligent system is adopted. The system assists the diagnosis decision from two aspects: image and disease index. By adjusting the weights of the two, the two aspects are finally combined to determine the clinical stage of cancer and provide appropriate treatment strategies for physicians. Big data decision-making intelligence system can complete the phased decision of disease quickly and effectively and at the same time can provide timely and effective treatment adjust-

ment to the patient to ensure effective treatment of the disease. In summary, the contributions made by the research are summarized as follows:

- (1) To construct a new model of prostate cancer based on perceptron neural network, which combines a series of indicators and related images to determine whether there is cancer and to determine the severity of the cancer, up to the clinical stage
- (2) According to the images obtained by positron emission computed tomography (PET) or computed tomography (CT) taken by the hospital technicians [18–21], the images can be screened, and the most effective part of the images can be screened for doctors' reference, so as to reduce the burden of doctors and improve work efficiency
- (3) The intelligent system assists in medical diagnosis and through statistical analysis data can be used as an auxiliary diagnosis and automatically provide a fast and accurate treatment plan for doctors
- (4) In order to ensure that the learning neural network can obtain accurate weights and ensure the accuracy of decision-making, the research is based on 2,156,528 hospitalization information of three hospitals in China. The experimental results show that the intelligent medical decision-making system can improve the efficiency of doctors and reduce the rate of misdiagnosis

The rest of this paper is organized as follows: in the second part, introducing the current state-of-the-art related to our work; the third section proposes and constructs a big data decision model; the fourth section gives a detailed description of the experimental performance; the conclusion is given in the last part.

2. Material and Methods

The intelligent auxiliary diagnosis system is a new type of intelligent system based on machine learning and its application has been recognized by the society to a large extent. Through a large number of precedents, the intelligent system can still greatly facilitate people's lives and reduce the related work and life burden. The emergence of the intelligent medical system is expected to improve the current medical and health conditions in developing countries. With the update and development of technology, in many countries and regions, the intelligent medical system can assist in diagnosing of the diagnosis of the patient's health care system information of patients and to help doctors to do some simple relative work, which greatly reduces the workload of doctors, and even able to diagnose the illness and make corresponding treatment decisions, making doctors' to have a more effective understanding of the health situation. On the other hand, in order to increase the accuracy and rationality of machine-assisted diagnosis, the article should not only use simple indicators to make the decision process but

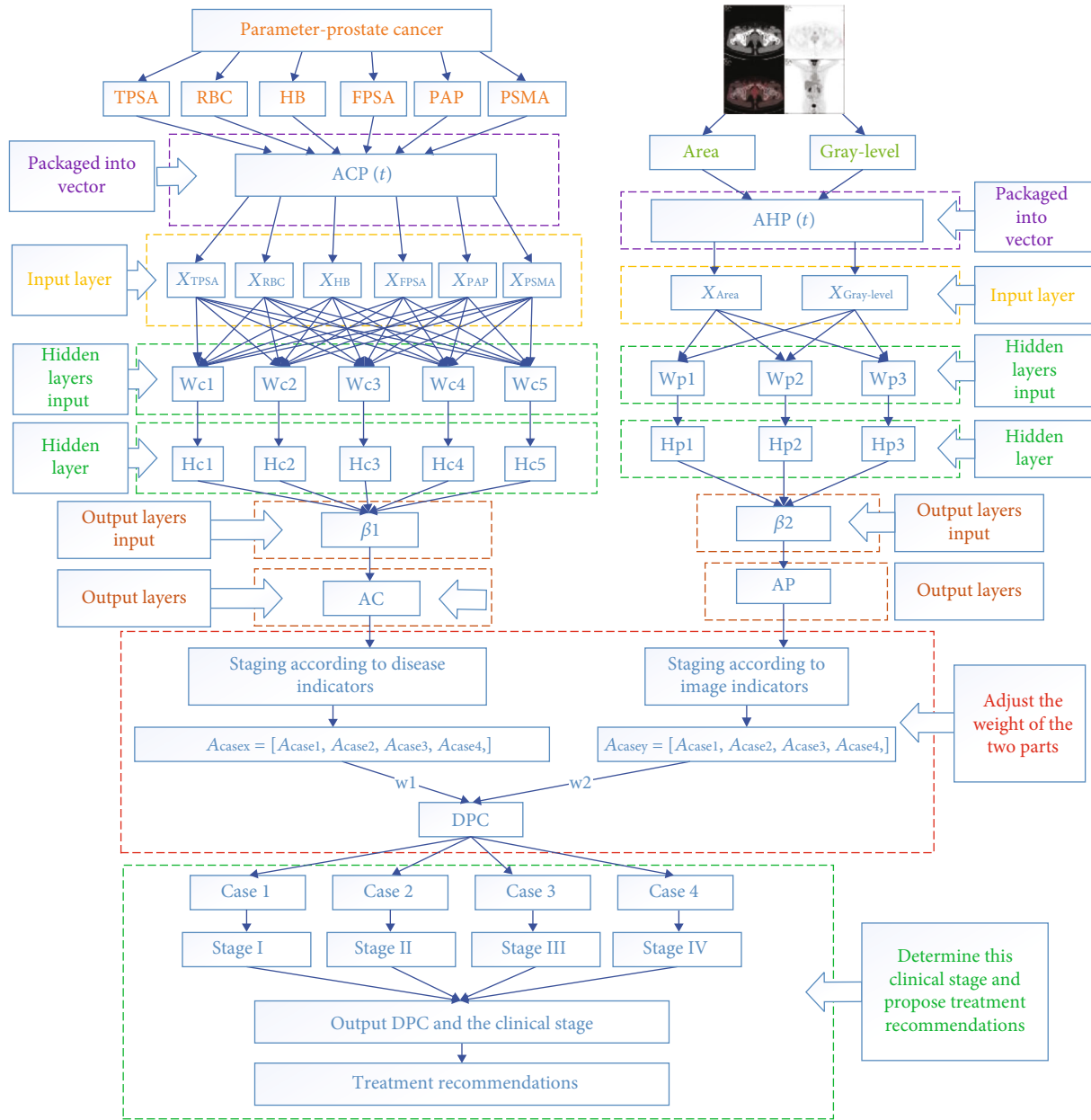


FIGURE 1: The overall structure of the big data-assisted diagnostic decision model.

also include more important PET-CT images to make the decision, because sometimes, the images can judge the disease more directly, accurately, and effectively. Therefore, the article establishes a new decision model based on the perceptron neural networks in developing countries.

2.1. General Framework of the Machine-Aided Diagnosis System Model of Prostate Cancer in the Context of Big Data. During the real medical diagnosis of prostate cancer, doctor makes the corresponding diagnosis through a variety of indicators such as PAP, PSMA, HB, PSMA, TPSA, RBC, PAP, FPSA and patients of positron emission computed tomography (PET) and computed tomography (CT). In real life, doctors will diagnose and treat according to their own knowledge and treatment experience.

In a big data environment, our medically assisted diagnostic systems are also based on big data. Through machine learning, a large amount of data is learned and a mature neural network algorithm is constructed to perform machine diagnosis through medical data and image information. As shown in Figure 1, it is the overall structure of the big data decision model of the intelligent medical system, and it introduces the system's operation process and data flow and processing in detail and clearly.

Finally, because of the strong concurrency of the hospital's intelligent system, it is able to process a large number of patient information at the same time. The system is mainly divided into five stages. The first stage is the pre-processing model of medical images before the input of

the machine-aided diagnosis system of prostate cancer under the big data environment. Finally, it achieved two goals. One is accurate and effective screening of medical images. In the current hospital system, the PET-CT scans generate about 640 images per patient. If the image is processed and screened in a manual form, a large amount of labor, material, and human resources will be wasted. If all images are stored and processed, 640 images need about 2M bytes of space, and only part of the 640 images can be used as effective processing information, which will result in the waste of medical system resources. The other is to calculate the concentration value of the prostate cancer lesion area and carry out edge detection and processing on the effective images to obtain the corresponding lesion area, which can be used as the input information of prostate cancer machine-assisted diagnosis system under the big data environment.

The second stage is the input model of the machine-assisted diagnosis system for prostate cancer under the big data environment. All the obtained digital information will be preprocessed properly, and the irregular, unreasonable, and invalid information is transformed into qualified and effective digital information. Six different disease indexes (tumor markers) can be used to evaluate the clinical staging of prostate cancer produced by two kinds of medical image indexes: the lesion area and lesion area density; then, these two kinds of data integration generate input vector $ACPI$ and $ACP2$, to the different data models.

The third stage is the neural network model of the machine-assisted prostate cancer diagnosis system under the big data environment, which is the operation process of the input layer, hidden layer, and output layer of the neural network. The whole neural network model combined with the diagnosis process of prostate cancer is divided into two parts. The first part is a large number of prostate cancer cases collected and counted in the early stage as the data information of model training, with the purpose of constantly adjusting the weight and bias between various points. Positron emission computed tomography (PET) and computed tomography (CT) patients with six different disease indicators (tumor markers): Hb, PSMA, FPSA, RBC, PAP and TPSA, and two medical image indicators. The lesion area is the density of the lesion area generated by the two vectors as the input of the input layer. The disease index and the medical image are finally found to be the return error of the two-part model parameter adjustment.

The fourth stage is the decision-making model of the machine-aided diagnosis system for prostate cancer under the environment of big data. The final output value is obtained by combining the prostate cancer disease index and the output of the prostate cancer medical image. Combine two different output values, that is, combine the prostate cancer disease index and the prostate cancer medical imaging index to adjust the weight. Because the clinical diagnosis of prostate cancer mainly relies on serum PSA, when digital rectal examination through rectal ultrasound and pelvic MRI are performed through digital rectal examination, CT is less sensitive to the diagnosis

of early prostate cancer. Therefore, the weight of the image in the early diagnosis of prostate cancer should be less than that in the later stage. This requires a corresponding algorithm to combine the two, and finally draw a DPC system output. Then according to the final output value (DPC), the cancer stage is assisted in diagnosis. According to the Tumor Node Metastasis (TNM) guidelines, prostate cancer can be divided into four stages (I, II, III, and IV): prophase, early stage, middle stage, and late stage. The clinical stage of prostate cancer can be comprehensively evaluated according to the final combined output value (DPC).

2.2. The Machine-Aided Diagnosis System for Prostate Cancer Inputs Medical Images for Preprocessing in the Context of Big Data. The clinical manifestations of prostate tumors are the obvious enlargement of tumor and the diffusion of water molecules in living tissues, the increase of cell density, the change of the macromolecular protein content, and the increase of cell density in malignant lesions of cancer cells, which leads to the increase of tumor plane area and darker color of medical image. Therefore, the article uses the lesion area and its gray level of medical images as indicators to assist in the diagnosis.

In the context of big data. Machine-assisted diagnosis of prostate cancer is based on six different disease indicators (tumor markers): HB, PSMA, FPSA, RBC, PAP and TPSA, as well as PET and CT. These two vectors are the lesion area and the density of the lesion area generated as input to the input layer. In terms of medical images of prostate cancer, we are not able to directly obtain the input of two image indexes through PET-CT images. Therefore, this will require preliminary preprocessing of medical images.

First of all, in some medical images, there may be some color labeling, which is called the noise of the image. Therefore, we performed gray-scale processing of prostate cancer medical images to obtain a complete gray-scale cancer image:

$$P_{Gray} = P_R * \alpha + P_G * \beta + P_B * \gamma. \quad (1)$$

P_R , P_G , and P_B , respectively, correspond to the three colors of the original RGB color pattern in the original image. According to the characteristics of prostate cancer medical images, the value range of parameter α is 0.25~0.35, the value range of β is 0.55~0.65, and the value range of γ is 0.1~0.2.

Our ultimate goal is to get the focus area of the medical image of prostate cancer. And because of that, the medical image has a lot of noise; we want to remove the noise of the cancer image. The purpose of Gaussian blur is mainly to reduce the overall image noise. The purpose is to calculate the image gradient and edge more accurately P_{Gray} :

$$h(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2)$$

where σ determines the Gaussian variance for the input parameter in order to determine the degree of blurring of the cancer image.

Then, Gaussian smoothing was performed on the prostate cancer image to obtain the processed prostate cancer image $f(x, y)g(x, y)$:

$$g(x, y) = h(x, y, \sigma) * f(x, y) \quad (3)$$

where $*$ represents convolution, which will be converted into a two-dimensional template for the operation of convolution on cancer images $h(x, y, \sigma)$.

The gradient and amplitude of prostate cancer medical images were then calculated. Image gradient, the direction of the gradient is the direction where the function $f(x, y)$ changes fastest. When there is an edge in the image, there must be a large gradient value. On the contrary, when there is a relatively smooth part in the image, the gray value changes little, and the corresponding gradient is also small. In image processing, the mode of gradient is referred to as gradient, and the image composed of image gradient is called gradient image. The edge of the image has two properties, direction and magnitude. The pixel changes gently along the edge direction and dramatically perpendicular to the edge direction. This change on the edge can be detected by the differential operator, usually using the first or second derivatives to detect edge. We calculate the amplitude and direction of the gradient with the first-order finite difference.

The gradient of the medical image of prostate cancer after smoothing and filtering can be used to calculate the partial derivatives of x and y with the first-order finite difference approximation of $2 \times 2g(x, y)f'_x(x, y)f'_y(x, y)$:

$$\begin{cases} f'_x(x, y) \approx G_x = \frac{f(x+1, y) - f(x, y) + f(x+1, y+1) - f(x, y+1)}{2}, \\ f'_y(x, y) \approx G_y = \frac{f(x, y+1) - f(x, y) + f(x+1, y+1) - f(x+1, y)}{2}. \end{cases} \quad (4)$$

In a rectangular coordinate system, the relation between coordinates, amplitude, and azimuth is

$$\begin{cases} M(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}, \\ Q(x, y) = \arctan\left(\frac{G_x(x, y)}{G_y(x, y)}\right). \end{cases} \quad (5)$$

For prostate medical image, there is a need to get the focus area and need to get the edge image of the focus area, where $M(x, y)$ represents the amplitude, which reflects the edge intensity of the cancer image; $Q(x, y)$ represents the azimuth, which reflects the direction of the gradient; therefore, when $M(x, y)$ obtained the local maximum, its corresponding gradient direction $Q(x, y)$ reflects the direction of the edge (the edge direction is perpendicular to the gradient direction), as shown in the figure.

In order to obtain more accurate edge images of cancer lesion areas, nonmaximal suppression of gradient amplitude is also required at last; that is, the larger the value in the current gradient amplitude matrix is, the corresponding edge point of the cancer image is at this point, which can only indicate that the gradient value of this point is larger. In this way, the step of nonmaximum suppression is completed. A large part of nonedge points are removed, and binary image output is finally obtained, as shown in the figure.

At this point, after detecting the edge of the prostate cancer lesion area, we will try to obtain one of the input parameter areas. This system uses the eight-direction Freeman chain code to calculate the irregular area.

The reason why it is called eight-direction chain code is that it starts from the starting pixel point and scans the next adjacent pixel point, then looks at the relative position of the original pixel point and the adjacent pixel point, encodes according to the eight-direction diagram, and scans successively to obtain the Freeman chain code.

Then, carry out front vector and back vector annotation on the chain code:

$$\begin{aligned} s1[i] &= \begin{cases} -1, & i > 0, & pi \times 1[i] \cdot y > pi \times [i-1] \cdot y, & i = 0, & pi \times 1[0] \cdot y > pi \times [N-1] \cdot y, \\ 0, & i > 0, & pi \times 1[i] \cdot y = pi \times [i-1] \cdot y, & i = 0, & pi \times 1[0] \cdot y = pi \times [N-1] \cdot y, \\ 1, & i > 0, & pi \times 1[i] \cdot y < pi \times [i-1] \cdot y, & i = 0, & pi \times 1[0] \cdot y < pi \times [N-1] \cdot y, \end{cases} \\ s2[i] &= \begin{cases} -1, & i < N-1, & pi \times 1[i+1] \cdot y > pi \times [i] \cdot y, & i = N-1, & pi \times 1[N-1] \cdot y > pi \times [0] \cdot y, \\ 0, & i < N-1, & pi \times 1[i+1] \cdot y = pi \times [i] \cdot y, & i = N-1, & pi \times 1[N-1] \cdot y = pi \times [0] \cdot y, \\ 1, & i < N-1, & pi \times 1[i+1] \cdot y < pi \times [i] \cdot y, & i = N-1, & pi \times 1[N-1] \cdot y < pi \times [0] \cdot y. \end{cases} \end{aligned} \quad (6)$$

Add the former vector annotation and the latter vector annotation, and the result is $s[i]$:

$$s[i] = \begin{cases} -1, & a = 1, 2, 3, \\ 0, & a = 0, 4, \\ 1, & a = 5, 6, 7. \end{cases} \quad (7)$$

The final target area can be expressed as

$$P_{\text{Area}} = \sum_{i=0}^{N-1} \text{pix1}[i]x * s[i] + N_1. \quad (8)$$

After converting the original image to grayscale image, calculate another parameter of the model input:

$$P_{\text{Gray-level}} = \frac{1}{\text{nl}} \sum_{i=1}^{\text{nl}} V_{\text{pi}}, \quad (9)$$

where nl is the total pixel point in the prostate cancer lesion area and is the pixel value corresponding to the pixel point V_{pi} .

In this module, we have two purposes: first, the model needs to screen the medical images of prostate cancer through two indicators; second, the need to finally obtain the input and $P_{\text{Area}}P_{\text{Gray-level}}$.

2.3. Input Model of the Machine-Aided Diagnosis System for Prostate Cancer in the Context of Big Data. The input model of the machine-aided diagnosis system of prostate cancer in the context of big data needs to properly process all the digital information obtained and convert irregular, unreasonable, and invalid information into qualified and effective digital information. In particular, such digital information cannot be faulted, and it is more necessary to process irregular and unreasonable invalid information into effective input, so as not to affect the operation of the system.

In the prostate cancer machine-assisted diagnosis system in the context of big data, the input is divided into two parts. One part is the input of disease indicators into $ACP(t)$. Among these test results, tumor markers are the most valuable signals for diagnosis, treatment, and prognosis. Therefore, the prostate disease index input consists of six disease indexes, namely, HB, PSMA, FPSA, RBC, PAP, and TPSA. The vector $ACP(t)$ that combines these six disease indexes is the input to the system:

$$ACP(t) = [X_{\text{TPSA}}, X_{\text{RBC}}, X_{\text{HB}}, X_{\text{FPSA}}, X_{\text{PAP}}, X_{\text{PSMA}}]. \quad (10)$$

Although the disease index of prostate cancer is an important indicator to assist the diagnosis, it is obviously too superficial and insufficient to simply rely on the disease index as the judgment and decision. Therefore, based on taking the disease index as the input of the system to make the diagnosis decision, the system also adds a part of the input image index input $AHP(t)$. $AHP(t)$ is two kinds of medical image indexes, the lesion area and lesion area density, pro-

duced by PET-CT taken by patients. And these two indicators form a vector $AHP(t)$ as another input to the system $P_{\text{Area}}P_{\text{Gray-level}}$:

$$ACP(t) = [P_{\text{Area}}, P_{\text{Gray-level}}]. \quad (11)$$

The input model is not only the beginning of the theoretical machine-aided diagnosis system but also the beginning of the two input neural network models in the input layer of neural network.

2.4. Neural Network Model of Machine-Aided Diagnosis System for Prostate Cancer under Big Data Environment. This part is the study of prostate cancer by six different disease indexes (tumor markers), Hb, PSMA, FPSA, RBC, PAP, TPSA, and PET-CT of two kinds of medical image metrics, the lesion area P_{Area} and lesion area density $P_{\text{Gray-level}}$ generated by the two vectors as input, and then adjusted by weight to study the diagnosis process of prostate cancer by disease index and medical image, respectively. Therefore, this stage of research should be divided into two parts: the first part is the reasonable evaluation and allocation of the weight of disease indicators and medical image indicators; the second part is the diagnosis process of the intelligent diagnosis system, and a preliminary diagnosis result is obtained by adjusting the weight and bias of tumor markers.

2.4.1. Reasonable Evaluation and Allocation of Disease Index and Image Index Weight of Prostate Cancer. In this study, the inpatient information of more than 8,000 prostate patients in three hospitals in China was studied, and about 2,156,528 data items were collected and statistically analyzed. Therefore, under such a large data volume, we can continuously adjust and evaluate the weight through experiments. The neural network model of the machine-aided diagnosis system of prostate cancer in the big data environment is a process of reasonably adjusting the weight and deviation of each node of prostate cancer after input. According to the overall structure of intelligent medical big data decision model diagnosis system, the neural network model following the process of the diagnosis of prostate cancer is divided into two parts. On the left side is through early collection of statistics of a large number of cases of prostate cancer as a model of training data information; the goal is to continuously adjust the weights and bias between the various points. With six different disease indexes (tumor markers): Hb, PSMA, FPSA, RBC, PAP, and TPSA generated by the vector $ACP(t)$ as input; intelligent diagnosis system is ultimately getting the return error of disease index to adjust the parameters of the intelligent diagnosis system. The right part is a large number of prostate cancer case images collected and counted in the early stage, and two medical image indexes obtained through preliminary preprocessing are used as data information of model training, so as to constantly adjust the weight and bias between various points. And in patients with a type of PET-CT of two kinds of medical image indexes, the lesion area and lesion area density, generated vector $AHP(t)$ as input, intelligent diagnosis system is ultimately getting the return error of disease index of intelligent diagnosis system with

parameter adjustment $P_{\text{Area}}P_{\text{Gray-level}}$. Therefore, the input of the two parts can be expressed by vectors as follows:

$$\begin{aligned} \{X_{\text{TPSA}}, X_{\text{RBC}}, X_{\text{HB}}, X_{\text{FPSA}}, X_{\text{PAP}}, X_{\text{PSMA}}\} &= \text{ACP}(t), \\ \{X_{\text{Area}}, X_{\text{Gray-level}}\} &= \text{AHP}(t). \end{aligned} \quad (12)$$

When the $\text{ACP}(t)$ and $\text{AHP}(t)$ generated by six tumor markers and two medical image indicators are inputted into the intelligent diagnosis system, the weight between neurons is adjusted to form the input of the next neuron, and each neuron has a corresponding activation function, which is used to process the data.

The activation function selected is sigmoid function, which is a common S-shaped function in biology, also known as the S-type growth curve. Because its value range is (0, 1), the input vector $\text{ACP}(t)$ and $\text{AHP}(t)$ of prostate cancer can be mapped to the interval (0, 1). In the process of big data medical-aided diagnosis of prostate cancer, there is no linear connection between the upper and lower layers of the neural network, and the difference between these features is complex but not large. Therefore, sigmoid function is a suitable choice for the model of big data-assisted diagnosis of prostate cancer.

$$f = \frac{1}{1 + e^{-x}} \quad (13)$$

In a multilayer network, the output of the previous layer will serve as the input of the later layer:

$$\begin{aligned} a^{m+1}_{(\text{ACP}(t), \text{AHP}(t))} &= f^{m+1}(W^{m+1}a^m + b^{m+1}), \\ m &= 0, 1, \dots, M-1. \end{aligned} \quad (16)$$

M is the number of layers in the neural network. Six disease indicators and two medical image indicators can be calculated by weight to obtain a final cancer result, which we call the expected result (the desired result). Comparing the expected result and the actual result (actual results), the error function will reach the minimum value $\min E$ according to the gradient descent method. The purpose is to make the diagnosis decision result more and more close to the real value by adjusting the weight, so that the weight and bias can reach a state of "perfect". The error function of desired result and actual results are

$$E_{(\text{ACP}(t), \text{AHP}(t))}(w, b) = \frac{1}{2} \sum_{k=0}^{n-1} (D_{\text{Desired result}} - D_{\text{actual results}})^2. \quad (14)$$

Among them, the number of nodes n as the final output to diagnose prostate cancer under the large data of expected results with the real result would be a differentiation value on prostate cancer; we eventually have to reach the goal to get the value to minimize the differentiation of prostate cancer. When it reaches a certain small value is an optimal solution, and finally, we, through the optimal solution to adjust the weights and bias, optimize weights and bias:

$$\begin{aligned} w_{(\text{ACP}(t), \text{AHP}(t))} &= w_{(\text{ACP}(t), \text{AHP}(t))} - \eta_1 \frac{\partial E(w, b)}{\partial w_{(\text{ACP}(t), \text{AHP}(t))}} \\ &= w_{(\text{ACP}(t), \text{AHP}(t))} - \eta_1 \delta x, \\ b_{(\text{ACP}(t), \text{AHP}(t))} &= b_{(\text{ACP}(t), \text{AHP}(t))} - \eta_2 \frac{\partial E(w, b)}{\partial b_{(\text{ACP}(t), \text{AHP}(t))}} \\ &= b_{(\text{ACP}(t), \text{AHP}(t))} - \eta_2 \delta, \end{aligned} \quad (15)$$

η is learning efficiency, and diagnostic accuracy is the essential requirements of the whole model. The higher the accuracy is, the more perfect the intelligent diagnosis system is. We adjust the weight; the diagnosis accuracy is the purpose of increase through six tumor markers and two medical image indexes which generated $\text{ACP}(t)$ and $\text{AHP}(t)$ two inputs; through the distribution of the weight, get the final diagnosis decision results. Therefore, it is crucial to adjust the weights and bias to achieve a "perfect state".

2.4.2. The Diagnosis Process of Prostate Cancer Intelligent Diagnosis System. In the first stage, we have obtained the weight and bias between each node by training with a large amount of data. That is to say, the ownership weight of the model of big data intelligent medical auxiliary diagnostic system has been determined. Therefore, at this stage, the process of the diagnosis of intelligent system will be analyzed. First of all, the six different disease indexes (tumor markers): HB, PSMA, FPSA, RBC, PAP, TPSA, and PET-CT of two kinds of medical image metrics: two vectors $\text{ACP}(t)$ and $\text{AHP}(t)$ generated by Area and Gray-level are used as inputs.

The next section will introduce the relationship between disease indicators and image indicators and how to adjust the weight proportion between the comprehensive two diagnosis results.

2.5. Decision-Making Model of Machine-Aided Diagnosis System for Prostate Cancer in the Context of Big Data

2.5.1. Decision Model of Expert System Based on Rule Reasoning. From the previous step, the article has obtained the auxiliary diagnosis results derived from prostate disease indicators and medical imaging indicators. But the final decision needs to combine the two to make a decision diagnosis of prostate cancer. First, you need to obtain a comprehensive index and then make a decision diagnosis, so that the decision can be more accurate and reliable. Therefore, this section introduces the relationship between the disease index and the image index and how to adjust the weight ratio between the two comprehensive diagnosis results.

Among the decision models, the rule reasoning model is a new weight adjustment algorithm introduced by us, which is developed from the rule-based expert system, a branch of the expert system in the field of artificial intelligence. In the decision-making model of the machine-aided diagnosis system of prostate cancer in the context of big data, the article uses a simpler expert system hair model based on rule reasoning.

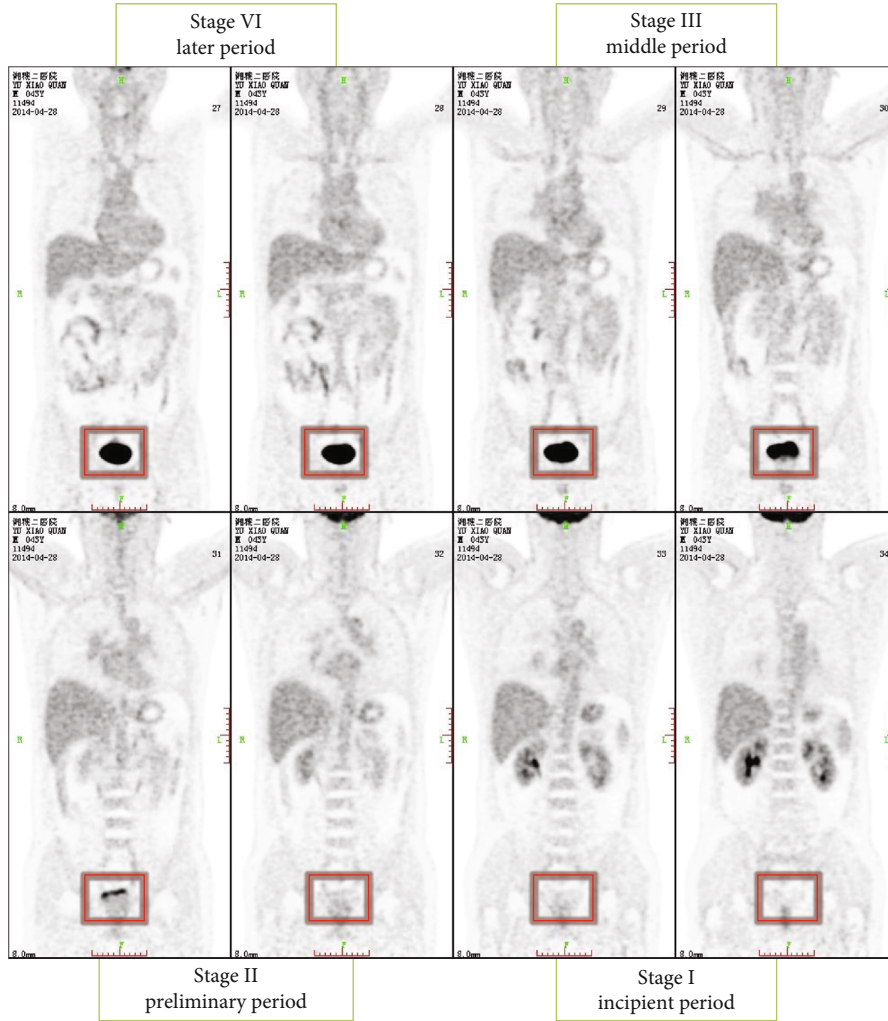


FIGURE 2: The clinical staging marker image of prostate cancer.

TABLE 1: Reasonable distribution of weights corresponding to different staging results.

	Case1	Case2	Case3	Case4
Case1	Weight (C1, C1)	Weight (C1, C2)	Weight (C1, C3)	Weight (C1, C4)
Case2	Weight (C2, C1)	Weight (C2, C2)	Weight (C2, C3)	Weight (C2, C4)
Case3	Weight (C3, C1)	Weight (C3, C2)	Weight (C3, C3)	Weight (C3, C4)
Case4	Weight (C4, C1)	Weight (C4, C2)	Weight (C4, C3)	Weight (C4, C4)

The rule-based inference model is a knowledge set expressed by rules, including the knowledge needed for reasoning execution. As shown in Figure 2, in the process of the diagnosis of prostate cancer, low-level prostate cancer clinical symptoms are not particularly obvious, but in a high level of clinical symptoms of prostate cancer is obvious and in the condition of prostate cancer research found that when patients clinical staging of prostate cancer in stage III or IV stage image characteristics than I or II when more obvious, therefore, if the step we have got prostate disease index and auxiliary diagnosis of medical image index to are still in the stage I or II, this may cause the index weight of the disease to be greater than the image index.

In the working memory of the rule-based inference model, the complete data set required for inference execution will be obtained. And the analysis results can be roughly obtained by analyzing the changes of the index characteristics corresponding to different clinical stages as shown in Table 1.

Finally, it is the setting of an inference engine. When two diagnostic results are taken as conditions, in order to complete the inference process, it is necessary to determine which rule in the current cycle needs to be activated. In the decision-making model of the machine-assisted prostate cancer diagnosis system in the context of big data, each diagnosis is a definite value, and each set of conditions has a

corresponding data in the workspace. Therefore, in the decision-making model of the machine-aided diagnosis system of prostate cancer in the context of big data, when conditions are determined, the corresponding rules can be determined, and usually there will be no rule conflict.

2.5.2. Staging Management and Treatment Suggestions of Prostate Cancer Machine-Assisted Diagnosis System under Big Data Environment. In the decision-making model of the machine-assisted prostate cancer diagnosis system in the context of big data, get the final result through rule-based reasoning model DPC. Based on this result, perform cancer staging and then cancer machine-assisted diagnosis staging based on the final DPC. According to the tumor node metastasis (TNM) guidelines, prostate cancer is mainly divided into four stages (I, II, III, and IV) in the early, middle, and late. Based on the final output value (DPC), the clinical staging of prostate cancer was comprehensively evaluated. Table 2 shows the staging threshold of the machine-assisted diagnosis system for prostate cancer. At the same time, through the comprehensive evaluation of clinical staging judgment and phase estimation, the machine-aided system can recommend different treatment schemes to doctors, such as active monitoring, drug therapy, resection, endocrine therapy, radiotherapy, and chemotherapy.

3. Experimental Performance

3.1. Experimental Data Collection, Classification, and Preprocessing

- (1) Before the experiment, collect all the medical information needed from three first-class hospitals in China: Xiangya Hospital, Xiangya Second Hospital, and Xiangya Third Hospital, and classify these information. Besides, the data center collects various information items from different aspects to help the system-assisted diagnosis
- (2) The team collects a large number of prostate cancer medical data and makes corresponding classification statistics and preprocessing and then does the corresponding recording work. Because of the large amount of medical data collected, although it is medical data related to prostate cancer, only a part of the medical data is used in this system. Therefore, in order to ensure the smooth progress of the experiment and its rationality and accuracy, we extracted more than 8,000 prostate patients' data from the data collected by the three hospitals and extracted 16,143 items of systematic structured and effective information from the data of these patients. Figure 3 shows the classification items of medical data information about prostate cancer collected by the three hospitals

3.2. Initialize System Data Variables. (1) After testing and analyzing more than 8,000 cases of prostate patients as a training set, some parameters of the intelligent medical system are set as follows: initial weight determination; initial weight should be set to different values. Therefore, as for

TABLE 2: The staging threshold of the machine-assisted diagnosis system for prostate cancer.

Clinical stage	Stages I	Stages II	Stages III	Stages IV
Threshold (DPC)	(0, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1.00]

the initialization of the weights, design a random occurrence program to generate a set of 0~0.5 random numbers as the initial weight of the network; as for the minimum training rate, the initial value can be set to 0.9 since the training rate will be automatically adjusted; as for the dynamic coefficient, its selection is generally 0.6~0.8; the allowable error is 0.001~0.00001. According to clinical medical standards, some standard medical indicators data are shown in Table 3.

(2) The intelligent system image index initialization information is preprocessed according to two medical image indicators generated by PET-CT, and the area of the lesion area and the lesion area are obtained. Gray-level two image indicators, the resulting vector $ACP(t)$, as the initial input to the intelligent system.

In this process, because prostate cancer images have a lot of noise, preexperiment experiments were carried out through a large number of images of prostate cancer before the experiment, and various parameters suitable for processing prostate cancer images were calculated and grayed out by filters and double thresholds. The postprocessing of the image is shown in Figure 1. The edge intensity of the lesion area is compared with the surrounding edge intensity. The resulting edge range will be more accurate and ensure that the final intelligent system correctly diagnoses the diagnosis. Figures 4 and 5 show the comparison of the intensity of edge pixels after detection with the intensity of surrounding pixels. It can be clearly seen that the edge intensity is much higher than the surrounding pixel intensity. In Figure 5, the x -axis and y -axis is the length and width of the image, respectively, and the z -axis corresponds to the gray level of the grayscale image. The 3D effect shows that the gray scale of the prostate cancer lesion area is much higher than that of other regions.

The whole preprocessing process is to first perform grayscale processing on the image and then calculate the gradient amplitude of the grayscale image, nonmaximum value suppression, and the threshold processing finally obtains the binary image output and obtains a binary value suitable for calculating the area of the lesion area. The image is calculated by the Freeman chain code for the irregular connected graph, as shown in Figures 6 and 7:

After obtaining the two-part data index, the system uses the data index generation vector as the initial input of the system model. Only accurate indicator data can make an accurate auxiliary diagnosis.

(3) In the big data environment of more than 8,000 patients with prostate cancer, the intelligent medical system trained the data to get the corresponding weight. The intelligent system has a specific numerical division of the four clinical stages of prostate cancer. In terms of medical disease indicators, the smart medical system is through six prostate cancer disease indicators and two medical image indicators by patients were used for assisted diagnosis. Therefore, in

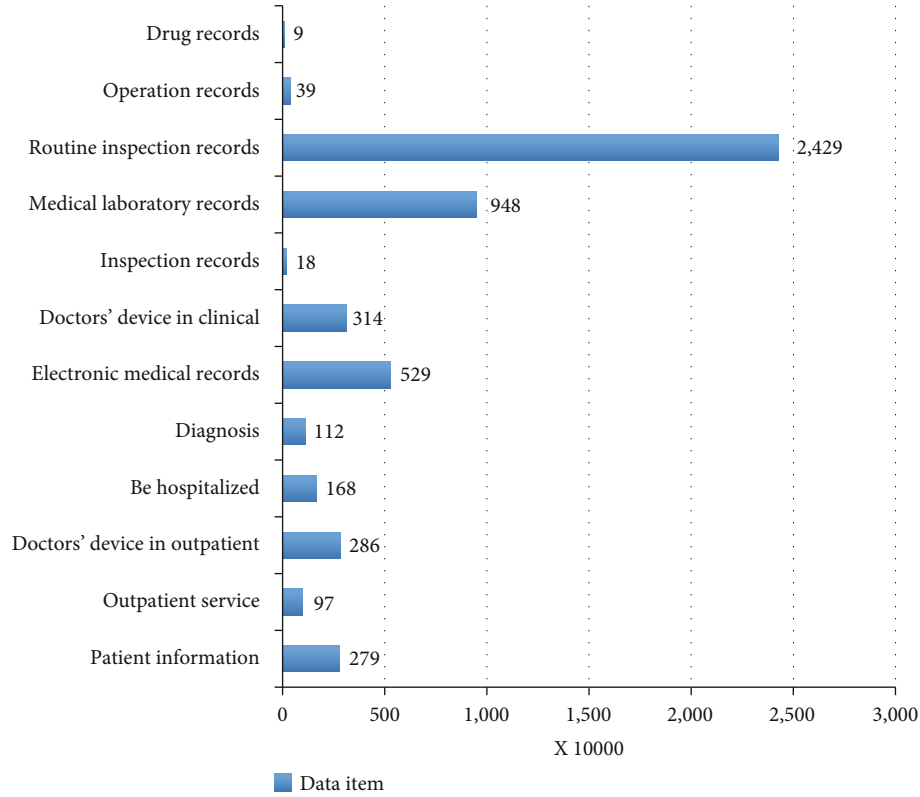


FIGURE 3: The classification of items for collecting medical data for prostate cancer patients.

TABLE 3: The standard indicator data for normal people.

Indicator	Normal range
Total prostate-specific antigen (TPSA)	0-4.0 ng/ml
Free prostate-specific antigen (FPSA)	4-20 μ g/L
Hemoglobin (HB)	120~165 g/L
Red blood cell (RBC)	12~15 g/100 ml
Acid phosphatase (PAP)	0~3.5 ng/ml
Prostate-specific membrane antigen (PSMA)	0~4 ng/ml
Area of the lesion-area	<15 mm
Gray-level of the prostate	<180

the two modules of prostate cancer disease indicators and medical image indicators, according to the weight of each disease indicator and medical image indicators plus the training result data, the system will give the weight between every two nodes reasonable. The weight distribution of prostate cancer disease indicators and image indicators is shown in Figure 8.

(4) In the big data intelligent medical system decision model, in order to accurately determine the clinical stage of prostate cancer, the system combines the disease index and the medical image index to assist the diagnosis. However, prostate cancer often occurs in the posterior lobe, and there are no obvious symptoms in the early stage. Patients with prostate cancer often have symptoms such as frequent urination, difficulty in urinating, thinning of urine, prolonged uri-

nary tract, dysuria, and urinary retention. Prostatic hyperplasia is the same, so it is difficult to diagnose prostate cancer based on this, mainly relying on digital rectal examination to make judgments. Therefore, in the rule-based reasoning algorithm, set the rules to the early, middle, and late three weights. Because medical images are not very obvious in early prostate cancer, if the prostate cancer disease index model and the image index model result are in the early stage, the image model will occupy a smaller proportion. If the prostate cancer index model and the image index model results are at a later stage, the proportion of medical image indicators is gradually increasing. Table 4 shows the rational distribution of the corresponding weight $W_{\text{disease_indicator}}$ of the disease indicator. The corresponding medical image indicator is expressed as $W_{\text{image_indicator}}$:

$$W_{\text{image_indicator}} = 1 - W_{\text{disease_indicator}} \quad (16)$$

As shown in Table 4, since medical imaging has no obvious characteristics in early prostate cancer, the reasonable allocation weights of disease indexes and imaging indexes are divided into 4 cases to set the weights reasonably. (a) When both the prostate disease index and the medical imaging index point to stages I and II, the weight of the disease index is set to be slightly larger than the weight of the medical imaging index. (b) When both the prostate disease index and the medical imaging index point to the third and fourth periods, the weight of the disease index is set to be slightly smaller than the weight of the medical imaging index. The

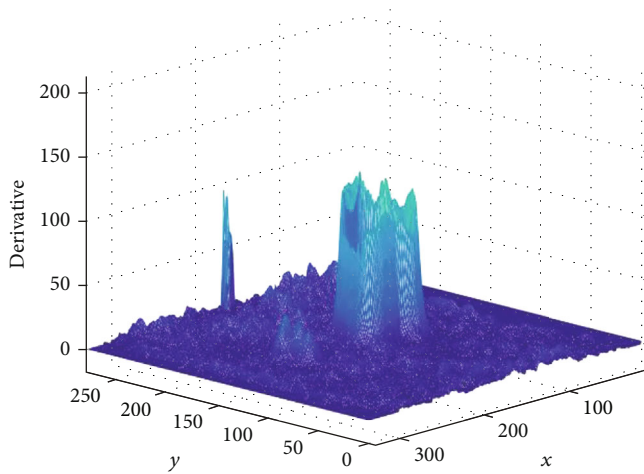


FIGURE 4: Three-dimensional image of pixel intensity in prostate cancer lesions.

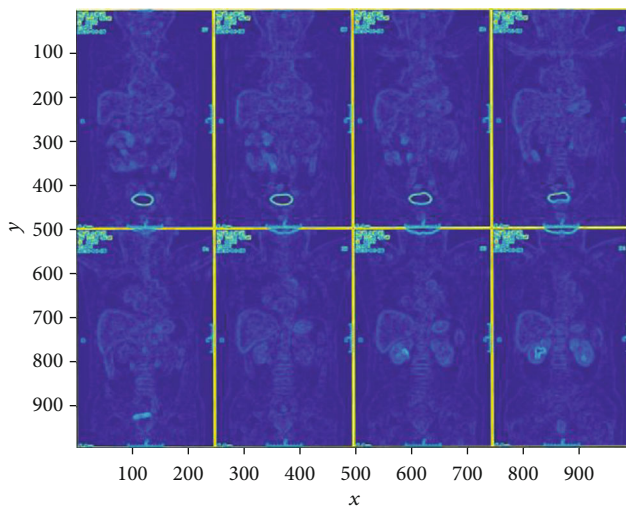


FIGURE 5: Plane display of pixel intensity in prostate cancer lesions.

other two situations are somewhat special. The clinical stage of the disease index model and the medical image model are completely different. Because in the early stages of prostate cancer, medical image detection is indeed very convincing. Therefore, in order to ensure accuracy and rationality, our weight design has been processed as follows. (c) When the medical image index points to stages I and II and the prostate disease index points to stages III and IV, the weight of the disease index is set to be much larger than the weight of the image index. (d) When the medical image index points to stages III and IV, and the prostate disease index points to stages I and II, the weight of the disease index is set to be much smaller than the weight of the image index.

3.3. Data Evaluation and Analysis of Prostate Cancer in the Big Data Environment. TPSA is a specific marker of prostate cancer, and it is also meaningful for the diagnosis of prostate cancer with no obvious symptoms in the early stage. As shown in Figure 9, in the five years from 2011 to 2015, the

average value of TPSA from 2011 to 2013 increased from 18.63 ng/ml to 45.2 ng/ml, reaching its peak in 2013, but declined from 2013 to 2015. The trend in the past five years indicates that the condition of prostate patients is effectively controlled. It is generally believed that TPSA is less than 4 ng/ml normal value, and TPSA is greater than 10 ng/ml, which indicates an increased risk for prostate cancer patients. When the average level exceeds 50 ng/ml, the patient is likely to have prostate cancer.

FPSA/TPSA, another important diagnostic indicator for the big data intelligent assisted diagnostic system, is also an important basis for assisted diagnosis and treatment. As shown in Figure 10, the normal range of FPSA/TPSA is around 0.25, but for the five years from 2011 to 2015, the FPSA/TPSA values are below 0.25 and are in a decline from 2011 to 2014. The value of FPSA/TPSA has fallen to 0.05 in 2014, and when FPSA/TPSA is below 0.1, the incidence of cancer will reach a high level of 56%. This shows that most patients with prostate cancer have been deteriorating during this period time and are not well-controlled. In these structured medical data, we can further realize that in these five years, most of the patients in the three hospitals are in the stages III and IV of cancer.

So far, the causes of prostate cancer have been divided into two major modules, one of which is genetically leading to prostate cancer. To prove this claim, we extracted data from all major prostate patients from 2011 to 2015 and made the analysis for a dangerous proportion of data, setting the number of uninherited patients to 1, and the genetic patients are the multiples of uninherited patients. From the tens of thousands of data, we can roughly draw two conclusions. One is that the number of patients with genetic disease is more than five times of the total number of patients, which indicates that heredity is one of the major causes of prostate cancer; another point is that it can show from the trend of data that this multiple relationship is constantly rising between 2011 and 2015, and the proportion of genetic diseases is increasing at the same time as the number of patients increases. This shows that the number of genetic patients is increasing year by year.

The other is diet that leads to prostate cancer. First, people who love smoking have a higher chance of developing cancer. Many people know that smoke contains a lot of dirty things, and in the process of smoking, those substances will be sucked into the body. Second, vitamin A is a trace element required by the human body. Absorption of a certain amount of vitamin A can promote the healthy development of human body, and excessive amount will lead to a large increase in saturated fatty acids in the human body, thereby causing lesions in prostate cells. Third, carotene has a certain promoting effect on the pathogenesis of prostate cancer. Fourth, a high-fat diet is also a cause of prostate cancer because it contains many animal fats and some fatty acids. If people eat more of these things, some of the meat in the body is digested into different types of fatty acids; so in the long run, the chance of developing prostate cancer will be greatly improved. So this is why many doctors say that patients who have cancer have to eat less meat, because eating more meat will hurt their body's immune system.

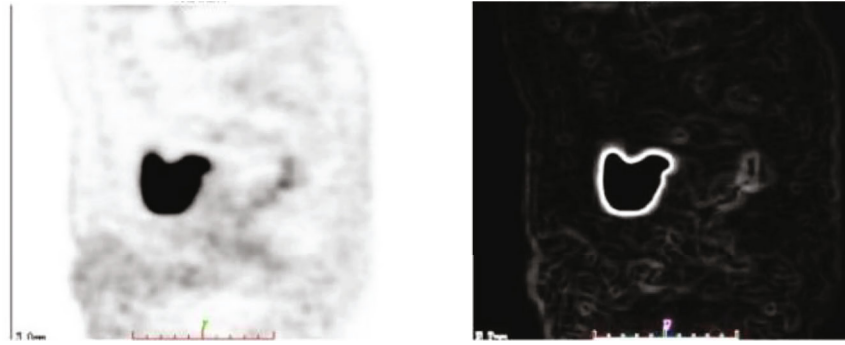


FIGURE 6: Comparison of grayscale image and gradient amplitude image.



FIGURE 7: Nonmaximum suppression image and binary image output.

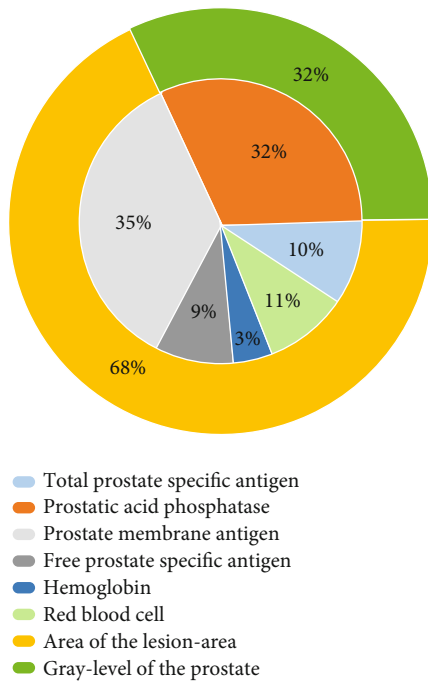


FIGURE 8: Weight distribution of various indicators of prostate cancer.

The incidence and mortality of prostate cancer are second only to lung cancer and ranks second in cancer mortality. As shown in Figure 11, the first-stage cure rate of prostate cancer is over 95%, the second-stage cure rate is over 80%, the third-stage cure rate is less than 40%, and the fourth-stage cure rate is only 25%. The prostate cancer intelligent diagnosis system provides patients with corresponding

treatment recommendations based on the system diagnosis results, as shown in Table 5:

$$\text{True positive rate} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}, \tag{17}$$

$$\text{False positive rate} = \frac{\text{false positive}}{\text{false positive} + \text{true negative}}.$$

The prostate cancer machine-assisted diagnosis system proved feasible in the experiment. As shown in Table 6 and Figure 12, the receiver operating curve (ROC), the ROC is the true positive rate – false positive rate curve: the x-axis represents true; positive rate, the y-axis represents the false positive rate. From Figure 12, we can see that the curve is very close to (0, 1) point [22–24]. Therefore, the classification accuracy of the prostate cancer machine-assisted diagnosis system is still very high. We can compare this accuracy with the expert doctor. The results are as follows:

The main role of the prostate cancer intelligent diagnosis system is self-evident, mainly for the preliminary diagnosis for prostate patients, and to collect statistical information according to the data. During the past five years, the vast majority of the more than 8,000 patients in the three hospitals have gradually recovered under the treatment of doctors. In this experiment, the patient enters the hospital for diagnosis, and the doctor will generate a diagnosis book for the first diagnosis. Our data are from three famous Xiangya hospitals of Central South University. The accuracy of the diagnosis is defined as the accuracy of the first patient diagnosis, not the accuracy of the patient after multiple diagnoses. Figure 13 shows the comparison of the diagnostic accuracy between doctors and intelligent systems. The blue bar represents the doctor’s diagnostic accuracy, and the green bar represents

TABLE 4: Weight distribution of combining disease indicators and image indicators.

Disease indicators Image indicators	Stage I	Stage II	Stage III	Stage IV
Stage I Stage II	$W_{\text{image_indicator}} < W_{\text{disease_indicator}}$		$W_{\text{image_indicator}} \ll W_{\text{disease_indicator}}$	
Stage III Stage IV	$W_{\text{image_indicator}} \gg W_{\text{disease_indicator}}$		$W_{\text{image_indicator}} > W_{\text{disease_indicator}}$	

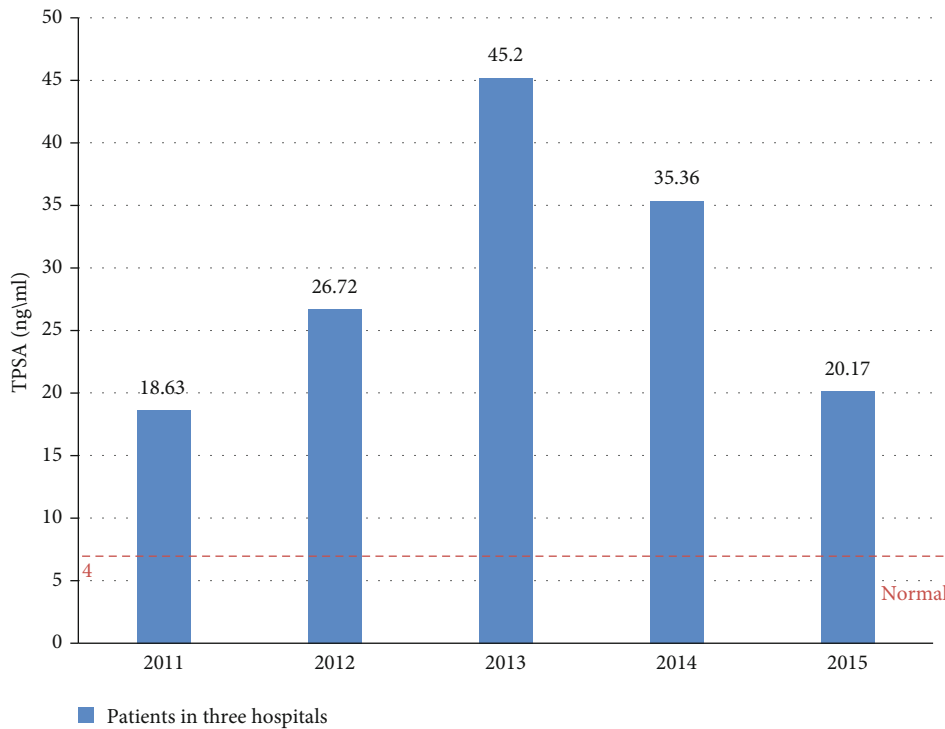


FIGURE 9: The TPSA growth trend of the three hospitals from 2011 to 2015.

the machine-assisted diagnostic accuracy. When the number of cases is relatively low, it is obvious that the diagnosis accuracy of artificial experts is much higher than that of intelligent systems. However, as the number of cases increases, the diagnostic accuracy of artificial experts declines, while that of intelligent systems increases. According to the specific data, when the number of cases increased from 200 to 8000, the accuracy of manual expert diagnosis decreased from 97% to 81%, while the diagnostic accuracy of intelligent systems increased from 61% to 87%. Therefore, it can be concluded that when the diagnostic data reaches a certain amount, the accuracy of the intelligent diagnosis system for prostate cancer will exceed that of the artificial expert. Therefore, the intelligent auxiliary diagnosis system can assist the artificial expert as an auxiliary diagnosis.

4. Discussion

In recent years, more and more people have proposed the concept of intelligent medical system. Many medical systems have implemented different methods and techniques and

involve many related technologies such as big data medical treatment and deep learning. Models for assisting prediction, diagnosis, and treatment have also become hot topics in medical applications and intelligence. Next, the article will introduce the research status related to our works in detail.

In the social discussion of the smart medical system, many people in the smart medical industry still hold an optimistic attitude and believe that the smart medical market is greatly guided by policies. With the development of society, smart medical will inevitably become a major trend. The medical reform is now divided into three stages: the first is digitization, which is to digitize a lot of information including medical records, medical history, and information for detecting a certain disease; the second stage is the hospitalization of mobile, this stage is more advanced than digital on the first floor, the change brought about is that the original patient and caregiver's space inside the hospital is not limited; the third stage is remoteization.

In fact, the three stages of medical intelligence are all related to the development of technology, and the reliability of technology makes medical institutions to trust enough.

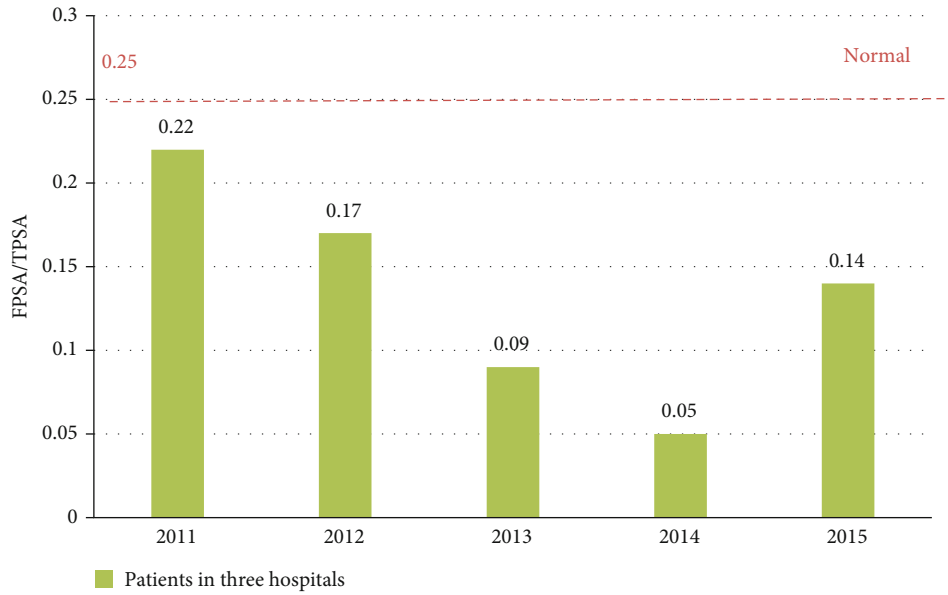


FIGURE 10: The changes in FPSA/TPSA trends in the three hospitals from 2011 to 2015.

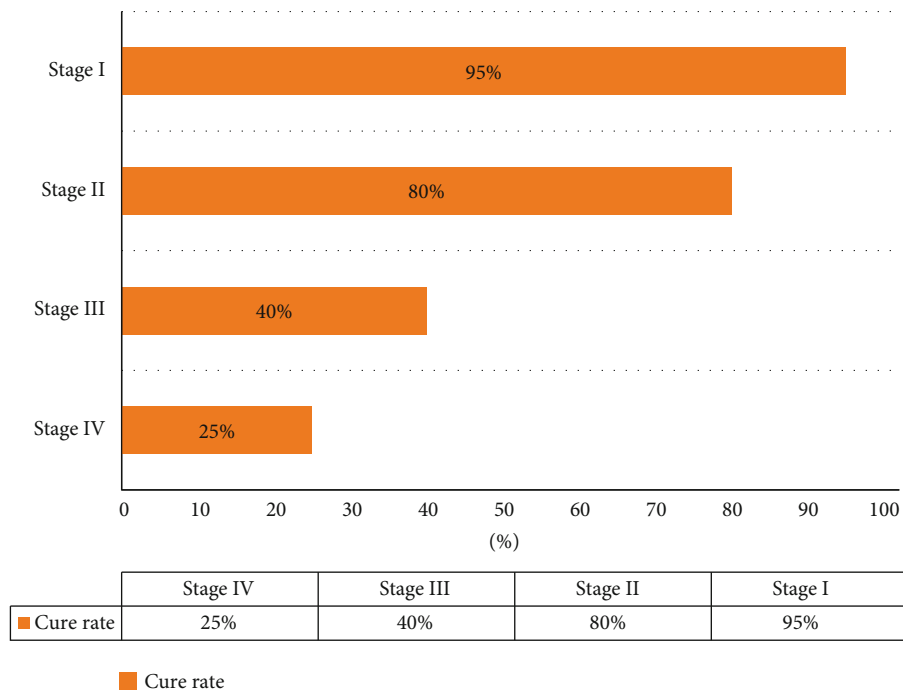


FIGURE 11: The cure rate of four stages of prostate cancer.

For example, electronic medical records are the development of smart terminals and networks in the hospital, while mobile hospitals rely on the progress of Wifi technology, and telemedicine relies on the development of 3G and 4G networks and mobile Internet and involves semiconductor manufacturing processes. Technology allows devices to work better with lower efficiency, such as the rapid transmission of image information after detection and the development of the Internet of Things, making the connections between devices and devices and people more smooth.

In recent years, the overall development of smart medical systems has been slow. The main reason lies in two points. The first point is that big data processing does not have an overall architecture. The second point is that the medical data model is messy and untargeted.

In the study of the article, the system will be able to directly connect to the CDR data interface in the hospital and batch process the medical big data. Among them, internal operations related to premedical data processing, data cleaning, and other operator operations are involved. For

TABLE 5: Treatment recommendations of four stages for prostate cancer.

Clinical stage	Treatment advice
Stage I	Pay attention to physical changes and observe carefully
Stage II	Prostatectomy, radioactive implantation
Stage III	Endocrine therapy, orchiectomy (nonsteroidal antiandrogen intermittent therapy)
Stage IV	Chemotherapy (CTX, 5-FU, ADM, VLB, and PTX)

TABLE 6: Parameter-lists of true and predicted parameter for machine-assisted diagnosis of prostate cancer.

	Actual value		Total
Predictive value	True positive (TP)	False positive (FP)	P'
	False negative (FN)	True negative (TN)	N'
Total	P	N	

medical big data processing architecture design, we will use Hadoop and HDFS big data processing framework platforms, based on operator operations and SQL layer processing. In the process of concurrent processing, part of the model processing can be performed in parallel, thereby reducing a lot of time costs.

4.1. Research on Cancer Based on Neural Network. At present, there are more and more researches on cancer diagnosis based on neural network methods. However, the current neural network-based decision-making methods are relatively simple, basically based on disease indicators or according to medical image indicators to do auxiliary diagnosis and lack of sufficient medical data information. The weight of the decision-making stage is unreasonable. Georgia et al. proposed a convolutional neural network- (CNN-) based intelligent diagnosis system for breast cancer diagnosis and related processing, through the related image processing, with images as input, diagnose breast cancer, and classify breast cancer histopathology images [25–27].

Besides, Zhang et al. proposed an artificial neural network for the selection and accurate diagnosis of miRNA biomarkers for colorectal cancer [28] and proposed an artificial neural network (ANN) model. The designed ANN model can accurately classify sample data as cancerous or noncancerous. By using ANN, circulating miRNAs can be used as non-invasive, sensitive, and specific diagnostic markers. Li et al. [29] proposed an expert level for lung cancer detection and classification using the method of deep convolutional neural networks. A three-dimensional convolutional neural network (CNN) was designed to detect pulmonary nodules and classify them as malignant or benign diseases according to pathology and laboratory-confirmed results.

4.2. Research on Cancer Based on Machine Learning. Disease diagnosis tests based on deep learning models are also widely used [30]. The DenseNet neural network model was proposed to classify benign and malignant mammography images, improve the DenseNet neural network model, and invent a new DenseNet-II neural network model. In

Wojciech et al., to predict clinical outcomes by analyzing time-series CT images of patients with locally advanced non-small-cell lung cancer (NSCLC), a deep learning prediction of lung cancer treatment response to a series of medical images was proposed [31].

In addition, Jung-Hoon et al. proposed a new machine learning method for early detection of hepatocellular carcinoma to help doctors solve clinical problems by combining genetic algorithms, support vector machines and feature optimization [32]. Geeitha et al. and Thangamani et al. proposed a support vector machine risk scoring system for ovarian cancer patients [33, 34]. By selecting miRNA sets, support vector machine (SVM) classifiers were constructed to analyze miRNA and clinical factors independently related to prediction, and a risk scoring system was constructed.

4.3. Research on Cancer Based on Fuzzy Reasoning. The diagnosis and probabilistic decision-making problem based on non-small-cell carcinoma (NSCLC) medical system has also become a hot issue. Wu et al. divided the evolution process of disease diagnosis parameters by clinical research and model movement of non-small-cell carcinoma and then established parameters and evaluated the selection stage of non-small-cell lung cancer [35]. Finally, combined with clinical data statistical analysis and probability, the system can provide accurate, rapid clinical data analysis and decision-making recommendations. In another study, [36] proposed a clustering range model based on image recognition and verified the accuracy of treatment by establishing a predictive decision algorithm.

Literature [37] introduced a new type of prostate cancer auxiliary diagnosis model based on fuzzy inference system, combined with statistical analysis of intelligent assisted diagnosis system and big data medical data decision, and finally combined with doctors' diagnosis to provide prostate cancer patients with disease diagnosis and the corresponding treatment plans. In addition, Liu et al. also proposed a breast cancer detection model based on fuzzy reasoning technology, which can determine whether the tumor classification is benign or malignant, and developed a two-layer, high-success rate classifier based on type-2 fuzzy reasoning that combines expert doctors' opinions to classify tumors in the BI-RADS category as benign or malignant [38]. Amin et al. proposed a model that was also based on fuzzy reasoning [22]. However, this study was based on a fuzzy reasoning model based on the change of time grayscale combined with the texture information of cervical images to classify patients with the risk of cervical intraepithelial neoplasia.

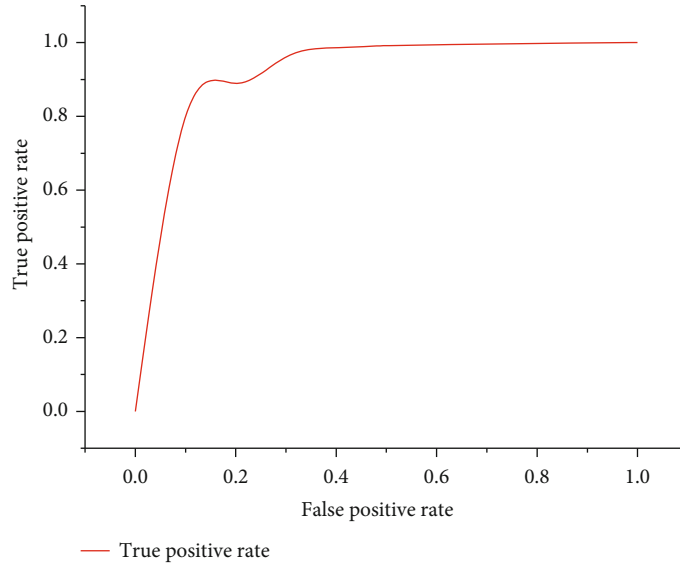


FIGURE 12: Receiver operating curve (ROC) for the classification rate of machine-assisted diagnosis of prostate cancer.

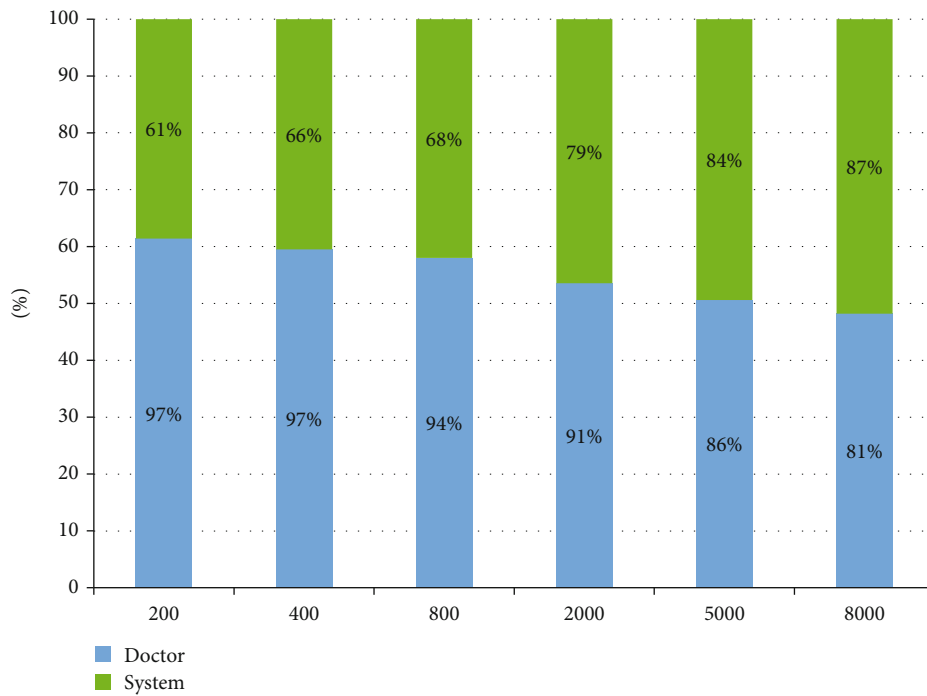


FIGURE 13: Comparison of diagnostic accuracy between doctors and machine-assisted systems.

5. Conclusions

In this study, a machine-aided diagnosis system for prostate cancer based on perceptron neural network and big data was proposed to solve the problem of scarce medical resources caused by the large population and underdeveloped medical level in developing countries. The intelligent system takes the combination of six disease indicators and two medical image indicators as the input and continuously adjusts the weight and deviation of the neural network in the context of medical big data to form an intelligent diagno-

sis model based on the neural network. The model allows doctors to diagnose prostate patients with greater accuracy. Although the intelligent system can provide diagnostic information and decision-making results, it can only serve as an auxiliary diagnosis decision-making system for doctors and cannot completely replace doctors. However, it can largely reduce the work burden of doctors, improve the efficiency of hospitals, and greatly improve the accuracy of doctors' diagnostic. Through the study of hospitalization information of more than 8,000 prostate patients in three hospitals, about 2,156,528 data items were collected and compiled for

experiment. Experimental data shows that when the patient base increases from 200 to 8,000, the accuracy of the machine-assisted diagnostic system will increase from 61% to 87%, and the doctor's diagnostic accuracy will drop to 81%, so the smart medical system can help doctors and medical experts make more effective treatments. From the study, it is concluded that when the patient base reaches a certain number, the diagnostic accuracy of the machine-assisted diagnosis system will exceed the doctor's expert.

In the future work, we will transfer the main work centers to big data and intelligent identification and constantly try new algorithm research, such as SVM and decision tree algorithms, or deep learning algorithms, by comparing the pros and cons between them, continuously optimizing the system model to continuously improve the accuracy of the diagnosis.

Abbreviations

PCA:	Prostate cancer
PET:	Positron emission computed tomography
CT:	Computed tomography
PAP:	Prostatic acid phosphatase
PSMA:	Prostate membrane antigen
HB:	Hemoglobin
TPSA:	Total prostate-specific antigen
RBC:	Red blood cell
FPSA:	Free prostate-specific antigen
MRI:	Magnetic resonance imaging
TNM:	Tumor node metastasis
ROC:	Receiver operating curve
NSCLC:	Non-small cell lung cancer.

Data Availability

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

Disclosure

Jun Zhang, Zhihang Chen, Jia Wu, and Kanghuai Liu current address is at Central South University, Changsha 410075, China.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

J.Z., Z.C., and J.W. conceived the idea of the paper. J.Z., Z.C., J.W., and K.L. designed and performed the experiments; W.Z. and K.L. analyzed the data; Z.C. contributed reagents/materials/analysis tools; J.Z. wrote and revised the paper. Jun Zhang, Zhihang Chen, Jia Wu, and Kanghuai Liu contributed equally to this work.

Acknowledgments

This work was supported partially by "Mobile Health" Ministry of Education-China Mobile Joint Laboratory. This

research was funded by the Major Program of National Natural Science Foundation of China (grant number 71633006) and the National Natural Science Foundation of China (grant number 61672540).

References

- [1] C. D. Mathers, D. Fat, D. M. Fat, M. Inoue, C. Rao, and A. D. Lopez, "Counting the dead and what they died from: an assessment of the global status of cause of death data," *Bulletin of the World Health Organization*, vol. 83, pp. 171–177, 2005.
- [2] F. Cavalli, "Cancer in the developing world: can we avoid the disaster?," *Nature Clinical Practice Oncology*, vol. 3, no. 11, pp. 582–583, 2006.
- [3] P. Vineis and C. P. Wild, "Global cancer patterns: causes and prevention," *The Lancet*, vol. 383, no. 9916, pp. 549–557, 2014.
- [4] C. Dariane, "Active surveillance for prostate cancer," *Progres en urologie: journal de l'Association francaise d'urologie et de la Societe francaise d'urologie*, vol. 25, no. 14, pp. 884–887, 2015.
- [5] A. E. Giuliano, J. L. Connolly, S. B. Edge et al., "Breast Cancer-Major Changes in the American Joint Committee on Cancer Eighth Edition Cancer Staging Manual," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 4, pp. 290–303, 2017.
- [6] D. K. Pauler and D. M. Finkelstein, "Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes," *Statistics in Medicine*, vol. 21, no. 24, pp. 3897–3911, 2002.
- [7] S. Neupane, F. Bray, and A. Auvinen, "National economic and development indicators and international variation in prostate cancer incidence and mortality: an ecological analysis," *World Journal of Urology*, vol. 35, no. 6, pp. 851–858, 2017.
- [8] B. K. Edwards, A. M. Noone, A. B. Mariotto et al., "Annual Report to the Nation on the status of cancer, 1975–2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer," *Cancer*, vol. 120, no. 9, pp. 1290–1314, 2014.
- [9] R. C. Ribeiro and C. H. Pui, "Saving the children improving childhood cancer treatment in developing countries," *New England Journal of Medicine*, vol. 352, no. 21, pp. 2158–2160, 2005.
- [10] A. A. Kotwal and M. A. Schonberg, "Cancer screening in the elderly: a review of breast, colorectal, lung, and prostate cancer screening," *Cancer Journal*, vol. 23, no. 4, pp. 246–253, 2017.
- [11] S. Chakraborty, M. Ghosh, and T. Maiti, "Bayesian neural networks for bivariate binary data: an application to prostate cancer study," *Statistics in Medicine*, vol. 24, no. 23, pp. 3645–3662, 2005.
- [12] M. Wanner, A. Richard, K. Matthes et al., "Trends in prostate cancer incidence between 1996 and 2013 in two Swiss regions by age, grade, and T-stage," *Cancer Causes & Control*, vol. 29, no. 2, pp. 269–277, 2018.
- [13] S. Pilleron, D. Sarfati, M. Janssen-Heijnen et al., "Global cancer incidence in older adults, 2012 and 2035: a population-based study," *International Journal of Cancer*, vol. 144, no. 1, pp. 49–58, 2019.
- [14] C. Proust-Lima, V. Philipps, and J.-F. Dartigues, "A joint model for multiple dynamic processes and clinical endpoints: application to Alzheimer's disease," *Statistics in Medicine*, vol. 5, 2019.

- [15] H. D. Zacho, J. B. Nielsen, A. Afshar-Oromieh et al., "Prospective comparison of 68Ga-PSMA PET/CT, 18F-sodium fluoride PET/CT and diffusion weighted-MRI at for the detection of bone metastases in biochemically recurrent prostate cancer," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 45, no. 11, pp. 1884–1897, 2018.
- [16] N. Hirmas, A. Al-Ibraheem, K. Herrmann et al., "[68Ga]PSMA PET/CT improves initial staging and management plan of patients with high-risk prostate cancer," *Molecular Imaging and Biology*, vol. 21, no. 3, pp. 574–581, 2019.
- [17] C. Öbek, Members of Urooncology Association, T. D. Turkey et al., "The accuracy of 68Ga-PSMA PET/CT in primary lymph node staging in high-risk prostate cancer," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 11, pp. 1806–1812, 2017.
- [18] J. Meng, Y. Liu, S. Guan et al., "The establishment of immune infiltration based novel recurrence predicting nomogram in prostate cancer," *Cancer Medicine*, vol. 20, 2019.
- [19] L. Budäus, S.-R. Leyh-Bannurah, G. Salomon et al., "Initial Experience of 68Ga-PSMA PET/CT Imaging in High-risk Prostate Cancer Patients Prior to Radical Prostatectomy," *European Urology*, vol. 69, no. 3, pp. 393–396, 2016.
- [20] C. Zamboglou, V. Drendel, C. A. Jilg et al., "Comparison of 68Ga-HBED-CC PSMA-PET/CT and multiparametric MRI for gross tumour volume detection in patients with primary prostate cancer based on slice by slice comparison with histopathology," *Theranostics*, vol. 7, no. 1, pp. 228–237, 2017.
- [21] S. Guan and M. Loew, "Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks," in *14th International Workshop on Breast Imaging (IWBI 2018)*, Atlanta, Georgia, United States, 2018.
- [22] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," *Mobile Information Systems*, vol. 2018, 21 pages, 2018.
- [23] A. U. Haq, J. P. Li, M. H. Memon et al., "Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings," *IEEE Access*, vol. 7, pp. 37718–37734, 2019.
- [24] A. U. Haq, J. P. Li, J. Khan et al., "Intelligent Machine Learning Approach for Effective Recognition of Diabetes in E-Healthcare Using Clinical Data," *Sensors*, vol. 20, no. 9, 2020.
- [25] G. V. Ionescu, M. Fergie, M. Berks et al., "Prediction of reader estimates of mammographic density using convolutional neural networks," *Journal of Medical Imaging*, vol. 6, no. 3, 2019.
- [26] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2560–2567, Vancouver, BC, 2016.
- [27] P. AR, L. P. Xue, S. Kalra, K. K. Pillai, R. B. Amin, and Mahul, "Triple negative breast cancer - differentiation of TNBC (ER-, PR-, and HER2/neu-) using new miRNA biomarker panel," 2013.
- [28] C. Zhang, X. Sun, K. Dang et al., "Toward an expert level of lung cancer detection and classification using a deep convolutional neural network," *The Oncologist*, vol. 24, no. 9, pp. 1159–1165, 2019.
- [29] H. Li, S. Zhuang, D.-a. Li, J. Zhao, and Y. Ma, "Benign and malignant classification of mammogram images based on deep learning," *Biomedical Signal Processing and Control*, vol. 51, pp. 347–354, 2019.
- [30] Y. Xu, A. Hosny, R. Zeleznik et al., "Deep learning predicts lung cancer treatment response from serial medical imaging," *Clinical Cancer Research*, vol. 25, no. 11, pp. 3266–3275, 2019.
- [31] W. Książek, M. Abdar, U. R. Acharya, and P. Pławiak, "A novel machine learning approach for early detection of hepatocellular carcinoma patients," *Cognitive Systems Research*, vol. 54, pp. 116–127, 2019.
- [32] J. H. Sung, S. H. Kim, W. I. Yang et al., "miRNA polymorphisms (miR-146a, miR-149, miR-196a2 and miR-499) are associated with the risk of coronary artery disease," *Molecular Medicine Reports*, vol. 14, no. 3, pp. 2328–2342, 2016.
- [33] S. Geeitha and M. Thangamani, "Incorporating EBO-HSIC with SVM for gene selection associated with cervical cancer classification," *Journal of Medical Systems*, vol. 42, no. 11, 2018.
- [34] J. Wu, P. Guan, and Y. Tan, "Diagnosis and data probability decision based on non-small cell lung cancer in medical system," *IEEE Access*, vol. 7, pp. 44851–44861, 2019.
- [35] J. Wu, Y. Tan, and Z. Chen, "Decision based on big data research for non-small cell lung cancer in medical artificial system in developing country," *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 87–101, 2018.
- [36] K. Liu, Z. Chen, J. Wu et al., "Big medical data decision-making intelligent system exploiting fuzzy inference logic for prostate cancer in developing countries," *IEEE Access*, vol. 7, pp. 2348–2363, 2019.
- [37] U. Esma and G. Volkan, "A novel classifier model for mass classification using BI-RADS category in ultrasound images based on Type-2 fuzzy inference system," *Sādhanā*, vol. 43, no. 9, pp. 138–140, 2018.
- [38] J. Liu, Y. Peng, and Y. Zhang, "A fuzzy reasoning model for cervical intraepithelial neoplasia classification using temporal grayscale change and textures of cervical images during acetic acid tests," *Ieee Access*, vol. 7, pp. 13536–13545, 2019.

Research Article

Capacity Evaluation of Diagnostic Tests For COVID-19 Using Multicriteria Decision-Making Techniques

Murat Sayan,^{1,2} Figen Sarigul Yildirim,³ Tamer Sanlidag,^{2,4} Berna Uzun,^{2,5}
Dilber Uzun Ozsahin ^{2,6} and Ilker Ozsahin ^{2,6}

¹Faculty of Medicine, Clinical Laboratory, PCR Unit, Kocaeli University, Kocaeli, Turkey

²DESAM Institute, Near East University, Nicosia/TRNC, Mersin-10, 99138, Turkey

³Health Science University, Antalya Education and Research Hospital, Department of Infectious Diseases and Clinical Microbiology, Antalya 07050, Turkey

⁴Department of Medical Microbiology, Manisa Celal Bayar University, Manisa, Turkey

⁵Department of Mathematics, Near East University, Nicosia/TRNC, Mersin-10, 99138, Turkey

⁶Department of Biomedical Engineering, Faculty of Engineering, Near East University, Nicosia/TRNC, Mersin-10, 99138, Turkey

Correspondence should be addressed to Ilker Ozsahin; ilker.ozsahin@neu.edu.tr

Received 11 May 2020; Revised 22 June 2020; Accepted 7 July 2020; Published 6 August 2020

Guest Editor: Plácido R. Pinheiro

Copyright © 2020 Murat Sayan 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.

In December 2019, cases of pneumonia were detected in Wuhan, China, which were caused by the highly contagious coronavirus. This study is aimed at comparing the confusion regarding the selection of effective diagnostic methods to make a mutual comparison among existing SARS-CoV-2 diagnostic tests and at determining the most effective one. Based on available published evidence and clinical practice, diagnostic tests of coronavirus disease (COVID-19) were evaluated by multi-criteria decision-making (MCDM) methods, namely, fuzzy preference ranking organization method for enrichment evaluation (fuzzy PROMETHEE) and fuzzy technique for order of preference by similarity to ideal solution (fuzzy TOPSIS). Computerized tomography of chest (chest CT), the detection of viral nucleic acid by polymerase chain reaction, cell culture, CoV-19 antigen detection, CoV-19 antibody IgM, CoV-19 antibody IgG, and chest X-ray were evaluated by linguistic fuzzy scale to compare among the diagnostic tests. This scale consists of selected parameters that possessed different weights which were determined by the experts' opinions of the field. The results of our study with both proposed MCDM methods indicated that the most effective diagnosis method of COVID-19 was chest CT. It is interesting to note that the methods that are consistently used in the diagnosis of viral diseases were ranked in second place for the diagnosis of COVID-19. However, each country should use appropriate diagnostic solutions according to its own resources. Our findings also show which diagnostic systems can be used in combination.

1. Introduction

After cases of pneumonia of unknown cause were detected in Wuhan, China, in December 2019, a new coronavirus was isolated from human airway epithelial cells and was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is responsible for coronavirus disease (COVID-19) [1]. SARS-CoV-2 is also a member of the coronavirus family that includes Middle East Respiratory Syndrome- (MERS-) CoV and SARS-CoV, which infect humans [1, 2]. Wild animals are the source of the infection.

According to phylogenetic analysis of full genome sequencing, the coronavirus that causes COVID-19 is a beta-coronavirus in the same subgenus clade as SARS-CoV-2. The structure of the receptor binding site for cell entry is similar and uses the angiotensin-converting enzyme 2 receptor found in the epithelial cells of the alveoli used by SARS-CoV-2 [3]. The International Committee on Taxonomy of Viruses has proposed that this virus be designated as SARS-CoV-2 [3, 4].

The main mode of transmission is via person-to-person spread. When an infected person coughs, sneezes, or speaks,

the virus released in respiratory secretions can infect another person if it comes into direct contact with the mucous membranes through droplet delivery. Also, infection can occur if a person touches an infected surface and then their eyes, nose, or mouth [5]. Infected but asymptomatic people can transmit the virus to others [6].

The most common serious sign of infection is pneumonia: it is characterized by fever, cough, shortness of breath, and bilateral infiltrates in chest imaging [7–9]. In severe cases, patients can quickly experience acute respiratory syndrome, septic shock, metabolic acidosis, and coagulopathy [10, 11]. The mortality rate is 1-2%, but this rate may increase up to 14%, especially in elderly patients with comorbidities such as hypertension, diabetes mellitus, or cardiovascular diseases [10]. Due to the strong infectivity of SARS-CoV-2, it is necessary to identify, isolate, and treat patients as soon as possible, which can reduce mortality rates while reducing the risk of public contamination. In order to be able to treat patients, it is imperative that the disease is diagnosed quickly and accurately. The diagnosis is based on real-time reverse transcription-polymerase chain reaction (rRT-PCR) positivity for the presence of coronavirus [12]. With nucleic acid isolation processing, rRT-PCR results usually require 5 to 6 hours. In addition, it remains unclear whether rRT-PCR is the gold standard and whether false-positive or false-negative results are common. The Centers for Disease Control and Prevention recommends that nasopharyngeal and oropharyngeal swab specimens should be collected to test for SARS-CoV-2 [13]. Although false-positive tests are generally possible, a positive test for SARS-CoV-2 confirms the diagnosis of COVID-19. False-negative results can be obtained from the upper respiratory samples, so if the first test is negative and the patient continues to suspect COVID-19, it is recommended that the test be repeated [14]. Serological tests, on the other hand, can be accessed and evaluated more easily and can identify patients with existing or previous infections but who have negative rRT-PCR tests [15]. These tests can also be used because they are easier to access in places that do not have access to the rRT-PCR test. The technique of the test is quite simple. It gives results in a short time [16]. Although samples taken from the patient are not evaluated for viral culture, they do have diagnostic value [17].

The other methods that are used in the diagnosis of COVID-19 are imaging methods. The diagnostic sensitivity of viral pneumonia by chest radiography is relatively low [18], whereas computerized tomography (CT) has high sensitivity for diagnosis of COVID-19 which makes it a primary tool for COVID-19 detection in epidemic areas [19]. In COVID-19 patients, the sensitivity of chest CT scan is 97% but its specificity is 25% with rRT-PCR results as reference, which is why it is not used as a primary screening test [19].

Chest CT scan abnormalities have also been identified in patients prior to the development of symptoms and even prior to the detection of viral RNA from upper respiratory specimens [20, 21]. In the article published by Huang et al. in the *Lancet*, it was stated that it is necessary to have signs of infection and/or a positive nucleic acid test for chest CT scan to be used in diagnosis; otherwise, CT should not be

recommended for screening or early diagnosis [22]. They said that CT should not be recommended for screening or early diagnosis. However, when the rRT-PCR results of patients were considered to be clinically and epidemiologically negative for COVID-19, chest CT scan became more valuable than rRT-PCR in the early stage of the disease. Shi et al. showed that combining the evaluation of chest CT scan imaging features with clinical and laboratory findings may facilitate the early diagnosis of COVID-19 pneumonia [20, 23].

Thus far, seven different diagnostic tests have been used for COVID-19: rRT-PCR, cell culture, CoV-19 antigen detection, serological tests (CoV-19 antibody IgM, CoV-19 antibody IgG), chest X-ray, and chest CT. From the onset of the pandemic, real-life experience has shown that rRT-PCR and chest CT are the most preferred methods for the diagnosis of COVID-19. The suitability of SARS-CoV-2 diagnostic tests can be prioritized with the parameters of high sensitivity, high specificity, low false positivity, low false negativity, high usability, low cost, etc. In this study, we used the fuzzy preference ranking organization method for enrichment evaluation (fuzzy PROMETHEE) and fuzzy technique for order of preference by similarity to ideal solution (fuzzy TOPSIS) techniques to compare the confusion regarding the selection of effective diagnostic methods to make a mutual comparison between the aforementioned seven SARS-CoV-2 diagnostic tests and hence determine the best one. We have decided to use fuzzy PROMETHEE because fuzzy part can handle fuzzy data very well, and PROMETHEE part can handle the data when there are too many parameters to be set properly. One other advantage of using PROMETHEE is that it is a user-friendly outranking method and successfully adapted to real-life problems. Fuzzy TOPSIS, on the other hand, is simple, rational, and comprehensive, as well as able to measure relative performance of alternatives.

2. Materials and Methods

Fuzzy PROMETHEE and fuzzy TOPSIS techniques, which are widely used by researchers as MCDM methods, were applied for the evaluation of the SARS-CoV-2 diagnostic tests in order to provide the best alternatives. The fuzzy PROMETHEE technique is a hybrid model that provides decision analysis for nonnumerical data and also enables decision-makers to define vague conditions mathematically [24]. The PROMETHEE technique was developed by Brans et al. [25, 26], and it is aimed at giving a comprehensive ranking for a finite set of alternatives corresponding to their criteria and the importance level of each criterion. By providing various preference functions, including V-shaped function, level function, Gaussian function, U-shaped function, and linear function, to the criteria for the comparison of the alternatives, the PROMETHEE technique differs from the other MCDM methods and can provide more sensitive results. The PROMETHEE and TOPSIS methods are unable to analyze fuzzy data (linguistic or vague data) in the actual decision-making environment. Thus, we have applied the fuzzy-based PROMETHEE and TOPSIS techniques, which

are the hybrid methods that use a fuzzy scale. The fuzzy scale supports the decision-maker in decision analysis if the data is within a range, vague, or qualitative.

In this study, a linguistic fuzzy scale has been used to define the parameters of the SARS-CoV-2 diagnostic tests. These values were converted to triangular fuzzy numbers, and then the Yager index was applied for defuzzification of the data. Then, the PROMETHEE decision lab program with a Gaussian preference function was applied for the determination of the net ranking result.

The PROMETHEE method comprises 5 steps that are applied for the MCDM analysis and to rank the alternatives as follows [25–28]:

- (i) The preference function P_j of each criterion j should be defined
- (ii) Importance weights of each criterion $w_j = (w_1, w_2, \dots, w_K)$, where $(j = 1, 2, \dots, K)$ should be defined as

$$\sum_{j=1}^K w_j = 1, \quad (1)$$

where K is the number of criteria.

- (iii) For each of the alternative pairs $a_t, a_{t'} \in A$, the outranking relation $\pi(a_t, a_{t'})$ should be determined by

$$\pi(a_t, a_{t'}) = \sum_{j=1}^K w_j \cdot \left[P_j \left(f_j(a_t) - f_j(a_{t'}) \right) \right], \quad AXA \longrightarrow [0, 1], \quad (2)$$

where $f_j(a_t)$ denotes the value of the j^{th} criterion of the alternative a_t and $\pi(a_t, a_{t'})$ denotes the preference indices, which show the preference intensity for an alternative a_t in comparison to an alternative $a_{t'}$ while counting all criteria simultaneously.

- (iv) The positive outranking flow and negative outranking flow should be determined as follows:

- (a) A positive outranking flow of the alternative a_t :

$$\Phi^+(a_t) = \frac{1}{n-1} \sum_{\substack{t'=1 \\ t' \neq t}}^n \pi(a_t, a_{t'}). \quad (3)$$

- (b) A negative outranking flow of the alternative a_t :

$$\Phi^-(a_t) = \frac{1}{n-1} \sum_{\substack{t'=1 \\ t' \neq t}}^n \pi(a_{t'}, a_t). \quad (4)$$

n denotes the number of alternatives. The $\Phi^+(a_t)$ defines the strength of alternative $a_t \in A$,

while the negative outranking flow $\Phi^-(a_t)$ defines the weakness of alternative $a_t \in A$.

PROMETHEE I determines a partial preorder of the alternatives based on the positive and negative outranking flows, and PROMETHEE II determines a complete preorder of the alternatives based on a net flow. The partial preorder of the options can be determined based on the following statements:

Via PROMETHEE I, alternative a_t is selected to alternative $a_{t'} (a_t Pa_{t'})$ if it satisfies either of the statements given below as

$$\begin{cases} \Phi^+(a_t) \geq \Phi^+(a_{t'}) \text{ and } \Phi^-(a_t) < \Phi^-(a_{t'}), \\ \Phi^+(a_t) > \Phi^+(a_{t'}) \text{ and } \Phi^-(a_t) = \Phi^-(a_{t'}). \end{cases} \quad (5)$$

a_t is indifferent to alternative $a_{t'} (a_t Ia_{t'})$ if

$$\Phi^+(a_t) = \Phi^+(a_{t'}) \text{ and } \Phi^-(a_t) = \Phi^-(a_{t'}). \quad (6)$$

And a_t is incomparable to $a_{t'} (a_t Ra_{t'})$ if

$$\begin{cases} \Phi^+(a_t) > \Phi^+(a_{t'}) \text{ and } \Phi^-(a_t) > \Phi^-(a_{t'}), \\ \Phi^+(a_t) < \Phi^+(a_{t'}) \text{ and } \Phi^-(a_t) < \Phi^-(a_{t'}). \end{cases} \quad (7)$$

The net outranking flow can be calculated for each alternative by using the following:

$$\Phi^{\text{net}}(a_t) = \Phi^+(a_t) - \Phi^-(a_t). \quad (8)$$

Via PROMETHEE II, the complete order with net flow can be determined as

$$\begin{aligned} a_t \text{ is preferred to } a_{t'} (a_t Pa_{t'}) & \quad \text{if } \Phi^{\text{net}}(a_t) > \Phi^{\text{net}}(a_{t'}), \\ a_t \text{ is indifferent to } a_{t'} (a_t Ia_{t'}) & \quad \text{if } \Phi^{\text{net}}(a_t) = \Phi^{\text{net}}(a_{t'}). \end{aligned} \quad (9)$$

The higher $\Phi^{\text{net}}(a_t)$ value provides the better alternative.

In this study, triangular fuzzy scale was used for defining the linguistic data of the criteria of SARS-CoV-2 diagnostic tests as shown in Table 1. Zadeh [29] has defined the fuzzy sets in order to define the linguistic or vague data mathematically based on the membership degree. Triangular fuzzy sets can be represented with 3 specific points (a , b , and c) where a is the initial point (left bound), b is the peak point of the triangle which contain the maximum membership degree, and c is the right bound of the triangle. The linguistic fuzzy scale indicates the linguistic terms of very high, high, medium, and low with their associated fuzzy sets. Furthermore, the importance level of each has been defined with the triangular fuzzy scale by the experts for analyzing the effectiveness criteria of SARS-CoV-2 diagnostic tests. The criteria selected for the analysis include cost, accessibility, sensitivity, and specificity.

For the validation of the fuzzy-based PROMETHEE output, fuzzy TOPSIS, a different type of MCDM technique, has been used for the same dataset. TOPSIS method was first

TABLE 1: Linguistic fuzzy scale.

Linguistic scale for evaluation	Triangular fuzzy scale	Importance ratings of criteria
VH	(0.75, 1, 1)	No symptom-window period, early stage of infection, active phase of infection, late of recurrent stage of infection, cost, accessibility, false positivity, false negativity
H	(0.50, 0.75, 1)	Sensitivity, specificity
M	(0.25, 0.50, 0.75)	Past infection, usability, equipment, education
L	(0, 0.25, 0.50)	Recovery stage of infection
VL	(0, 0, 0.25)	

VH: very high; H: high; M: medium; L: low; VL: very low.

proposed by Yoon and Hwang [30] and commonly applied for the decision-making problems of the conflicting criteria [31, 32] in wide disciplines. It compares a set of alternatives by identifying weights for each criterion, normalizing scores for each criterion and calculating the geometric distance between each alternative and the ideal alternative, which is the best score in each criterion. It based on the concept that the chosen alternative should have the shortest geometric distance from the positive ideal solution (PIS) and the longest geometric distance from the negative ideal solution (NIS).

There are 6 steps of the TOPSIS method to be applied for the MCDM analysis and rank the alternatives as follows [31].

- (1) The decision matrix which includes the parameters (criteria) of the alternatives and the importance weight of the criteria, $j(w_j)$, should be defined as

$$\sum_{j=1}^K w_j = 1, \quad (10)$$

where K is the number of criteria.

- (2) The normalized decision matrix (n_{ij}) should be calculated by one of the equations below:

$$n_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^m X_{ij}^2}},$$

$$n_{ij} = \frac{X_{ij}}{\max_i X_{ij}},$$

$$n_{ij} = \begin{cases} \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}, \\ \frac{\max_i x_{ij} - x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}, \end{cases} \quad (11)$$

where x_{ij} is the value of the i^{th} alternative and j^{th} criterion.

TABLE 2: Positive and negative ideal solution sets.

Criteria	Positive ideal solution	Negative ideal solution
No symptom-window period	0.473	0.158
Early stage of infection	0.473	0.158
Active phase of infection	0.473	0.158
Late of recurrent stage of infection	0.473	0.158
Past infection	0.257	0.086
Recovery stage of infection	0.129	0.043
Cost	0.086	0.257
Accessibility	0.473	0.158
Usability	0.257	0.086
Equipment	0.257	0.086
Education	0.257	0.086
Sensitivity	0.386	0.129
Specificity	0.386	0.129
False positivity	0.158	0.473
False negativity	0.158	0.473

- (3) The values of the weighted normalized matrix (v_{ij}) should be calculated by applying the following:

$$v_{ij} = w_j n_{ij}. \quad (12)$$

- (4) And positive ideal solution (A^+) and negative ideal solution (A^-) should be determined by using Equation (13) and Equation (14), respectively

$$A^+ = (v_1^+, v_2^+, \dots, v_n^+) = \left[\left[\max_i v_{ij} | j \in I \right], \left[\min_i v_{ij} | j \in J \right] \right], \quad (13)$$

$$A^- = (v_1^-, v_2^-, \dots, v_n^-) = \left[\left[\min_i v_{ij} | j \in I \right], \left[\max_i v_{ij} | j \in J \right] \right], \quad (14)$$

where I and J are associated with benefit criteria and cost criteria, for $i = 1, \dots, m$; $j = 1, \dots, n$.

Positive ideal solution (A^+) is the set of the values that maximizes the benefit criteria while minimizing the cost criteria. Negative ideal solution (A^-) can be considered an opposite of the positive ideal solution.

TABLE 3: Weighted normalized data of the SARS-CoV-2 diagnostic tests.

Criteria	Importance weight	Max/Min	Naso/oropharyngeal swab PCR	Cell culture	CoV-19 antigen detection	CoV-19 antibody IgM	CoV-19 antibody IgG	Chest X-ray	Chest CT
No symptom-window period	VH	Max	0.47	0.47	0.47	0.16	0.16	0.16	0.32
Early stage	VH	Max	0.47	0.47	0.47	0.47	0.16	0.16	0.47
Active phase	VH	Max	0.47	0.47	0.47	0.47	0.47	0.16	0.47
Recurrent stage	VH	Max	0.16	0.16	0.16	0.16	0.47	0.32	0.47
Past infection	M	Max	0.09	0.09	0.09	0.09	0.26	0.17	0.26
Recovery stage	L	Max	0.04	0.04	0.04	0.09	0.13	0.04	0.13
Cost	M	Min	0.09	0.09	0.26	0.09	0.09	0.09	0.09
Accessibility	VH	Max	0.47	0.16	0.16	0.47	0.47	0.47	0.47
Usability	M	Max	0.26	0.09	0.26	0.26	0.26	0.26	0.26
Equipment	M	Max	0.17	0.09	0.26	0.09	0.09	0.09	0.26
Education	M	Max	0.17	0.09	0.09	0.26	0.26	0.17	0.26
Sensitivity	H	Max	0.13	0.39	0.13	0.13	0.13	0.26	0.39
Specificity	H	Max	0.39	0.39	0.13	0.26	0.26	0.26	0.39
False positivity	VH	Min	0.16	0.16	0.47	0.47	0.47	0.47	0.47
False negativity	VH	Min	0.47	0.16	0.47	0.47	0.47	0.47	0.47

VH: very high; H: high; M: medium; L: low.

In the negative ideal solution, the benefit criterion is minimized while the cost criterion is maximized. According to the TOPSIS method, the most suitable option is the one which is closer to the positive ideal solution and further to the negative ideal solution [33].

- (5) Then, the separation measures between alternatives and the positive ideal solution should be calculated by using Equation (15) and the separation measures between alternatives and the negative ideal solution should be calculated by using Equation (16) in order to obtain the distance between the i^{th} alternative and the positive ideal solution (d_i^+) and the distance between the i^{th} alternative and the negative ideal solution (d_i^-) as shown below:

$$d_i^+ = \left(\sum_{j=1}^n (v_{ij} - v_j^+)^p \right)^{1/p}, \quad i = 1, 2, \dots, m, \quad (15)$$

$$d_i^- = \left(\sum_{j=1}^n (v_{ij} - v_j^-)^p \right)^{1/p}, \quad i = 1, 2, \dots, m, \quad (16)$$

where $p \geq 1$.

- (6) Then, the relative closeness to the positive ideal solution (R_i) should be calculated for every alternative by using Equation (17) as seen below:

$$R_i = \frac{d_i^-}{d_i^- + d_i^+}, \quad (17)$$

where $0 \leq R_i \leq 1, i = 1, 2, \dots, m$.

Positive and negative ideal solution sets of the SARS-CoV-2 diagnostic tests have been obtained as shown in Table 2 below.

In TOPSIS technique, relative closeness to the positive ideal solution (R_i) determines the ranking results of the alternatives. The higher (R_i) is a more preferred option. The weighted normalized dataset of the SARS-CoV-2 diagnostic tests can be seen in Table 3.

3. Results and Discussion

Table 4 indicates the results of the complete ranking for the SARS-CoV-2 diagnostic tests, where each alternative pair is compared numerically based on each criterion and their importance weights. The power of each alternative can be thought numerically as the positive outranking flow, while the weakness of the alternatives can be thought as the negative outranking flow. However, the net flow gives the net ranking results. The most effective alternative is the one with the higher net flow. With the maximum positive outranking flow and minimum negative outranking flow, the chest CT is the best available SARS-CoV-2 diagnostic test, followed by naso/oropharyngeal swab rRT-PCR. Even if naso/oropharyngeal swab rRT-PCR has a lower positive flow than cell culture, it also has lower negative flow than the cell culture. Therefore, the second-best option is naso/oropharyngeal swab rRT-PCR with 0.0258 net flow, and the third one is the cell culture with the 0.0098 net flow. The last two options are CoV-19 antigen detection and chest X-ray as seen in Table 4.

The strength and the weakness of the available SARS-CoV-2 diagnostic tests are shown in Figure 1. The criteria for each SARS-CoV-2 diagnostic tests are listed

TABLE 4: Complete ranking of SARS-CoV-2 diagnostic tests with fuzzy PROMETHEE.

Complete ranking	Diagnostic tests	Positive outranking flow (ϕ^+)	Negative outranking flow (ϕ^-)	Net flow (ϕ^{net})
1	Chest CT	0.0666	0.0109	0.0557
2	Naso/oropharyngeal swab PCR	0.0478	0.0220	0.0258
3	Cell culture	0.0635	0.0537	0.0098
4	CoV-19 antibody IgG	0.0384	0.0429	-0.0045
5	CoV-19 antibody IgM	0.0228	0.0391	-0.0163
6	CoV-19 antigen detection	0.0277	0.0543	-0.0267
7	Chest X-ray	0.0176	0.0615	-0.0439

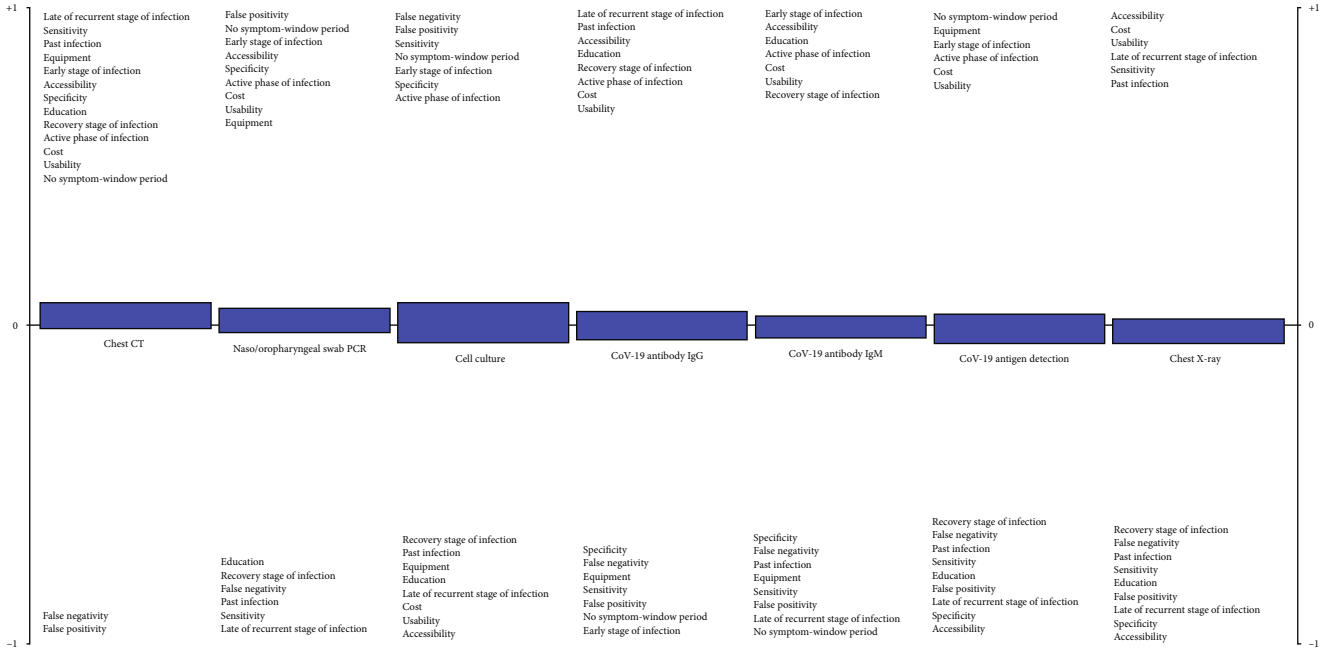


FIGURE 1: Positive and negative aspects of each technique obtained by fuzzy PROMETHEE.

above or below the zero-threshold level based on their effect on the performance of the alternatives. If the criteria are in advantage of the technique, they are listed above, and if the criteria are in disadvantage of the technique, they are listed below the zero-threshold level. As represented, the most preferred technique is the chest CT, with almost all the criteria listed above the threshold level, while the last effective technique for the SARS-CoV-2 diagnosis is the chest X-ray, with most of the criteria listed above the threshold level. These results can be useful for hospitals, doctors, governments, and patients or anyone involved in the decision-making processes of the SARS-CoV-2 diagnosis.

The SARS-CoV-2 diagnostic tests ranking results obtained with fuzzy TOPSIS have been presented in Table 5. With the 0.6403 value of the closeness to the ideal solution (R_i), chest CT is the best alternative, naso/oropharyngeal swab rRT-PCR is the second best alternative with 0.5865 R_i , and the cell culture with 0.5718 R_i is the third best option between the alternatives. The CoV-19 antigen detection and chest X-ray were found as the least effective options.

These results are consistently similar with the fuzzy-based PROMETHEE outputs. Hence, the results from both

TABLE 5: The relative closeness to positive ideal solution with the ranking of the SARS-CoV-2 diagnostic tests.

Ranking	Alternatives	R_i
1	Chest CT	0.6403
2	Naso/oropharyngeal swab PCR	0.5865
3	Cell culture	0.5718
4	CoV-19 antibody IgG	0.4815
5	CoV-19 antibody IgM	0.4638
6	CoV-19 antigen detection	0.4460
7	Chest X-ray	0.3815

techniques could give evidence to the decision-makers about the effectiveness of the available SARS-CoV-2 diagnostic tests.

According to the findings of a study conducted with more than 1,000 patients by Ai et al. [19], CT should be the primary diagnosis tool used to screen for COVID-19. However, their finding of 97% sensitivity for the chest CT was based on positive rRT-PCR test results. Chest CT scan for COVID-19 diagnosis can be a good complement to rRT-PCR [34].

According to the study of Fang et al. [34], the sensitivity of CT for COVID-19 was higher compared to rRT-PCR (98% vs. 71%, respectively). Similarly, we also showed mathematically that CT should be the first preferred diagnostic test. Furthermore, differentiating COVID-19 from non-COVID-19 pneumonia on chest CT, which is a relatively harder task, was studied in [35], and the authors reported that the findings of CT are considered nonspecific despite its high sensitivity. In addition, one of the greatest issues faced is the need for an expert radiologist to interpret the radiography images. This procedure is even more difficult when attempting to differentiate COVID-19 from other types of pneumonia since the visual indicators can be subtle [35].

Therefore, the main limitation of our study is the lack of analysis of the combination of different diagnostic tests. Hence, in future studies, once the available techniques for diagnosing COVID-19 are more robust, we will apply the MCDM methods to the combination of different diagnostic tests. In addition, it is sensible to apply the MCDM methods for specific countries, regions, or even hospitals considering specific conditions such as their available diagnostic toolkits, scanners, and experts.

4. Conclusions

This study analyzed the seven selected available COVID-19 diagnostic tests using the fuzzy-based MCDM methods. The evaluations of the effectiveness of the alternatives were made based on the selected criteria, and their corresponding weights were determined by the experts. As a result of the evaluation, chest CT was found to be the most effective diagnostic test. Moreover, the results indicated that the top three most effective COVID-19 diagnostic tests are chest CT, rRT-PCR, and cell culture. On the other hand, chest X-ray is the last effective diagnostic test. It is interesting to note that the methods that are always used in the diagnosis of viral diseases were ranked second for the diagnosis of COVID-19.

In this study, during the process of evaluating the effectiveness of the given COVID-19 diagnostic tests, each technique was considered and compared based on their general properties collected from the latest available guidelines. The use of fuzzy-based MCDM methods can be also adapted based on the patient's specific condition.

Data Availability

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

Conflicts of Interest

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

References

- [1] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, "The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2," *Nature Microbiology*, vol. 5, no. 4, pp. 536–544, 2020.
- [2] N. Zhu, D. Zhang, W. Wang et al., "A novel coronavirus from patients with pneumonia in China, 2019," *The New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020.
- [3] F. A. Rabi, M. S. Al Zoubi, G. A. Kasasbeh, D. M. Salameh, and A. D. Al-Nasser, "SARS-CoV-2 and coronavirus disease 2019: what we know so far," *Pathogens*, vol. 9, no. 3, p. 231, 2020.
- [4] A. E. Gorbalenya, S. C. Baker, R. S. Baric et al., *Severe acute respiratory syndrome-related coronavirus: the species and its viruses – a statement of the Coronavirus Study Group*, bioRxiv, 2020.
- [5] World Health Organization, *Novel coronavirus situation reports -2*, 2020, March 2020, <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200122-sitrep-2-2019-ncov.pdf>.
- [6] D. Chang, M. Lin, L. Wei et al., "Epidemiologic and clinical characteristics of novel coronavirus infections involving 13 patients outside Wuhan, China," *JAMA*, vol. 323, no. 11, article 1092, 2020.
- [7] W. J. Guan, Z. Y. Ni, Y. Hu et al., "Clinical characteristics of coronavirus disease 2019 in China," *New England Journal of Medicine*, vol. 382, no. 18, pp. 1708–1720, 2020.
- [8] C. Huang, Y. Wang, X. Li et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.
- [9] N. Chen, M. Zhou, X. Dong et al., "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," *The Lancet*, vol. 395, no. 10223, pp. 507–513, 2020.
- [10] Y. Yi, P. N. P. Lagniton, S. Ye, E. Li, and R. H. Xu, "COVID-19: what has been learned and to be learned about the novel coronavirus disease," *International Journal of Biological Sciences*, vol. 16, no. 10, pp. 1753–1766, 2020.
- [11] L. L. Ren, Y. M. Wang, Z. Q. Wu et al., "Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study," *Chinese Medical Journal*, vol. 133, no. 9, pp. 1015–1024, 2020.
- [12] V. M. Corman, O. Landt, M. Kaiser et al., "Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR," *Euro-surveillance*, vol. 25, no. 3, 2020.
- [13] A. Patel, D. B. Jernigan, 2019-nCoV CDC Response Team et al., "Initial Public Health Response and Interim Clinical Guidance for the 2019 Novel Coronavirus Outbreak — United States, December 31, 2019–February 4, 2020," *Morbidity and Mortality Weekly Report*, vol. 69, no. 5, pp. 140–146, 2020.
- [14] World Health Organization, "Coronavirus disease (COVID-19) technical guidance: surveillance and case definitions," March 2020, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/surveillance-and-case-definitions>.
- [15] L. Guo, L. Ren, S. Yang et al., "Profiling early humoral response to diagnose novel coronavirus disease (COVID-19)," *Clinical Infectious Diseases*, 2020.
- [16] I. Cassaniti, F. Novazzi, F. Giardina et al., "Performance of VivaDiag COVID-19 IgM/IgG rapid test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department," *Journal of Medical Virology*, 2020.
- [17] W. B. Park, N. J. Kwon, S. J. Choi et al., "Virus isolation from the first patient with SARS-CoV-2 in Korea," *Journal of Korean Medical Science*, vol. 35, no. 7, article e84, 2020.

- [18] T. Franquet, "Imaging of pulmonary viral pneumonia," *Radiology*, vol. 260, no. 1, pp. 18–39, 2011.
- [19] T. Ai, Z. Yang, H. Hou et al., "Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," *Radiology*, 2020.
- [20] H. Shi, X. Han, N. Jiang et al., "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study," *The Lancet Infectious Diseases*, vol. 20, no. 4, pp. 425–434, 2020.
- [21] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing," *Radiology*, 2020.
- [22] Y. Huang, W. Cheng, N. Zhao, H. Qu, and J. Tian, "CT screening for early diagnosis of SARS-CoV-2 infection," *The Lancet Infectious Diseases*, 2020.
- [23] C. Long, H. Xu, Q. Shen et al., "Diagnosis of the coronavirus disease (COVID-19): rRT-PCR or CT?," *European Journal of Radiology*, vol. 126, article 108961, 2020.
- [24] D. Uzun Ozsahin and I. Ozsahin, "A fuzzy PROMETHEE approach for breast Cancer treatment techniques," *International Journal of Medical Research & Health Sciences*, vol. 7, no. 5, pp. 29–32, 2018.
- [25] J. P. Brans, B. Mareschal, and P. Vincke, "PROMETHEE: a new family of outranking methods in MCDM," *Oper Res IFORS 84*, J. P. Brans, Ed., pp. 477–490, 1984.
- [26] J. P. Brans and P. Vincke, "Note—a preference ranking organisation method," *Management Science*, vol. 31, no. 6, pp. 647–656, 1985.
- [27] J. P. Brans, P. Vincke, and B. Mareschal, "How to select and how to rank projects: the PROMETHEE method," *European Journal of Operational Research*, vol. 24, no. 2, pp. 228–238, 1986.
- [28] J. Geldermann, T. Spengler, and O. Rentz, "Fuzzy outranking for environmental assessment. Case study: iron and steel making industry," *Fuzzy Sets and Systems*, vol. 115, no. 1, pp. 45–65, 2000.
- [29] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [30] C. L. Yoon and K. Hwang, *Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, New York, NY, USA, 1981.
- [31] J. Huang, "Combining entropy weight and TOPSIS method for information system selection," *2008 IEEE International Conference on Automation and Logistics*, 2008, pp. 1281–1284, Qingdao, China, September 2008.
- [32] F. Ye and Y. Li, "An extended TOPSIS model based on the possibility theory under fuzzy environment," *Knowledge-Based Systems*, vol. 67, pp. 263–269, 2014.
- [33] C.-R. Wu, C. T. Lin, and P. H. Tsai, "Financial service of wealth management banking: balanced scorecard approach," *Journal of Social Sciences*, vol. 4, no. 4, pp. 255–263, 2008.
- [34] Y. Fang, H. Zhang, J. Xie et al., "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," *Radiology*, 2020.
- [35] H. X. Bai, B. Hsieh, Z. Xiong et al., "Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT," *Radiology*, 2020.

Research Article

Applying Queuing Theory and Mixed Integer Programming to Blood Center Nursing Schedules of a Large Hospital in China

Li Luo,¹ Xiaofei Liu ,¹ Xinyuan Cui,¹ Yuanjun Cheng ,¹ Xinzhu Yu,¹ Yue Li,¹ Li Jiang,² and Mingying Tan³

¹Business School, Sichuan University, Chengdu, China

²Sichuan Provincial People's Hospital, Chengdu, China

³West China Hospital, Chengdu, China

Correspondence should be addressed to Yuanjun Cheng; tomcyj@scu.edu.cn

Received 25 March 2020; Revised 20 May 2020; Accepted 27 May 2020; Published 1 July 2020

Guest Editor: Mirian C. D. Pinheiro

Copyright © 2020 Li 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.

Blood centers in large hospitals in China are facing serious problems, including complex patient queues and inflexible nursing schedules. This study is aimed at developing a flexible scheduling method for blood center nurses. By systematically analyzing the constraints that affect scheduling, a flexible scheduling model is established based on queuing theory and mixed integer programming. This combined model can reasonably determine the number of nurses required during a given working period and flexibly arrange nursing schedules while ensuring sufficient rest periods for individual nurses. Results of numerical studies conducted using data from a large hospital in China show a significant improvement in patient waiting time performance metrics over the hospital's current practice. In addition, the nurses' workloads and rest periods are well balanced, indicating that the proposed method can effectively and flexibly arrange nursing shifts in blood centers.

1. Introduction

Healthcare resources in China are currently under pressure because of high demand and low supply, with 20% of the world's population having to cope with no more than 3% of global healthcare resources [1]. Human staff is a vital medical resource; in particular, nursing staff considerably impact clinical outcomes. Nurse scheduling is presently the most demanding scheduling problem affecting hospital personnel on a daily basis [2]. Many hospital nursing managers still rely on manual scheduling to create nurses' shift patterns. Although experienced managers can easily perform this task in small hospitals, it can become a very difficult and time-consuming task in large hospitals. In addition, fairly distributing the workload among nurses can directly affect their job satisfaction.

The number of patients arriving at different times in the blood center substantially fluctuates. Such fluctuations often cause two unfortunate situations: (1) the service's peak capacity is insufficient to meet the demand, causing long

queues, and (2) the service's capacity during quiet periods is much higher than the demand, wasting human resources. In addition, heavy workloads, high pressure, and lack of necessary rest periods make it easy for blood collection nurses to get stressed and make mistakes, further affecting the quality of service. Although many service industries have introduced part-time workers to ease work pressure during peak hours, hospitals around the world rarely employ such staff [3]. We introduce part-time nurses to ease the stress on regular nurses during peak hours. In addition, we have reduced the number of nurses working in low hours needed to avoid waste of resources.

The most important aspect of blood center nurse scheduling is to accurately calculate the demand for nurses. How to properly determine the number of nurses who need to work for providing the most convenient medical services for patients? The operational research model offers a systematic approach to problem solving and allows for the characterisation of activities of an existing system using mathematical modelling. Saville et al. believed that the

introduction of operational research techniques into mainstream nurse research can enhance decision-making on nurse staffing and discussed the application of these techniques in the medical field [4]. One approach, based on mathematical models, that successfully addresses problems in healthcare systems is the use of queueing models [5]. Izady and Worthington proposed an iterative scheme that uses an infinite server network, square root staffing, and simulation to set the minimum medical staffing levels required to achieve government goals [6]. Jahn et al. used queueing theory to simulate patient arrivals and assess the resulting economic effects, proving that queueing theory produced better results [7]. According to the fact that the patients' arrival rate is dynamic and random, Liu et al. and Lin et al. used the M/M/C queueing theory to estimate the patient's waiting time in the system of dynamic demand and then constructed the mixed integer programming model for EMR scheduling to obtain a flexible shifting scheme [8, 9]. Chen and Ping used a queueing theory model to study the queueing system and its effects on a hospital's blood collection hall and determined the optimal number of nursing staff required to meet the needs of both patients and the staff themselves [10].

Although some hospitals may use queueing theory to create nursing schedules, this theory only calculates the number of nurses required to prevent patients' queuing for a long time and does not consider the broader issues involved in creating suitable nursing schedules. The scheduling problem of medical staff is a combinatorial optimization problem. Manual scheduling is tedious and time-consuming; therefore, the nurse rostering problem (NRP), also known as the nurse scheduling problem (NSP), has attracted considerable research attention.

In these research studies, some soft and hard constraints are considered, including the policies of the state, rules of the hospital, and personal needs of doctors and nurses [11]. The 0-1 integer programming model, mixed integer programming model, and the goal programming model are constructed, and the exact algorithm and heuristic algorithm are used to solve them [12, 13]. In order to obtain the scientific and reasonable scheduling shifts, Tan et al. consider various management rules in a hospital, physicians' personal preferences, and the time requirements of their personal learning and living and take the minimum deviation variables from the soft constraints as the objective function to construct a mixed integer programming model with the doctor group as the scheduling unit [14]. In addition, some scholars use heuristic algorithms to solve the problem of nurses' scheduling. Wu et al. proposed a particle swarm optimization method to solve highly complex NSPs to generate nursing schedules that meet all requirements and consider fairness [15]. Lin designed an adaptive scheduling heuristic algorithm that reduces the patients' waiting times [16]. Zhong et al. proposed a two-stage heuristic algorithm for nurse scheduling to achieve fairness and flexibility goals [17].

This study is aimed at nurse scheduling models for accommodating different patient needs as well as for satisfying the working rules and regulations, thereby generating nursing rosters that adhere to the hospital fairness require-

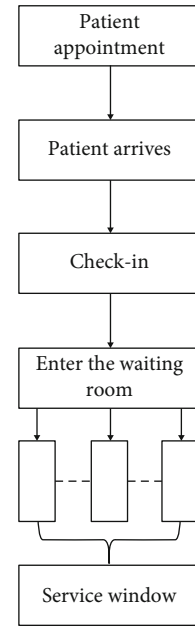


FIGURE 1: Patient consultation flow chart.

ments. Then, we introduce part-time nurses to ease the stress on regular nurses during peak hours. A queueing model is used herein to calculate the labor demands for each period during the day. This model calculates the minimum staffing levels required for each period to meet the hospital's needs. Based on these minimum staffing requirements, a mixed integer programming (MIP) model is proposed to determine the most effective nursing roster. The proposed model is numerically investigated to validate the model's feasibility.

2. Materials and Methods

2.1. Study Hospital. Our survey of a hospital in Chengdu revealed that the blood center, an important outpatient service, has always been unable to satisfy the increasing demand because of the increasing number of patients; this results in long queues and affects patient satisfaction. Patients wait for an average of 35 min and maximum up to 67 min. Because of the long waiting time, patients have to simultaneously wait for testing their blood.

As we all know, each nurse has his own role in a department to complete a series of operations. However, our research is limited to a blood collection center. In the blood collection center, the nurse has only one operation for blood collection. Some things like supporting employees and preparing records are all done by smart devices. Figures 1 and 2 provide the flow chart of patient consultation and the flow chart of doctor service, respectively. Patient appointments can also be made through online appointments or offline machines. When the patient registers, the patient only needs to swipe the medical card to record all the information. As for the nurse's job, only blood collection and labeling of blood containers are required. In addition, the labeling process is assisted by intelligent machines.

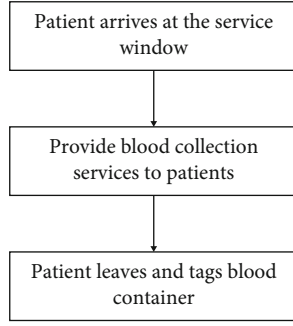


FIGURE 2: Doctor service flow chart.

2.2. Datasets. Our models are based on a dataset that includes 119302 patient blood collection records collected from hospital H's blood center between October 9, 2017, and January 31, 2018. Herein, work begins at 7:00 a.m. and finishes at 5:00 p.m. After processing and analyzing these data using R, we calculated the patient arrival rates and average service time. Then, we used a goodness-of-fit test to determine that the patient arrival rates obeyed the Poisson distribution.

2.3. Methods. The proposed models involve two stages. First, we calculate the most reasonable number of staff by a queuing model; then, we reasonably arrange their schedules by a mixed integer programming model.

2.3.1. Queuing Model. The patient arrival rates are subject to the Poisson distribution; therefore, queuing theory can be used to calculate the minimum number of staff required to meet the demand. Specifically, we propose the following solution to find the most reasonable number of windows to open during each time shift.

The blood center's service quality requirements state that at least 95% of patients should leave the service within 15 min. We use queuing theory to calculate the optimal number of open windows and establish a queuing model. We use the following notation:

- c : number of open windows in the blood center
- λ : patient arrival rate
- μ : average service rate
- ρ : service intensity
- L_q : number of patients in the queue
- W_q : average patient waiting time
- t_q : hospital's required maximum average patient waiting time
- t : hospital's required maximum average time that patients stay in the center
- p_q : probability of achieving the average waiting time requirement
- P : probability of achieving the average staying time requirement
- L : average number of patients in the queuing system
- W : average time patients stay in the blood center
- u : average utilization rate of the blood center
- \bar{c} : average number of busy windows

The number of open windows in the blood center must be a positive integer, and this value cannot be analytically

obtained. Therefore, we designed the following algorithm to calculate the value of the decision variable.

Step 1. For the queuing process to reach a steady state, $c = \lceil \lambda/\mu \rceil$ (an integer not exceeding $\lceil \lambda/\mu \rceil$).

Step 2. If $W_q \leq 15$ min, then terminate and return to c as the minimum number of open windows required; otherwise, go to Step 3.

Step 3. $c = c + 1$ and go to Step 2.

The queuing model is represented as follows:

$$\min c, \quad (1)$$

subject to

$$P\{W_q \leq t_q\} = 1 - P\{W_q > t_q\} = 1 - P_q > p_q m, \quad (2)$$

$$P\{W \leq t\} = 1 - P\{W > t\} = 1 - P \geq p, \quad (3)$$

$$c \in \mathbb{Z}^+, \quad (4)$$

$$P_q = P\{W_q > t_q\} = (1 - P\{W_q = 0\})e^{-c\mu(1-\rho)t_q}, \quad (5)$$

$$P\{W_q = 0\} = \sum_0^{c-1} p_n, \quad (6)$$

$$P = P\{W > t\} = e^{-\mu t} \left[1 + \frac{p_0(\lambda/\mu)^c}{c!(1 - (\lambda/c\mu))} \left(\frac{1 - e^{-\mu t(c-1-(\lambda/\mu))}}{c-1-(\lambda/\mu)} \right) \right], \quad (7)$$

$$p_0 = \left[\sum_{n=0}^{c-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^c}{c!} \left(\frac{1}{1 - (\lambda/c\mu)} \right) \right]^{-1}, \quad \text{if } \frac{\lambda}{c\mu} < 1, \quad (8)$$

$$p_0 = \begin{cases} \frac{(\lambda/\mu)^n}{n!} p_0, & 0 \leq n \leq c, \\ \frac{(\lambda/\mu)^n}{c! e^{-(\lambda/c\mu)}} p_0, & n \geq c, \end{cases} \quad (9)$$

$$L_q = \frac{\rho(\lambda/\mu)^c}{c!(1-\rho)^2} p_0, \quad \rho = \frac{\lambda}{c\mu}, \quad (10)$$

$$L = L_q + \frac{\lambda}{\mu}, \quad (11)$$

$$W_q = \frac{L_q}{\lambda}, \quad (12)$$

$$W = \frac{L}{\lambda}, \quad (13)$$

$$c_- = L - L_q, \quad (14)$$

$$u = \frac{\lambda}{c\mu}. \quad (15)$$

Herein, the objective function (1) minimizes the number of open windows, constraint (2) represents the waiting time

constraint, constraint (3) represents the staying time constraint, constraint (4) guarantees that c is an integer, constraints (5)–(9) calculate the number of open windows, and constraints (10)–(15) are various indicators used for evaluation.

2.3.2. Nurse Scheduling Model. The above queuing model can be used to calculate the number of blood centers that need to be opened during each period of the day. Results provide the number of staff members required. Based on the results, an MIP model is established to generate daily rosters for the blood collection nurses. Creating flexible schedules for hospital blood collection nurses involves two main factors: the nurses have to receive sufficient rest periods and their working hours should be approximately the same.

The hospital introduces part-time nurses to alleviate the stress during peak hours and cover absences and lunch time. Most peak working hours are in the morning; therefore, these part-time nurses can only work during two time shifts in the morning. In actual scheduling scenarios, the need to combine many constraints with the number of blood collection nurses required during each time shift dramatically increases the complexity of the problem, making manual scheduling difficult to achieve.

We use the following additional notation:

i : full-time blood collection nurse identifier

j : part-time blood collection nurse identifier

d : day number in the roster period

t : period number in the day

c_{dt} : minimum number of nurses required for time shift t on day d (obtained using the above queuing model)

nt : maximum number of time shifts for which a nurse can continuously work every day

nd_{\min} and nd_{\max} : minimum and maximum number of days, respectively, for which a full-time nurse can work continuously during a given roster period

z_{id} : work status of a full-time nurse i on day d

m_{jd} : work status of a part-time nurse j on day d

x_{idt} : indicates whether a full-time nurse i works during period t on day d

y_{jdt} : indicates whether a part-time nurse j works during period t on day d

nj : maximum number of part-time nurses working on a given day

$sumd_{\min}$ and $sumd_{\max}$: minimum and maximum number of days worked by a full-time nurse during one roster period, respectively

The nurse scheduling model is represented as follows:

$$\min \left[z_{id} + \left(\sum_{i=1}^I x_{idt} + \sum_{j=1}^J y_{jdt} \right) \right], \quad (16)$$

subject to

$$\sum_{i=1}^I x_{idt} + \sum_{j=1}^J y_{jdt} \geq c_{dt}, \quad \forall d, \forall t, \quad (17)$$

$$-n \leq \sum_{d=1}^D \sum_{t=1}^T x_{i_1 dt} - \sum_{d=1}^D \sum_{t=1}^T x_{i_2 dt} \leq n, \quad \forall i_1, \forall i_2, i_1 \neq i_2, \quad (18)$$

$$\sum_{t=k}^{k+nt} x_{idt} \leq nt, \quad \forall d, \forall i, \forall k = 1, 2, \dots, T - nt, \quad (19)$$

$$\sum_{t=k}^{k+nt} y_{jdt} \leq nt, \quad \forall d, \forall j, \forall k = 1, 2, \dots, T - nt, \quad (20)$$

$$nd_{\min} \leq \sum_{d=k}^{k+nd} z_{id} \leq nd_{\max}, \quad \forall i, \forall k = 1, 2, \dots, D - nd, \quad (21)$$

$$M \cdot mz_{id} \geq \sum_{t=1}^T x_{jdt}, \quad \forall i, \forall d, \quad (22)$$

$$M \cdot (z_{id} - 1) + 1 \leq \sum_{t=1}^T x_{jdt}, \quad \forall i, \forall d, \quad (23)$$

$$y_{idt} = 0, \quad \forall j, \forall d, t = T_j, \quad (24)$$

$$sumd_{\min} \leq \sum_{d=1}^D z_{id} \leq sumd_{\max}, \quad \forall i, \quad (25)$$

$$\sum_{j=1}^J m_{jd} \leq nj, \quad \forall d, \quad (26)$$

$$M \cdot m_{id} \geq \sum_{t=1}^T y_{jdt}, \quad \forall j, \forall d, \quad (27)$$

$$M \cdot (m_{id} - 1) + 1 \leq \sum_{t=1}^T y_{jdt}, \quad \forall j, \forall d, \quad (28)$$

$$x_{idt} = c(0, 1), \quad \forall i, \forall d, \forall t, \quad (29)$$

$$y_{jdt} = c(0, 1), \quad \forall j, \forall d, \forall t, \quad (29)$$

$$z_{id} = c(0, 1), \quad \forall i, \forall d, \quad (29)$$

$$y_{jd} = c(0, 1), \quad \forall j, \forall d. \quad (29)$$

Here, the objective function (16) minimizes the number of days worked by full-time nurses given that the total number of hours worked by all nurses is as small as possible. Constraint (17) indicates that a sufficient number of nurses must work daily to meet the patients' needs. Constraint (18) indicates that there should not be a considerable difference between the work periods of different full-time nurses in the scheduling cycle to ensure fairness. Constraints (19) and (20) limit the maximum number of consecutive hours for which the nurses work per day, ensuring that they have sufficient rest to maintain efficiency. Constraint (21) enforces the maximum and minimum number of consecutive working days for each full-time nurse. Constraints (22) and (23) ensure that the logical relations between the decision variables are satisfied, where M is a sufficiently large constant. Constraint (24) ensures that no part-time nurses are working during the given time shift. Constraint (25) enforces the

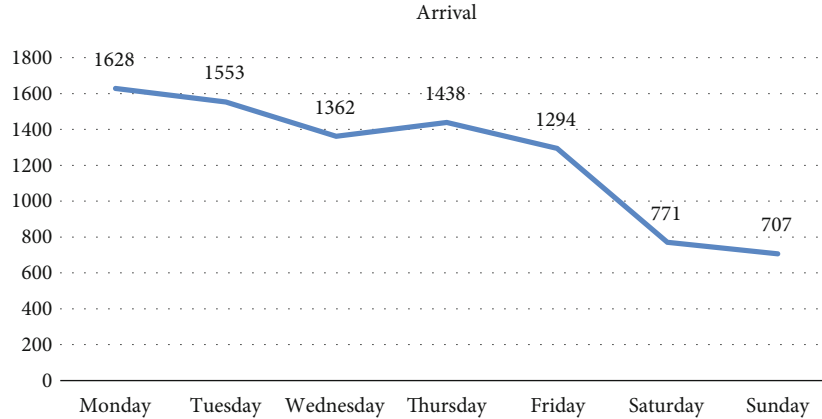


FIGURE 3: Patients arriving on different days.

TABLE 1: Patient arrival rates (people/min) for different time shifts and days.

Days	7 a.m.–10 a.m.	10 a.m.–1 p.m.	1 p.m.–5 p.m.
Monday, Tuesday	5.23	2.16	0.79
Wednesday, Thursday, Friday	4.56	1.61	0.67
Saturday, Sunday	2.45	0.89	0.42

minimum and maximum number of days nurses can work during the planning period to ensure reasonable working hours and rest periods. Constraint (26) limits the number of part-time nurses who can work on a given day. Constraints (27) and (28) ensure that the logical relations between decision variables are satisfied, where M is a sufficiently large constant. Finally, constraint (29) constrains the decision variables to binary values.

3. Results

3.1. Data Analysis. Figure 3 shows that different numbers of patients arrive on different days of the week, with peak numbers on Monday and Tuesday and relatively few on Saturday and Sunday. Based on these numbers, we can divide the week into three parts: Monday and Tuesday; Wednesday, Thursday, and Friday; and Saturday and Sunday. Table 1 shows that the patient arrival rate substantially fluctuates during the day. Hospitals typically prefer to roster a fixed number of staff to work, but such large demand fluctuations can easily lead to one of the two situations: either the staff struggles to meet the demand during peak periods, causing long queues, or they are assigned less tasks during quiet periods when the demand is low, resulting in overstaffing and wasted manpower.

3.2. Solution to the Models. The above queuing model is an M/M/C queuing system, which is used herein to solve using R. The proposed algorithm and model are verified using a practical example. This study assumes that patients arrive in the hospital’s blood collection hall according to the Pois-

son process and that the service time is subject to a negative exponential distribution with an average rate of 0.64 minute/person. Based on the hospital requirements quoted above, the example’s other main parameters are set to $t_q = 15$ min and $P_q = 0.95$. Finally, the results of the queuing theory model are calculated using R.

Table 2 shows the minimum number of open blood collection windows during different time shifts. For example, opening nine windows on Monday and Tuesday between 7 a.m. and 10 a.m. is sufficient to meet the demand. The queuing model calculates the probability of any patient’s queuing time exceeding the given time limit with a certain number of staff. Most patients have to wait less than 15 min, which is a substantial improvement over the previous situation.

After using the queuing model to calculate the nurses required during each time shift, the following data about hospital A’s blood collection nurses were obtained. A total of 15 full-time blood collection nurses and five part-time nurses are routinely available. Each nurse can work for at least three consecutive days, but no more than five days, and can work for up to three time shifts per day. Part-time nurses can only work in the first and second time shifts. Rosters are prepared for four-week periods. Full-time nurses work at least 20 days but less than 24 days. Using these data, we solved the nurse scheduling model.

Clearly, this is an MIP problem and feasible; approximately, optimal solutions are acceptable for nurse scheduling. Here, we used the mature CPLEX solver to solve this MIP problem. By combining the calculated number of blood collection windows that must be opened during each time shift with this example model, CPLEX obtained the most optimal scheduling plan.

According to the optimal solution, a total of 300 person-days are required during the four-week planning period. Table 3 shows part of the corresponding scheduling plan, including full-time and part-time nurses working during each time shift. For example, during the first period of the first day, full-time nurses 1, 4, 5, 6, 8, 10, 11, and 17 and a part-time nurse must work. Table 4 shows the days on which each nurse is working during the first week of the 28-day shift

TABLE 2: Predicted minimum number of open blood collection windows.

Date	λ	Period	c.min	W_q	L_q	P_q	W	L	P	c_-	u
Monday	5.23	1	9	1.33	6.98	0	2.9	15.15	0	8.17	0.91
	2.16	2	4	1.69	3.66	0	3.26	7.03	0	3.38	0.84
	0.79	3	2	0.96	0.76	0	2.52	1.99	0	1.23	0.62
Wednesday	4.56	1	8	1.21	5.51	0	2.77	12.63	0	7.12	0.89
	1.61	2	3	2.29	3.69	0.01	3.86	6.21	0.01	2.52	0.84
	0.67	3	2	0.59	0.4	0	2.15	1.44	0	1.05	0.52
Saturday	2.45	1	5	0.65	1.6	0	2.22	5.43	0	3.83	0.77
	0.89	2	2	1.46	1.3	0	3.02	2.69	0	1.39	0.7
	0.42	3	1	2.98	1.25	0.02	4.55	1.91	0.04	0.66	0.66

TABLE 3: Blood collection nurses working during different time shifts.

Date	Period	Nurses	Date	Period	Nurses
1	1	1, 4, 5, 6, 8, 10, 11, 17, 20	15	1	1, 2, 6, 8, 10, 11, 12, 16, 19
1	2	1, 5, 6, 8	15	2	1, 10, 14, 15
1	3	10, 11, 13	15	3	2, 6, 11
2	1	1, 3, 4, 6, 9, 11, 12, 14, 15	16	1	1, 3, 4, 5, 6, 7, 8, 9, 12, 15
2	2	1, 2, 7, 8	16	2	3, 5, 10, 13, 14

TABLE 4: Days that each blood collection nurse works.

Nurse	1	2	3	4	5	6	7	8	9	10
	1	2	2	1	1	1	2	1	2	1
	2	3	3	2	2	2	3	2	3	3
Days worked	5	4	4	4	4	3	5	3	4	4
	6	5	5	5	5	4	6	4	5	5
	7	8	7	6	6	7	7	5	6	7

period. For example, nurse 1 works on days 1, 2, 5, 6, and 7 during the first week. Tables 3 and 4 show only partial results; for more detailed results, please see Supplementary Materials S1 and S2.

Figure 4 describes each nurse’s average workload over the entire planning period (total number of hours worked/total number of time shifts in the planning period) according to the solution. Here, we can see that the nurses’ workloads do not considerably fluctuate, implying that the plan is fair.

4. Discussion

Based on a systematic analysis of the factors affecting the scheduling of nurses in large hospitals, a flexible scheduling model for blood collection nurses is established using queuing theory and MIP. This combined model can reasonably determine the number of nurses required during particular periods and flexibly arrange rosters while ensuring that they receive reasonable rest periods. The final results of the model denote that the patient’s waiting time is effectively alleviated and that a nurse’s rest time is ensured.

The large number of constraints in this example and the complexity of the decision variables imply that manual scheduling cannot find optimal solutions. In contrast, the

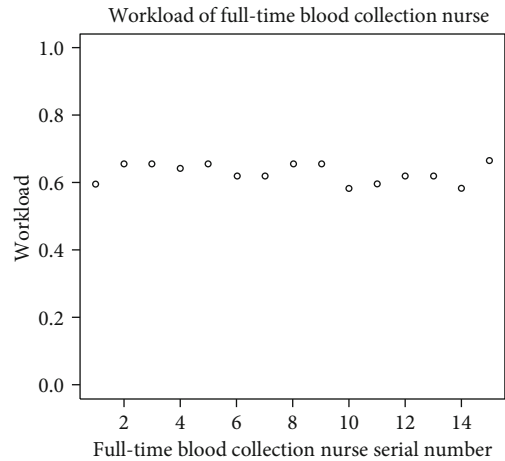


FIGURE 4: Workloads of the full-time nurses over the planning cycle.

proposed model can rapidly implement nurse scheduling. This flexible scheduling approach can adjust the number of nurses working based on dynamic changes in demand. During periods of low demand, fewer staff is scheduled to work, whereas part-time nurses can be introduced during peak periods to spread the workload. Adapting the model to different hospitals only requires changes to the input data and parameters, enabling managers to easily grasp and promote it. However, several issues pertaining to the proposed model must be addressed. For example, we did not consider an approach to minimizing the hospital’s costs when employing part-time nurses or to handling the nurses’ leave periods.

In summary, the proposed flexible scheduling model provides an effective solution for rostering nurses in blood centers. In China, hospitals, supermarkets, banks, and

government offices struggle with long queues and desire flexible scheduling. The proposed model that is based on queuing theory and MIP could be a flexible scheduling model for such tasks.

5. Conclusions

Queuing and nurse scheduling models were proposed herein to solve the issue of rostering nurses. First, the queuing model determined the minimum labor demand during each period of each day, which served as the primary input for the scheduling model. This model was then used to determine the nurses' shift schedules and organize the full-time nurses' shifts to ensure fairness. Results of numerical studies conducted using data from a large hospital in China show a significant improvement in patient waiting time performance metrics over the hospital's current practice. The proposed model could generate reasonable rosters to handle relatively large fluctuations in patient numbers. In summary, we proposed methods for making decisions about staff numbers and scheduling that can improve work efficiency.

Data Availability

The data used to support the findings of this study are restricted by Sichuan Provincial People's Hospital in order to protect patient privacy. Data are available from Sichuan Provincial People's Hospital for researchers who meet the criteria for access to confidential data.

Conflicts of Interest

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

Acknowledgments

The authors gratefully acknowledge the support from Sichuan Provincial People's Hospital. In addition, we would like to thank Becton, Dickinson and Company (BD) for their support of some blood collection medical equipment. This research was supported in part by the National Natural Science Foundation of China (Nos. 71532007 and 71131006) and Sichuan Province Science and Technology Support Project Plan (Nos. 2016FZ0080 and 2019YFS0385).

Supplementary Materials

S1 and S2 take up too much space; only part of the results is shown in the manuscript (Table 3 and Table 4). Table 3 shows part of the corresponding scheduling plan, including full-time and part-time nurses working during each time shift. Table 4 shows the days on which each nurse is working during the first week of the 28-day shift period. (*Supplementary materials*)

References

[1] Y. Chen, Z. Yin, and Q. Xie, "Suggestions to ameliorate the inequity in urban/rural allocation of healthcare resources in

China," *International Journal for Equity in Health*, vol. 13, no. 1, p. 34, 2014.

- [2] L. H. Aiken, "Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction," *JAMA*, vol. 288, no. 16, pp. 1987–1993, 2002.
- [3] L. Y. N. N. N. JAMIESON, L. E. O. N. I. E. M. O. S. E. L. WILLIAMS, W. I. L. L. I. A. M. LAUDER, and T. R. U. D. Y. DWYER, "The 'realities' of part-time nursing: a grounded theory study," *Journal of Nursing Management*, vol. 16, no. 7, pp. 883–892, 2008.
- [4] C. E. Saville, P. Griffiths, J. E. Ball, and T. Monks, "How many nurses do we need? A review and discussion of operational research techniques applied to nurse staffing," *International Journal of Nursing Studies*, vol. 97, pp. 7–13, 2019.
- [5] C. Lakshmi and S. A. Iyer, "Application of queuing theory in health care: A literature review," *Operations Research for Health Care*, vol. 2, no. 1-2, pp. 25–39, 2013.
- [6] N. Izady and D. Worthington, "Setting staffing requirements for time dependent queueing networks: the case of accident and emergency departments," *European Journal of Operational Research*, vol. 219, no. 3, pp. 531–540, 2012.
- [7] B. Jahn, E. Theurl, U. Siebert, and K. P. Pfeiffer, "Tutorial in medical decision modeling incorporating waiting lines and queues using discrete event simulation," *Value in Health*, vol. 13, no. 4, pp. 501–506, 2010.
- [8] Q. Liu et al., *Shifts Scheduling Method for Physicians in Emergency Department with Time-Varying Demand*, 2015.
- [9] L. D. Lin et al., *Physician Weekly Staffing for Emergency Departments with Time-Varying Demand*, 2017.
- [10] Y. Chen and L. I. Ping, "Application of queuing theory model in nurse staffing in blood collection rooms," *Journal of Nursing Administration*, 2012.
- [11] R. Stolletz and J. O. Brunner, "Fair optimization of fortnightly physician schedules with flexible shifts," *European Journal of Operational Research*, vol. 219, no. 3, pp. 622–629, 2012.
- [12] E. K. Burke, P. de Causmaecker, G. V. Berghe, and H. van Landeghem, "The state of the art of nurse rostering," *Journal of Scheduling*, vol. 7, no. 6, pp. 441–499, 2004.
- [13] M. Erhard, J. Schoenfelder, A. Fügner, and J. O. Brunner, "State of the art in physician scheduling," *European Journal of Operational Research*, vol. 265, no. 1, pp. 1–18, 2018.
- [14] M. Tan, J. Gan, and Q. Ren, "Scheduling emergency physicians based on a multiobjective programming approach: a case study of West China Hospital of Sichuan University," *Journal of Healthcare Engineering*, vol. 2019, 9 pages, 2019.
- [15] T.-H. Wu, J.-Y. Yeh, and Y.-M. Lee, "A particle swarm optimization approach with refinement procedure for nurse rostering problem," *Computers & Operations Research*, vol. 54, pp. 52–63, 2015.
- [16] C. K. Y. Lin, "An adaptive scheduling heuristic with memory for the block appointment system of an outpatient specialty clinic," *International Journal of Production Research*, vol. 53, no. 24, pp. 7488–7516, 2015.
- [17] X. Zhong, J. Zhang, and X. Zhang, "A two-stage heuristic algorithm for the nurse scheduling problem with fairness objective on weekend workload under different shift designs," *IIEE Transactions on Healthcare Systems Engineering*, vol. 7, no. 4, pp. 224–235, 2017.

Research Article

Auxiliary Medical Decision System for Prostate Cancer Based on Ensemble Method

Jia Wu ^{1,2}, Qinghe Zhuang,^{1,2} and Yanlin Tan ^{2,3}

¹School of Computer Science and Engineering, Central South University, Changsha 410083, China

²“Mobile Health” Ministry of Education-China Mobile Joint Laboratory, Changsha 410083, China

³PET-CT Center, The Second Xiangya Hospital of Central South University, Changsha 410083, China

Correspondence should be addressed to Jia Wu; jiawu5110@163.com and Yanlin Tan; tanyanlin@csu.edu.cn

Received 3 March 2020; Accepted 27 April 2020; Published 18 May 2020

Guest Editor: Plácido R. Pinheiro

Copyright © 2020 Jia Wu 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.

Prostate cancer (PCa) is one of the main diseases that endanger men’s health worldwide. In developing countries, due to the large number of patients and the lack of medical resources, there is a big conflict between doctors and patients. To solve this problem, an auxiliary medical decision system for prostate cancer was constructed. The system used six relevant tumor markers as the input features and employed classical machine learning models (support vector machine and artificial neural network). Stacking method aimed at different ensemble models together was used for the reduction of overfitting. 1,933,535 patient information items had been collected from three first-class hospitals in the past five years to train the model. The result showed that the auxiliary medical system could make use of massive data. Its performance is continuously improved as the amount of data increases. Based on the system and collected data, statistics on the incidence of prostate cancer in the past five years were carried out. In the end, influence of diet habit and genetic inheritance for prostate cancer was analyzed. Results revealed the increasing prevalence of PCa and great negative impact caused by high-fat diet and genetic inheritance.

1. Introduction

In 2018, morbidity and mortality of PCa accounted for 13.5% and 6.7%, respectively, in male patients. In 185 countries around the world, PCa has the highest morbidity in 105 countries and the highest mortality in 46 countries [1]. Undoubtedly, PCa has become one of the main threats to men’s health worldwide. Countries with high HDI (human development index) usually have high morbidity (68.0/100,000). Although countries with low HDI have relatively low morbidity (14.5/100,000), there is an obvious trend of growth and it increases fast [2]. Take China for example; in 1998, the rough morbidity was 3.25/100,000; however, in 2012, it increased to 8.14/100,000. By 2013, it has become 8.58/100,000 [3, 4]. The high morbidity in developed countries and the increasing incidence rate in developing countries have led to a huge number of prostate cancer patients worldwide.

In developing countries that lack medical resources, many patients cannot receive timely and effective diagnosis

and therapy, which will aggregate the conflict between doctors and patients. In China, there are only 2.59 practitioners for every 1000 people [5]. In Beijing, a small number of high-level hospital medical staff members have to provide medical services to more than 20 million people in Beijing, and many cancer patients in other regions also come here for the extraordinary medical treatment. Medical staff members have been overloaded for a long time [6–9]. This will inevitably have an impact on the accuracy of the diagnosis, thus leading to serious consequences.

Other developing countries may face similar problems:

- (i) Due to the large number of patients and less medical resources, it is difficult for patients to get timely and effective diagnosis and treatment
- (ii) The long-term workload of doctors reduces the efficiency and accuracy of diagnosis
- (iii) Many hospitals have poor medical equipment, which further increases the probability of misdiagnosis

- (iv) The per capita income in developing countries is pretty low, and most people cannot afford expensive but accurate diagnostic methods like PET-CT

These problems can be alleviated by building an auxiliary medical decision system. By analyzing a large number of data, the auxiliary medical decision system can learn a diagnostic model. When a new patient comes, it will provide doctors with suggestions relevant to diagnosis or treatment based on the learned model [10, 11]. Combing suggestions from the system and their own knowledge, doctors will give the final conclusion of diagnosis and treatment method. The auxiliary medical decision system can relieve the doctor's burden to some extent, thus alleviating the conflict between doctors and patients [12, 13]. In this work, we constructed an auxiliary medical decision system which can determine whether a patient has prostate cancer, judge the clinical stage, recommend treatment options, and evaluate the effectiveness of treatment options. Given the low income in developing countries, six tumor markers are selected with relatively low testing price and high relevance to PCa as the input features. Classical machine learning techniques and ensemble method are adopted to extract the knowledge inside data and improve performance.

The main contributions and innovations of this research include the following:

- (i) Appropriate features are selected for the construction of medical systems according to national conditions of developing countries
- (ii) The constructed auxiliary system can give treatment plan and evaluate its effectiveness
- (iii) The use of constructed ensemble method by a secondary learner improved the accuracy
- (iv) The system is trained based on a large amount of patient information from three high-level hospitals in China, and some factors affecting PCa via the constructed system are analyzed

The main structure of this article is as follows: the first part describes the background and contribution of the research, the second part introduces related research, the third part is a detailed description to the construction of the model, the fourth part is about the training process and analysis to experiments' results, and the fifth part is the conclusion.

2. Related Works

Medical diagnosis of cancer is usually a gradual transition which starts from simple, cheap, and harmless but with low-accuracy methods and ends with expensive and accurate methods. Compared with commonly used CT, MRI, PET-CT, and other methods, detection of tumor markers is a relatively basic and cheap method in the diagnosis of cancer, which makes constructing the auxiliary medical decision system with tumor markers suitable for developing countries with low capital medical expenditure, low medical level,

and poor medical facilities. Diagnosing cancer with a single tumor marker usually does not have good sensitivity and specificity [14]. Therefore, many related research studies [14–17] combine different tumor markers or biomarkers to predict some diseases and have acquired good results. Specifically, literature [15] detected serum levels of 17 tumor markers for 145 patients with pancreatic cancer and selected 9 tumor markers by backward elimination selection, scatter plots, and relative operating characteristic analysis. Based on these features, the pancreatic cancer determination system CAMPAS-P was established. The final results showed that the CAMPAS-P system was able to accurately distinguish malignant pancreatic cancer from benign pancreatobiliary disease but performed bad on the diagnosis of the unusual histologic type of pancreatic tumors and various digestive organ malignancies. Literature [17] used serum microRNA biomarkers to predict nonalcoholic fatty liver disease (NAFLD). Among all the selected biomarkers, nine of them were associated with NAFLD severity, and some of them appeared specific to NAFLD. These biomarkers showed good classification performance for nonalcoholic steatohepatitis (NASH). Literature [14] combined growth-related tumor markers and associated tumor markers for the diagnosis of cancer and acquired 80–90% sensitivity, 84–85% specificity, and 83–88% accuracy.

As one of the classic machine learning algorithms, the support vector machine (SVM) [18] shows good performance in many classification problems before the revival of deep learning. It is also widely used in medical diagnosis [19–23]. Literature [19] extracted features from mammograms by Hough transform and classified mammograms by SVM. Its accuracy reached 94% while other machine learning methods like linear discriminant analysis just had 86% accuracy. Literature [20] employed various machine learning techniques for the prediction of breast cancer in Wisconsin Breast Cancer (original) datasets. After considering accuracy, sensitivity, specificity, and precision, SVM got the best results. Literature [21] proposed a classification fuzzy-rough set with the SVM model featured as CA-125 and other amino acids to detect early-stage ovarian cancer. It performed quick learning and had good classification performance.

Ensemble learning technology is also widely used in medical assistant diagnosis. Literature [24] selected demographic, physiological, and vital signs and laboratory tests as features and built different models, finding that the ensemble learning random forest is the most effective in mortality prediction in the early hours of an ICU patient admission. The proposed EMPICU-RF framework based on the ensemble learning random forest outperformed many standard scoring systems in terms of AUROC (area under the curve) and time. Literature [25] proposed a model that combines the physicians' knowledge in the form of a rule-based classifier and supervises learning algorithms to detect asthma control level. Literature [23] constructed two different ensemble models by confidence-weighted voting method and the boosting ensemble technique for the diagnosis of breast cancer. The proposed CWV-BANN-SVM model reached the accuracy of 100%.

TABLE 1: Commonly used kernel functions.

Kernel functions	Formula
Linear kernel	$\kappa(x_i, x_j) = x_i^T x_j$
Polynomial kernel	$\kappa(x_i, x_j) = (x_i^T x_j)^d$
Gauss kernel	$\kappa(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$
Sigmoid kernel	$\kappa(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ }{\sigma}\right)$
Laplace kernel	$\kappa(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta)$

3. Design of the Auxiliary Medical Decision-Making System

3.1. Requirements and Framework of the System. The auxiliary medical decision system is aimed at offering some help to doctors. Its functions cover diagnosing the patients, staging the cancer, recommending the treatment plan, and evaluating the treatment plan. Diagnosing patients is to tell if the tumor is malignant or benign. Staging the cancer is to determine the clinical stage (I, II, III, or IV) for those diagnosed with malignant PCa. The above two requirements can be satisfied by building a classification model using the machine learning method. On the other hand, in order to give a cancer treatment plan and evaluate its efficacy at the same time, the whole problem is considered a regression problem. The system will finally output a value evaluating the malignancy of PCa, abbreviated as EM value. The larger the value, the higher the malignancy. If the value does not decrease after executing a certain treatment plan, it means that the treatment plan is not effective and another treatment plan needs to be selected. Meanwhile, the auxiliary medical decision-making system needs to have good parallelism and be able to process multiple patients' simultaneous diagnosis requests. It is worth noting that after the medical system is invested, the amount of data obtained will gradually increase over time. The decision model will be retrained to further improve the generalization performance.

3.2. Design of the Decision Model

3.2.1. Introduction to the Support Vector Machine. The support vector machine can acquire the global optimal solution in high-dimensional problems; thus, it is widely used in many situations [26].

For a linear separable binary classification problem, assume that the input dataset $S = \{x_1, x_2, x_3, \dots, x_m\}$ and the output label $y = \{y_1, y_2, y_3, \dots, y_m\}$, where x_i is the input vector of the i th sample and $y_i \in \{-1, 1\}$ is the corresponding label of x_i . SVM aims to find a hyperplane $H : W^T x + b = 0$ that separates the positive and negative samples and meanwhile maximizes their distance to the hyperplane. The optimizing process can be expressed as follows:

$$\begin{cases} \min_{w,b} \frac{1}{2\|W\|^2}, \\ y_i(W^T x + b) \geq 1, \quad i = 1, 2, \dots, m. \end{cases} \quad (1)$$

In order to compute the solution efficiently, the Lagrange multiplier is introduced, and according to the Wolfe duality theory, it is changed into an equivalent dual problem:

$$\text{s.t. } \sum_{i=1}^m y_i \alpha_i = 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \dots, m. \quad (2)$$

As for the linear inseparable problem, penalty parameter C and slack variable ξ_i are introduced, and the problem is expressed as follows:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m y_i y_j \alpha_i \alpha_j x_i^T x_j - \sum_{j=1}^n \alpha_j \\ \text{s.t.} \quad & \sum_{j=1}^m y_j \alpha_j = 0 \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, m, \end{aligned} \quad (3)$$

where C represents the interval of two classes, and the final decision function is as follows:

$$y(x) = \text{sgn}(f(x)) = \text{sgn} \left\{ \left(\sum_{i=1}^m \alpha_i^* y_i (x_i^T \cdot x) \right) + b^* \right\}. \quad (4)$$

For the nonlinear problem, kernel function $\varphi(x_i)$ that maps the nonlinear problem in low dimension into a linear problem in high dimension is introduced. The kernel functions defined in the input sample space satisfy the Mercer condition and have the following expression: $K(x_i, x_j) = (\varphi(x_i), \varphi(x_j)) = \varphi(x_i)^T \varphi(x_j)$. Commonly used kernel functions are listed in Table 1.

A typical support vector machine is usually used to deal with binary classification problems. In this medical decision system, SVM is firstly used to classify benign (labeled 1) and malignant (labeled -1) tumors. In order to stage possible malignant tumors, a four-class (I, II, III, or IV) classification task is completed by one-to-one method, which means to train $\binom{4}{2}$ SVMs simultaneously and integrate the results of each SVM by majority voting. In the training process of one-to-one method, each SVM only needs the data in two classes which will have smaller training cost and keep the generalization performance at the same time.

The selection of the kernel function is one of the main factors that influence the performance of SVM. Commonly used kernel functions include linear function, polynomial function, sigmoid function, and radial basis function. Here, multiple kernel functions are used simultaneously to construct the SVM-based multiclassifiers in case of poor generalization performance due to wrong kernel function selection.

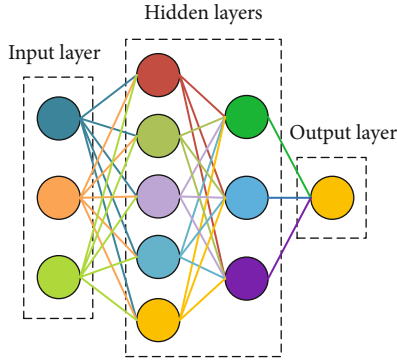


FIGURE 1: Schematic diagram of MLP.

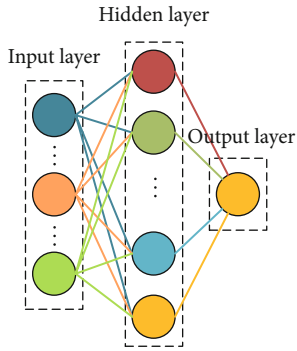


FIGURE 2: Schematic diagram of the RBF neural network.

3.2.2. Introduction to the Neural Network. The neural network (NN) is a model with strong fitting ability and is widely used by researchers in various disciplines. It is mainly composed of an input layer, hidden layers, and an output layer. According to the structure of the network, NN can be divided into multiple types: multilayer perceptron (MLP) neural network, radial basis function (RBF) neural network, adaptive resonance theory (ART) neural network, self-organizing map (SOM) neural network, etc. There are two main network models used in this medical decision system, MLP neural network and RBF neural network.

The MLP neural network contains one input layer, one or more hidden layers, and one output layer. Usually, every two adjacent layers are fully connected as shown in Figure 1.

The activation function of the MLP neural network is usually the rectified linear unit or the ReLU function which can be expressed as $f(x) = \max(0, x)$.

The RBF neural network is another widely used neural network [27]. It converges fast and has strong generalization ability. Unlike the MLP neural network, the RBF network contains only one hidden layer and uses radial basis function $f(x, c_i) = \exp(-\beta_i \|x - c_i\|^2)$ as the activation function, where c_i is the center of the i th unit in the hidden layer. The structure of the RBF neural network is shown in Figure 2.

3.2.3. Ensemble Learning. Ensemble learning is a method that integrates many base learners together to improve the overall learning ability. Commonly used ensemble learning methods include boosting, bagging, and stacking. The stacking method can generate all base learners in parallel

and uses a secondary learner to integrate the results of the base learners. This integration method has suitable training cost and strong generalization performance. In our medical decision-making system, in order to provide corresponding treatment plan recommendations and evaluate the efficacy of the plan after staging the cancer, the results of the classifier needs to be converted into a regression value. In this case, stacking is a good choice. To make ensemble method really work, base learners need to have some difference. This difference may come from different models, different input datasets, or different input features. Given that SVM and neural networks are less sensitive to the input dataset and our feature set is small, it is not suitable to use different input samples or different input features to improve generalization performance. However, there are many choices in the selection of the kernel function for SVM and structure for neural networks. Therefore, parameter perturbation is taken to enlarge the difference between base learners so that our decision model can be strengthened. Finally, the weights of base learners are learned through exponential linear regression (ELR) to obtain the evaluation of tumor malignancy or the EM value.

3.3. Detailed Description of the Medical Decision System. In the proposed medical decision system, six important tumor markers including prostate-specific antigen (PSA), prostate-specific membrane antigen (PSMA), total prostate-specific antigen (tPSA), red blood cell (RBC), hemoglobin (HB), and prostate acid phosphatase (PAP) are chosen for the diagnosis of PCa as the input features and SVM as the diagnosis model. Clinical stage determination of malignant PCa, treatment recommendation, and evaluation are completed by an ensemble model that combines SVM groups for four-class classification and neural networks with different structures. Figure 3 depicts the main flow of the auxiliary medical system.

First, relevant data from different hospital systems are collected. Then, six important tumor markers' levels are extracted from thousands of information items. After dropping samples with missing or abnormal value, an input vector $x = (x_{\text{PSA}}, x_{\text{PSMA}}, x_{\text{tPSA}}, x_{\text{RBC}}, x_{\text{HB}}, x_{\text{PAP}})$ is formed. Next, it will firstly use SVM to judge if the tumor is malignant. In clinical medicine, the increase in tumor marker level does not mean the development of a malignant tumor for sure. Many benign lesions or inflammations may also lead to an increase in tumor marker level, but the increase is not large. When the system determines that the tumor is benign, recommendations about the next examination and corresponding treatment will be listed.

If the tumor is judged to be malignant, the ensemble model will complete the stage division. The development of the malignant tumor will be divided into four stages: I, II, III, and IV; that is, to say, the system must complete a four-classification task. Since SVM is mainly used at binary classification problems, one-to-one strategy is taken and every six SVM models form a group of SVMs. The output of each group is a voting combination of six binary SVM classifiers in the group, which is represented by a four-dimensional one-hot vector. The differentiation of the SVM classifier is realized by choosing different kernel functions to improve the final performance. More explicitly, SVMs in the same

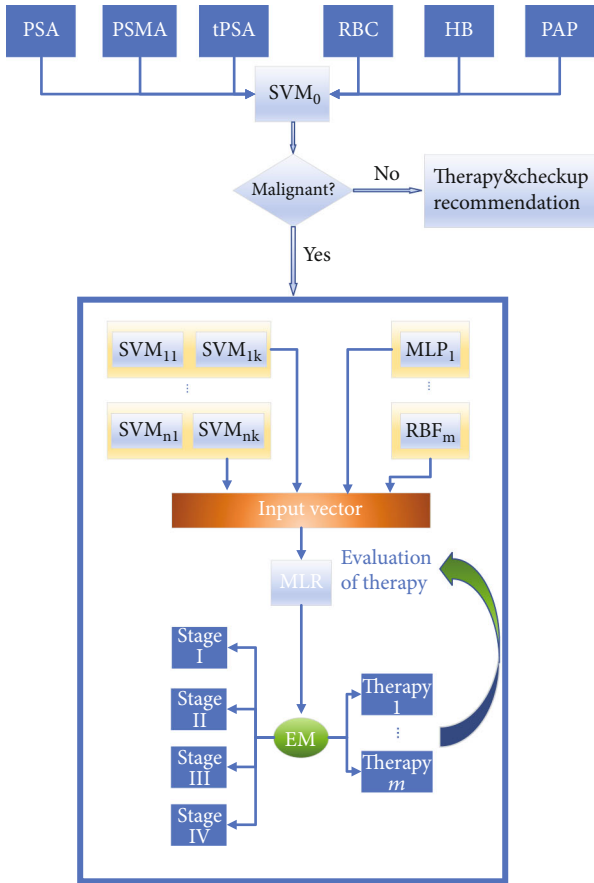


FIGURE 3: The main flow of the auxiliary medical decision system.

group use the same kernel function, and SVMs in different groups use different kernel functions. Three commonly used kernel functions: linear kernel, polynomial kernel, and Gaussian kernel are chosen to differentiate SVM groups.

While training, parameters in the kernel function and penalty parameters of each binary SVM are adjusted to reduce the generalization error below the threshold ϵ . In order to further reduce the risk, the widely used MLP neural network and RBF neural network are added into the system. Because 6 input features are selected and samples are classified into four classes, the input and output layers of the MLP and RBF networks are 6 units and 4 units, respectively. Three group MLP neural networks with different structures are selected: 6-9-7-4, 6-10-7-5-4, and 6-7-5-4 (these numbers follow the order of unit numbers in each layer). The ReLU function is used as the activation function in MLP neural networks. Similarly, three RBF networks with different structures are used. The hidden unit numbers in three networks are set as 10, 14, and 16, respectively. After clustering the samples by k -means algorithm, the center c_i of each hidden unit is determined. The activation function of the RBF neural networks is determined to be radial basis function. For the MLP and RBF networks, the hyperparameters are adjusted to reduce the generalization error below the threshold ϵ .

Finally, outputs of each SVM group and all MLP and RBF networks are connected into one vector, which will be the

input of the secondary learner. By observing the tumor marker level in the dataset, it is found that for benign tumors and patients in stage I, tumor marker levels are usually close to the normal range. But for patients in stage III and stage IV, the level of tumor markers deviates greatly from the normal range. Therefore, we assume that the growth of tumor markers in the development process of PCa conforms to the exponential law. This hypothesis is basically true in medicine. In the early stage, symptoms are very slight or not obvious. Tumors tend to be hard to find and grow slowly. However, in the middle and late stages, they grow savagely and spread throughout the body, making tumor marker levels really high.

Therefore, ELR is selected as the secondary learner to ensemble results of the SVM, MLP, and RBF models. Supervising output values 3, 4, 5, and 6 are added manually for input patient samples in stages I, II, III, and IV, respectively. What may be wired is that the output value of ELR is not set to start from 1. It is considered for the reason of improving the model's robustness to normal people and benign tumor cases. Finally, the evaluation value of PCa's malignancy (EM value) is output. Algorithm 1 shows the procedure that integrates the results of base learners by ELR.

The system determines the stage of malignant PCa according to the EM value and recommends the appropriate treatment method. Commonly used tumor treatment methods include chemotherapy, radiotherapy, excision, drug method, and hospital charge. After treatment methods are recommended by the system, doctors will decide to take it or abandon it or make modification based on it. What must be emphasized is that in the aspect of therapy recommendation, the system mainly gives a rough strategy to relieve doctors' pressure to some extent instead of replacing doctors completely. Concrete treatment is relevant to many factors which cannot be simply given by the system. If one patient has been treated for a while, the system will track the changes of his or her tumor marker levels, input the results of the tumor marker test into the system, and output the EM value to evaluate effect of the treatment. If the EM value changes a little, the treatment plan will be changed. If the EM value decreases greatly, this means it really works, so the original plan will be maintained. During this process, every effective therapy and corresponding EM value will be recorded in the database for further use.

4. Experiment

4.1. Dataset and Models' Training. We collected a large amount of data from three top-class hospitals in China: First Xiangya Hospital, Second Xiangya Hospital, and Third Xiangya Hospital. Relevant information about the data is shown in Table 2.

After screening and preprocessing the data, relevant records of the tumor markers (PSA, PSMA, tPSA, RBC, HB, and PAP) and diagnostic results (benign, stage I, stage II, stage III, and stage IV) are obtained.

Table 3 shows the normal range of six tumor markers related to PCa. Values of malignant patients' tumor marker are several times or even tens of times beyond the normal range.

```

Input:
Training set  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}, y_i \in \{I, II, III, IV\}$ 
Primary classifier  $S = \{SVM_1, SVM_2, SVM_3, MLP_1, MLP_2, MLP_3, RBF_1, RBF_2, RBF_3\}$ 
Output:
Second learning algorithm  $H(x)$ :  $\ln(y_{EM}) = w^T x + b$ 
/*
 $x$  is the input vector.
 $w \in R^{36 \times 1}$  is the weight vector, and each element represents the corresponding weight in  $x$ ;  $w^T$  is the transpose of  $w$ .
 $b \in R^{36 \times 1}$  is the bias vector.
 $y_{EM}$  is the output EM value.
*/
Begin
 $D' = \emptyset$ 
for  $i$  in  $D$  do
  for  $t$  in  $S$  do
     $z_{it} = S_t(x_i)$ ; /*  $z_{it}$  is a four-dimensional vector. */
  end for
   $y'_i = \text{map}(y_i)$ ; /* map function converts the class label into a numerical value. */
   $D' = D' \cup ((z_{i1}, z_{i2}, \dots, z_{it}), y'_i)$ ;
end for
use  $D'$  to train  $H(x)$ ;
output  $H(x)$ ;
End

```

ALGORITHM 1

TABLE 2: Type and number of collected data.

Data type	Number
Patient information	1,933,535 items
Outpatient service	691,238 people
Doctors' device in outpatient	24,021,298 items
Be hospitalized	1,149,187 people
Diagnosis	1,089,327 items
Electronic medical records	4,855,619 items
Doctors' device in clinical	25,757,699 items
Inspection records	157,426 items
Medical laboratory records	8,725,586 items
Routine inspection records	22,358,881 items
Operation records	318,022 items
Drug records	120,546 items

Figure 4 shows the training process. The datasets are divided into two parts: training set and test set, accounting for 80% and 20%, respectively. Each of them is then divided into $S_{\text{malignant}}$ and S_{benign} . First of all, $S_{\text{malignant}}$ and S_{benign} are used and the appropriate kernel function and penalty parameter are searched to train SVM_0 until the test error is below ϵ . Second, malignant samples are divided into four parts $S_{\text{malignant}} = \{S_I, S_{II}, S_{III}, S_{IV}\}$ according to their clinical stages. SVM and neural networks are not sensitive to data. What is more, arbitrary division of data is likely to lead to the problem of imbalanced data which means two datasets do not have the same distribution. Hence, the whole training set is used to train all base learners instead of dividing it into several parts. Each binary SVM is trained separately. Majority

TABLE 3: Normal range of different tumor markers.

Types of tumor marker	Normal range
Prostate-specific antigen	0-4.0 ng/mL
Total prostate-specific antigen	4-20 $\mu\text{g/L}$
Hemoglobin	120-165 g/L
Red blood cell	12-15 g/100 mL
Prostate acid phosphatase	0-9 U/L
Prostate-specific membrane antigen	0-4 ng/mL

voting is used to ensemble the output results of SVMs in the same group. For neural networks, the malignant samples are directly marked as $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$ by their stages. What needs to be emphasized is that while training RBF neural networks, the k -means clustering algorithm is performed to determine the centers of hidden layers. c_i in Figure 4 is a hyperparameter that needs to be tuned. Back propagation and gradient descent are performed to obtain good classification ability. Finally, the output of SVM groups and neural networks are reshaped into one vector, which is used as the input of the exponential linear regression model. Artificial labels y are added to train the ELR model. The loss function of ELR is selected as mean square loss, namely,

$$L(w, b) = \frac{1}{m} \sum_{i=1}^m (EM_i - EM'_i)^2, \quad (5)$$

where EM_i is the evaluation of the i th patient's tumor malignance and EM'_i is the manually set supervising value.

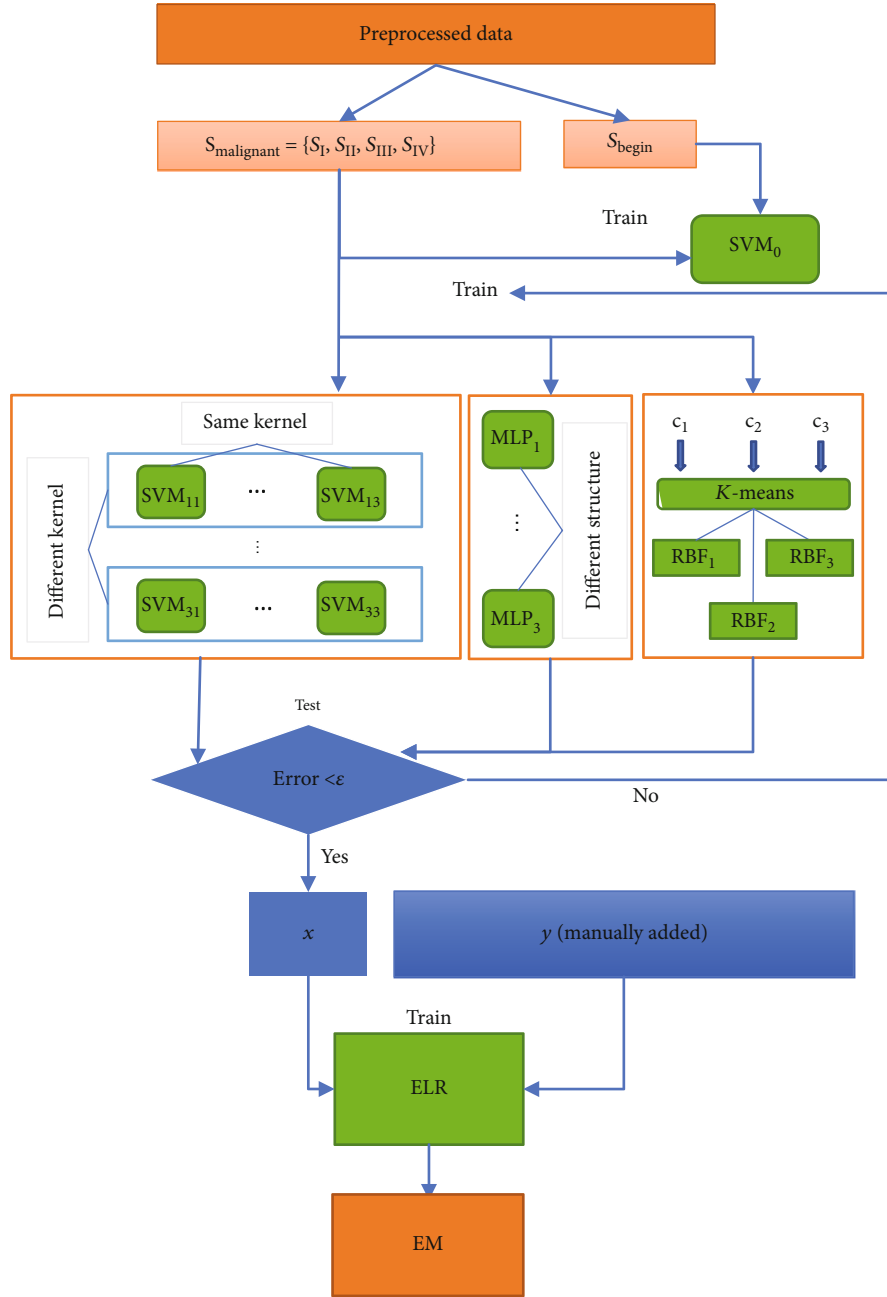


FIGURE 4: Training process of the proposed system.

TABLE 4: EM value of each stage of PCa.

Clinical stage of PCa	Range of ln EM
Stage I	2.7-3.6
Stage II	3.6-4.5
Stage III	4.5-5.3
Stage IV	>5.3

4.2. Analysis of the Results of Experiments. After the model was trained, all the malignant examples in different stages were input into the model and the range of their EM values was calculated, which are listed in Table 4. From Table 4, it

can be known that the EM values of all malignant examples have a rough 0.5 deviation around the supervising value set in advance. The model has good fitting ability on malignant samples of different stages, which indirectly proves our hypothesis that the tumor marker level increases exponentially with the development of tumor is credible.

To verify the effectiveness of our medical decision system, we compared the accuracy of the model on different scale datasets with the accuracy of doctors. As shown in Figure 5, when the amount of data is small, the accuracy of the auxiliary medical decision system is very low, close to 50%. In this circumstance, the accuracy of doctors is really high, almost 100%. However, as the amount of data increases,

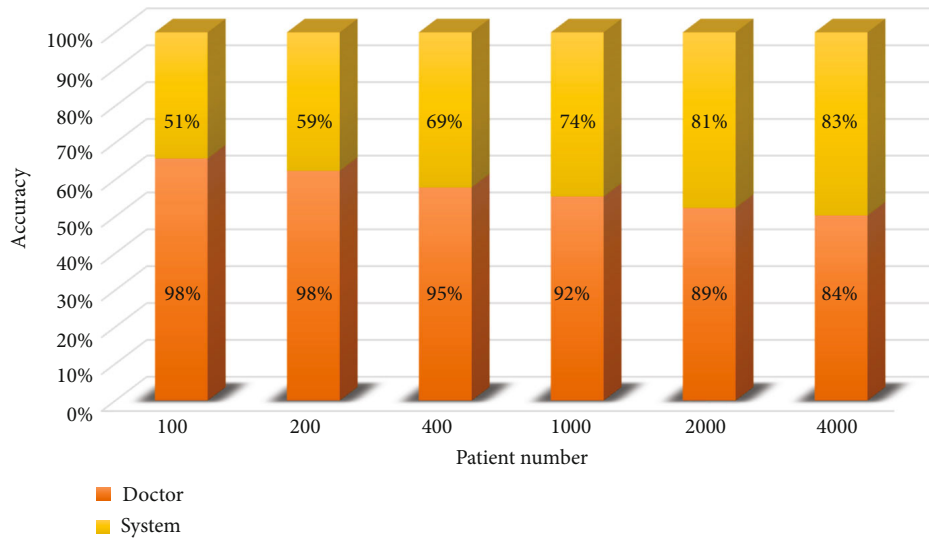


FIGURE 5: Comparison of the doctor and the system.

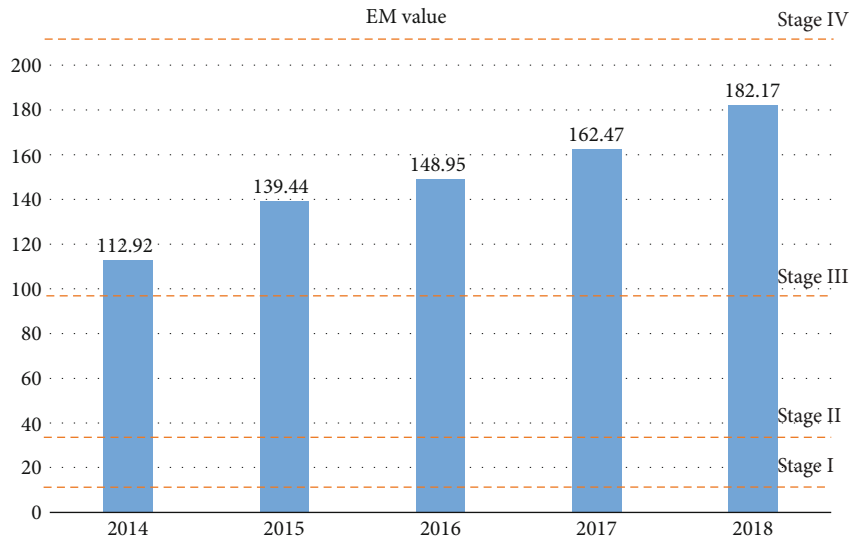


FIGURE 6: Average EM value in the past five years.

the accuracy of the medical decision system increases as well. Simultaneously, doctors' accuracy starts to decline because of the burden and cumulative errors. When the amount of data reaches 4000, the accuracy of the system is roughly the same with that of doctors. This indicates that our auxiliary diagnostic system can make use of the increasing amount of data to improve generalization performance.

We also calculated the average EM value of different years to explore the development trend of PCa in recent years. As shown in Figure 6, the mean EM value of patients from three hospitals has been gradually increasing since 2014. This implies an increase in the number or proportion of patients with malignant prostate cancer which will make medical resources scarcer, so it is necessary and urgent to establish an auxiliary medical decision system based on big data.

Because our medical decision system can quantitatively evaluate the malignancy of prostate cancer, it can easily judge

the efficacy of the treatment plan by its EM value change and recommend treatment methods to improve the condition of PCa patients according to their EM levels. Figure 7 shows the recommended treatment methods and changes of the EM value of a patient whose EM value is very high at first. In the end of the diagnosis interval, the patient's EM value is relatively low, which proves the tumor has been controlled by the recommended treatment plan. It can be concluded that the treatment methods recommended by the system can effectively improve the condition of cancer patients and prolong the survival time for patients in stage III or IV.

4.3. Relevant Analysis Based on the System. Since our medical decision system can evaluate the malignancy of tumors, by controlling different input variables, the influence of a certain factor on prostate cancer can be effectively evaluated. Here, relevant information of some patients was collated. Then,

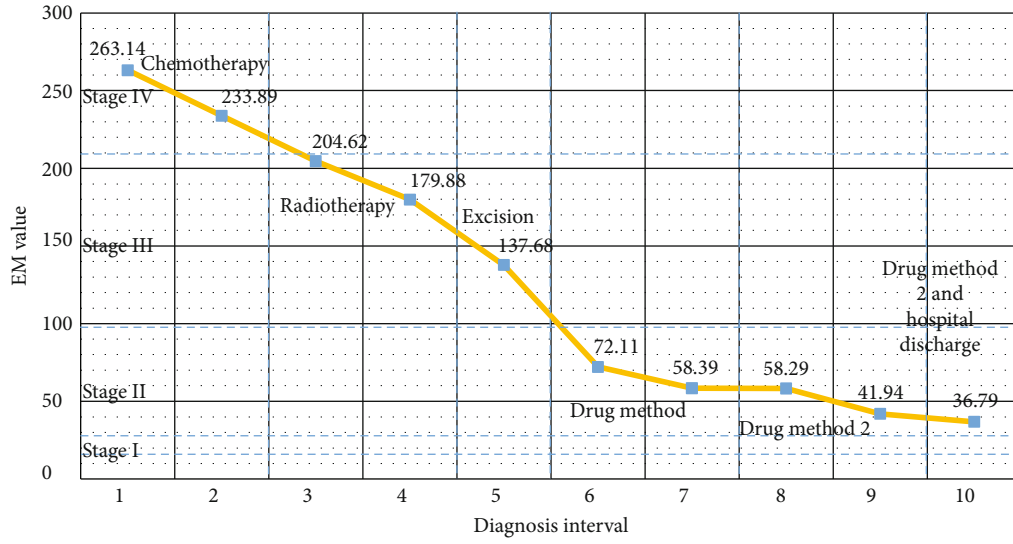


FIGURE 7: A typical treatment process of a PCa patient.

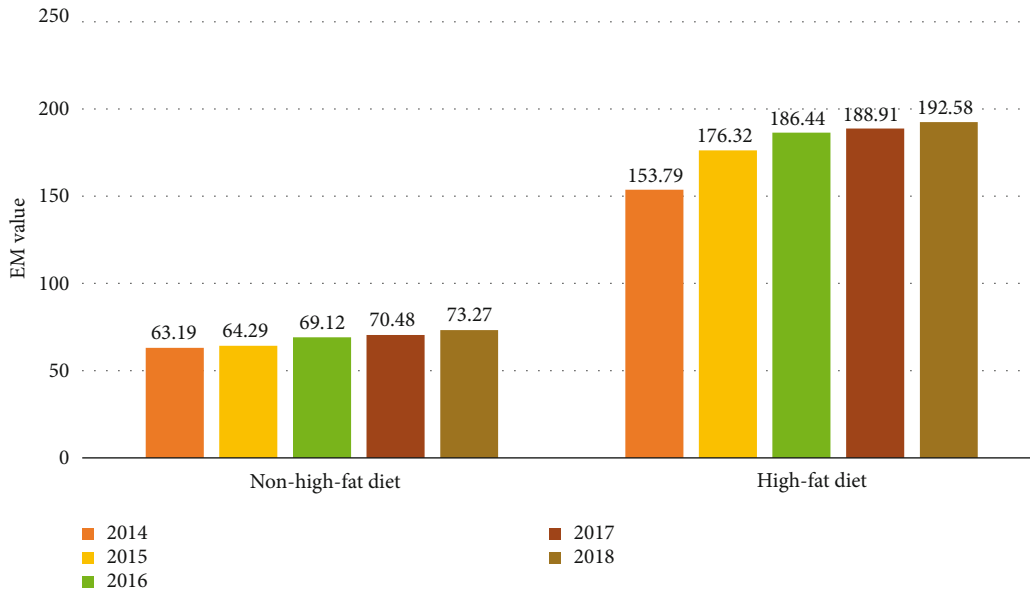


FIGURE 8: Contrast of people with different diet habits.

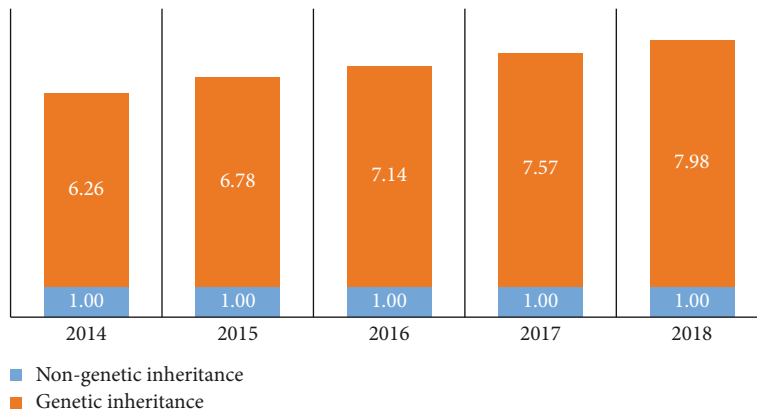


FIGURE 9: Contrast of people with or without genetic inheritance.

influence of patients' diet habits and genetic inheritance on prostate cancer was evaluated. Diet habits are mainly divided into high-fat diets and non-high-fat diets according to the description in the patient's medical history. From the data of 2014-2018, it can be seen that the condition of patients with high-fat diet tends to be more severe. The EM value for patients with high-fat diet is in the range of 150-190, while for those with non-high-fat diet, EM value is only 60-70, as shown in Figure 8.

Genetic inheritance is defined by a cancer case in the patient's family members. The results show that patients with genetic inheritance have a cancer malignancy that is 6 to 7 times that of patients without genetic inheritance, which can be seen in Figure 9.

5. Conclusion

This paper mainly builds an auxiliary medical decision system of PCa for developing countries that lack medical resources. The system is able to provide doctors with advice on the diagnosis, staging, and treatment method of prostate cancer. After training the system in big data environment, although its accuracy continues to rise, it still cannot replace professional doctors and can only be used as an auxiliary diagnostic system to relieve the burden of doctors. Based on this system, we have researched the development of prostate cancer in the past five years and found that the prevalence of prostate cancer is increasing. In addition, high-fat diet and genetic inheritance increase the severity of the disease. The next stage of this research will consider introducing other medical detection information, such as CT, MRI, and PET-CT, to further improve the accuracy and credibility of the system.

Data Availability

All medical data management and system come from Central South University. If readers are interested in those data, you can visit <http://www.xiangya.com.cn/english/>. All data analyzed during the current study are included in the submission.

Conflicts of Interest

The authors declare that they have no conflicts of interest.




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] M. C. S. Wong, W. B. Goggins, H. H. X. Wang et al., "Global incidence and mortality for prostate cancer: analysis of temporal patterns and trends in 36 countries," *European Urology*, vol. 70, no. 5, pp. 862–874, 2016.
- [3] J. Ferlay, I. Soerjomataram, R. Dikshit et al., "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012," *International Journal of Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.
- [4] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: a Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.
- [5] J. Wu and Z. Chen, "Data decision and transmission based on mobile data health records on sensor devices in wireless networks," *Wireless Personal Communications*, vol. 90, no. 4, pp. 2073–2087, 2016.
- [6] J. Wu, X. Tian, and Y. Tan, "Hospital evaluation mechanism based on mobile health for IoT system in social networks," *Computers in Biology and Medicine*, vol. 109, pp. 138–147, 2019.
- [7] J. Wu, Y. Tan, Z. Chen, and M. Zhao, "Decision based on big data research for non-small cell lung cancer in medical artificial system in developing country," *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 87–101, 2018.
- [8] J. Wu, P. Guan, and Y. Tan, "Diagnosis and data probability decision based on non-small cell lung cancer in medical system," *IEEE Access*, vol. 7, pp. 44851–44861, 2019.
- [9] J. Wu, Y. Tan, Z. Chen, and M. Zhao, "Data decision and drug therapy based on non-small cell lung cancer in a big data medical system in developing countries," *Symmetry*, vol. 10, no. 5, p. 152, 2018.
- [10] B. Malmir, M. Amini, and S. I. Chang, "A medical decision support system for disease diagnosis under uncertainty," *Expert Systems with Applications*, vol. 88, pp. 95–108, 2017.
- [11] J. Wu, Z. Chen, and M. Zhao, "Community recombination and duplication node traverse algorithm in opportunistic social networks," *Peer-to-Peer Networking and Applications*, pp. 1–8, 2020.
- [12] J. Wu, Z. Chen, and M. Zhao, "An efficient data packet iteration and transmission algorithm in opportunistic social networks," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, 2019.
- [13] A. Tashkandi, I. Wiese, and L. Wiese, "Efficient in-database patient similarity analysis for personalized medical decision support systems," *Big Data Research*, vol. 13, pp. 52–64, 2018.
- [14] T. Kobayashi, "A blood tumor marker combination assay produces high sensitivity and specificity for cancer according to the natural history," *Cancer Medicine*, vol. 7, no. 3, pp. 549–556, 2018.
- [15] S. Saito, K. Taguchi, N. Nishimura et al., "Clinical usefulness of computer-assisted diagnosis using combination assay of tumor markers for pancreatic carcinoma," *Cancer*, vol. 72, no. 2, pp. 381–388, 1993.
- [16] B. K. Singh, "Determining relevant biomarkers for prediction of breast cancer using anthropometric and clinical features: a comparative investigation in machine learning paradigm," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 2, pp. 393–409, 2019.
- [17] M. López-Riera, I. Conde, G. Quintas et al., "Non-invasive prediction of NAFLD severity: a comprehensive, independent validation of previously postulated serum microRNA biomarkers," *Scientific Reports*, vol. 8, no. 1, pp. 10606–10615, 2018.
- [18] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *COLT '92: Proceedings of the fifth annual workshop on Computational learning theory*, vol. 5, pp. 144–152, ACM Press, 1992.
- [19] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, "Classification of mammogram for early detection

- of breast cancer using SVM classifier and Hough transform,” *Measurement*, vol. 146, pp. 800–805, 2019.
- [20] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, “Using machine learning algorithms for breast cancer risk prediction and diagnosis,” *Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016.
- [21] F. A. Badria, N. Shoaip, M. Elmogy, A. M. Riad, and H. Zaghloul, “A framework for ovarian cancer diagnosis based on amino acids using fuzzy-rough sets with SVM,” in *Advanced Machine Learning Technologies and Applications. AMLTA 2014. Communications in Computer and Information Science*, vol. 488, A. E. Hassanien, M. F. Tolba, and A. Taher Azar, Eds., pp. 389–400, Springer, Cham, 2014.
- [22] Z. Tao, L. Huiling, W. Wenwen, and Y. Xia, “GA-SVM based feature selection and parameter optimization in hospitalization expense modeling,” *Applied Soft Computing*, vol. 75, pp. 323–332, 2019.
- [23] M. Abdar and V. Makarenkov, “CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer,” *Measurement*, vol. 146, pp. 557–570, 2019.
- [24] A. Awad, M. Bader-El-Den, J. McNicholas, and J. Briggs, “Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach,” *International Journal of Medical Informatics*, vol. 108, pp. 185–195, 2017.
- [25] R. Khasha, M. M. Sepehri, and S. A. Mahdavian, “An ensemble learning method for asthma control level detection with leveraging medical knowledge-based classifier and supervised learning,” *Journal of Medical Systems*, vol. 43, no. 6, 2019.
- [26] R. Punmiya and S. Choe, “Energy theft detection using gradient boosting theft detector with feature engineering-based pre-processing,” *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2326–2329, 2019.
- [27] G. Li, G. Zhao, C. Zhou, and M. Ren, “Stochastic Elastic Properties of Composite Matrix Material with Random Voids Based on Radial Basis Function Network,” *International Journal of Computational Methods*, vol. 15, no. 1, article 1750082, 2017.

Research Article

Identification of Upper-Limb Movements Based on Muscle Shape Change Signals for Human-Robot Interaction

Pingao Huang ^{1,2,3} **Hui Wang** ^{1,2} **Yuan Wang**^{1,2,3} **Zhiyuan Liu**^{1,2}
Oluwarotimi Williams Samuel^{1,2} **Mei Yu**^{1,2} **Xiangxin Li**^{1,2} **Shixiong Chen**^{1,2}
and Guanglin Li ^{1,2}

¹CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen 518055, China

²Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen 518055, China

³University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence should be addressed to Guanglin Li; gl.li@siat.ac.cn

Received 19 November 2019; Revised 21 February 2020; Accepted 6 March 2020; Published 14 April 2020

Guest Editor: Anna Visvizi

Copyright © 2020 Pingao Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Towards providing efficient human-robot interaction, surface electromyogram (EMG) signals have been widely adopted for the identification of different limb movement intentions. Since the available EMG signal sensors are highly susceptible to external interferences such as electromagnetic artifacts and muscle fatigues, the quality of EMG recordings would be mostly corrupted, which may decay the performance of EMG-based control systems. Given the fact that the muscle shape changes (MSC) would be different when doing various limb movements, the MSC signal would be nonsensitive to electromagnetic artifacts and muscle fatigues and maybe promising for movement intention recognition. In this study, a novel nanogold flexible and stretchable sensor was developed for the acquisition of MSC signals utilized for decoding multiple classes of limb movement intents. More precisely, four sensors were used to measure the MSC signals from the right forearm of each subject when they performed seven classes of movements. Also, six different features were extracted from the measured MSC signals, and a linear discriminant analysis- (LDA-) based classifier was built for movement classification tasks. The experimental results showed that using MSC signals could achieve an average recognition rate of about $96.06 \pm 1.84\%$ by properly placing the four flexible and stretchable sensors on the forearm. Additionally, when the MSC sampling rate was greater than 100 Hz and the analysis window length was greater than 20 ms, the movement recognition accuracy would be only slightly increased. These pilot results suggest that the MSC-based method should be feasible in movement identifications for human-robot interaction, and at the same time, they provide a systematic reference for the use of the flexible and stretchable sensors in human-robot interaction systems.

1. Introduction

In recent years, wearable devices [1, 2], such as exoskeletons and prostheses [3, 4], have shown a substantial promise in the fields of healthcare and rehabilitation that focus on restoring upper or lower extremity motor functions. More so, advances in technology have led to the development of wearable devices in the form of smart electronics that could continuously monitor different physiological parameters associated with the health status in humans [5, 6]. Although such wearable systems, especially the exoskeletons and

prostheses, have been well developed for decades with remarkable advancements, their commercial and clinical success are still marginal. One of the reasons for this issue should be that the motion intention recognition mechanism employed by the devices is inconsistently accurate, thus leading to poor control output when utilized in a real-life scenario. Meanwhile, accurate motion intention recognition mechanism constitutes an essential part of the devices. Surface electromyogram (sEMG) and electroencephalogram (EEG) have been commonly considered as potential sources of biosignals from which information for decoding human

limb movement intents can be seamlessly obtained, due to their noninvasiveness and ease of acquisition. Although these physiological signals have been widely utilized, they are relatively weak and susceptible to various kind of interferences. For instance, power line noise and motion artifacts would inevitably degrade the motion intention recognition accuracy of wearable systems that utilize sEMG or EEG signals as their sources of control. In an attempt to address this issue, researchers have sort alternative means from which motion intentions could be decoded which includes ultrasound [7], pressure [8], capacitance [9], muscle circumference [10], and muscle activation [11, 12]. However, some of the systems are relatively large in size and integrate sensors that lack flexibility and stretchability characteristics, which are the core requirements for developing smart miniaturized intelligent devices that could be easily adopted in practical applications. Therefore, there is a need to conduct further research in this direction that would lead to the development of a new sensing material for motion intention recognition with the capability to resolve the limitations of the existing sensing techniques in the context of wearable systems.

Recently, the use of flexible and stretchable sensing materials had attracted much attention in the bioelectrical signal recording and health monitoring domains [13, 14]. In this regard, various flexible and stretchable sensors have been developed [15], including strain, pressure [16–21], and tactile [22, 23] sensors. Interestingly, these sensors have been used for human motion monitoring [24–29], human-machine interfaces in the context of rehabilitation, and health monitoring [30–33]. Compared to the traditional sensors, some of these sensors are not only flexible and stretchable but also equipped with additional new features including self-power, self-cleaning, self-healing, and transparency, making them more convenient and feasible to adopt in the modern-day wearable systems. For instance, Song and Yang developed a self-power sensor with the capability to monitor human body movements while sleeping [34]. A self-healing strain sensor was developed by Cai et al. to detect the different joint movements in humans [35]. Trung et al. proposed a transparent hybrid sensor that could detect the temperature and strain associated with the human body [36]. In another study, Muth et al. developed a strain sensor that was mounted on a glove to detect the movements of human fingers [37], while Meyer et al. proposed the use of a textile pressure sensor for the detection of muscle activities in human [38]. It should be noted that the above work mainly focused on examining the electrical and physical properties of the sensors without systematic investigation and detailed experimental study of the sensors particularly in the context of human motion intention recognition, which constitutes a research gap.

To fill this research gap which may facilitate practical applications, this study firstly developed a new sensor based on nanogold flexible material to detect muscle shape change (MSC) information from which limb movement intentions could be adequately decoded. Secondly, a portable wireless acquisition system was built for the recording of the MSC signals picked up by the nanogold flexible and

stretchable sensors. Thirdly, the performance of the newly developed MSC-based sensor for motion intention recognition was extensively validated following a systematic study using datasets obtained from nine able-bodied subjects that observed seven classes of targeted upper-limb movements. Fourthly, we investigated the effects of sensor dimension, placement location, sampling rate, feature extraction method, and analysis window length, on the motion intention recognition accuracy (this is the main index for evaluating the performance of the system) of the proposed MSC-based sensor. Lastly, the stretchability and flexibility of the proposed motion intention recognition MSC-based sensor were also examined to determine the possibility of adopting it in real-life applications. In summary, we believe that this study would provide a symmetric guide on the selection of optimal core parameters (such as feature set, locations, sizes, sampling rates, and window lengths for data processing) required in the practical application of stretchable and flexible sensor in the context of motion intension recognition for human-robot interaction.

The rest of this paper is organized as follows. Section 2 describes the fabrication process of nanogold flexible and stretchable sensor and the portable wireless acquisition system and gives the systematic experimental protocols utilized in validating the sensor's characteristics. Section 3 presents the experimental results. Section 4 discusses the results. Finally, Section 5 presents the conclusion and future work.

2. Materials and Methods

2.1. Material. Gold is a well-known material with characteristics such as good conductivity, ductility, and biocompatibility, while polymers are soft and stretchable with good biocompatibility. With the aid of the state-of-the-art nano and microprocessing technology, we developed a soft and stretchable conductor using gold and polymer materials, which is conformal and biocompatible to detect biomechanical signal induced by the shape change of muscles. A detailed description of the fabrication process of the material is described in [39]. Meanwhile, the structure of the material used to fabricate the soft-stretchable sensor is presented in Figure 1(a). As shown in this figure, the top layer is made of the nanogold film while the bottom layer is a substrate known as polydimethylsiloxane (PDMS). Randomly distributed microcracks were observed inside the thin metal film on top of the polymer, as shown in Figure 1(b). When the film is subjected to tensile strain, the conductive path is still built up due to the randomly distributed microcracks. In addition, the conductivity of the material changes regularly with the opening and closing of the microcracks during the stretch/release process.

Since the conducting material is soft and stretchable and conformal with the texture of the human skin surface, the MSC can effectively induce the corresponding mechanical strain in the conductor. Additionally, the stretchable conductor maintains the conductivity regularly during the mechanical tensile strain, and by examining the conductivity

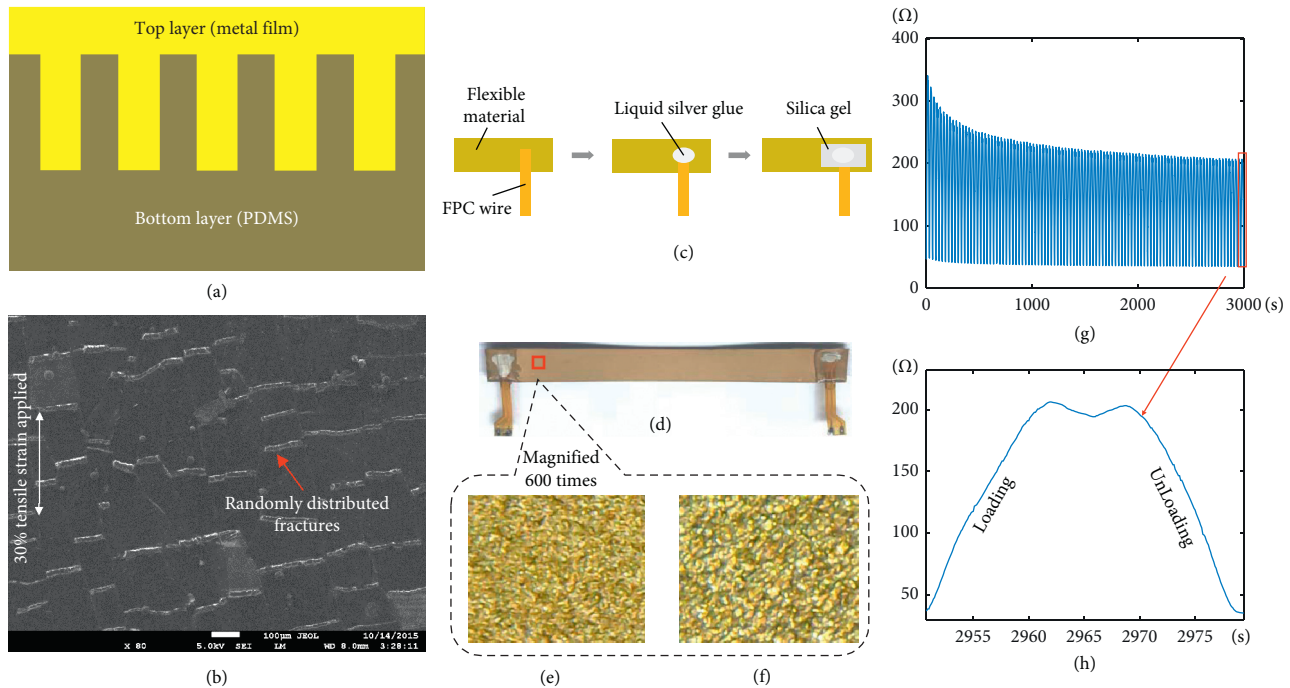


FIGURE 1: Fabrication of the sensor. (a) The structure of the material. (b) The microcracks of the gold film. (c) The process of fabrication. (d) The sensor. (e) The free state. (f) The stretched state. (g) One hundred circles of testing. (h) One circle of loading and unloading.

during the strain, the corresponding changes with respect to the muscles' shape can be detected.

2.2. Fabrication of the Sensors. The sensors utilized for acquiring the MSC signals in this study were fabricated as follows. Firstly, the nanogold material was cut into strips with each strip having a length and width of approximately 8 cm and 8 mm, respectively. Thereafter, flexible printed circuit board (FPC) wires were attached to the strips through a liquid silver gel at the two terminals, and then the strips were placed in an open space for about 10 minutes so that the wires could get glued to the strips properly. This procedure is represented in Figure 1(c). After the silver glue dried up, a silica gel was applied over the silver gel to enhance its adhesiveness which would protect the encapsulated regions of the strips. Finally, the fabricated sensors were placed in a curing oven at 60°C for six hours. A representative of the resulting sensor is shown in Figures 1(d)–1(f), characterized by a stretched surface area which is 600 times the original surface area of the strip. The stability of the fabricated sensors was examined by pulling the strips 100 cycles on a tensile machine (AG-X plus 100N, Shimadzu, Japan) and a multimeter (Keithley 2000, Tektronix, USA) to observe the stretchability against its resistance. The pulling process is described as follows: (a) the sensors were pulled at a constant speed of 1 mm/s along the direction of the sensors; (b) When the sensors were stretched at the speed of 1 mm/s to the predetermined elongation (20% of their lengths), the tensile machine held the state for 5 seconds. After that, the tensile machine relaxed at the speed of 1 mm/s until the sensors recovered to their original lengths. As shown in Figures 1 (g) and 1(h),

when the number of stretching cycles increases from 1 to 100, a corresponding decrease from an initial 340 ohms to a stable 210 ohms in the resistance of the sensor is observed. This stretching process makes the sensors more consistently stable for practical applications.

2.3. MSC Signal Acquisition System. A 4-channel acquisition system (length: 6.2 cm, width: 3.5 cm, height: 0.7 cm, and weight: 19 g) was developed to obtain the alteration in resistance of the sensors caused by the muscle shape change. As shown in Figure 2, the acquisition system was made up of three parts, namely, the analog front-end module, an MCU (WIFI transceiver included) module, and a computer. The ADS1292R (Texas Instruments, Texas, USA) is the analog front-end chip used to acquire bioelectrical signals, such as EMG signals. This analog front-end chip is also used for respiration resistance measurement. The respiration modulating module generates a 64 kHz square wave that is applied to the sensor, thus inducing a current that flows through the sensor. Thereafter, a voltage is produced by the current in the sensor and then amplified and demodulated by a respiration demodulating module. The demodulated signal is then digitalized by using a 24 bits sigma-delta ADC, and finally, the data are sent to the computer through WIFI module.

2.4. Setup of Experiments. To investigate the performance of the newly fabricated sensors, two different experimental sessions were designed for the collection of MSC signals associated with multiple classes of upper-limb movements in an offline mode. Information about the participants and the data acquisition procedure is provided as follows.

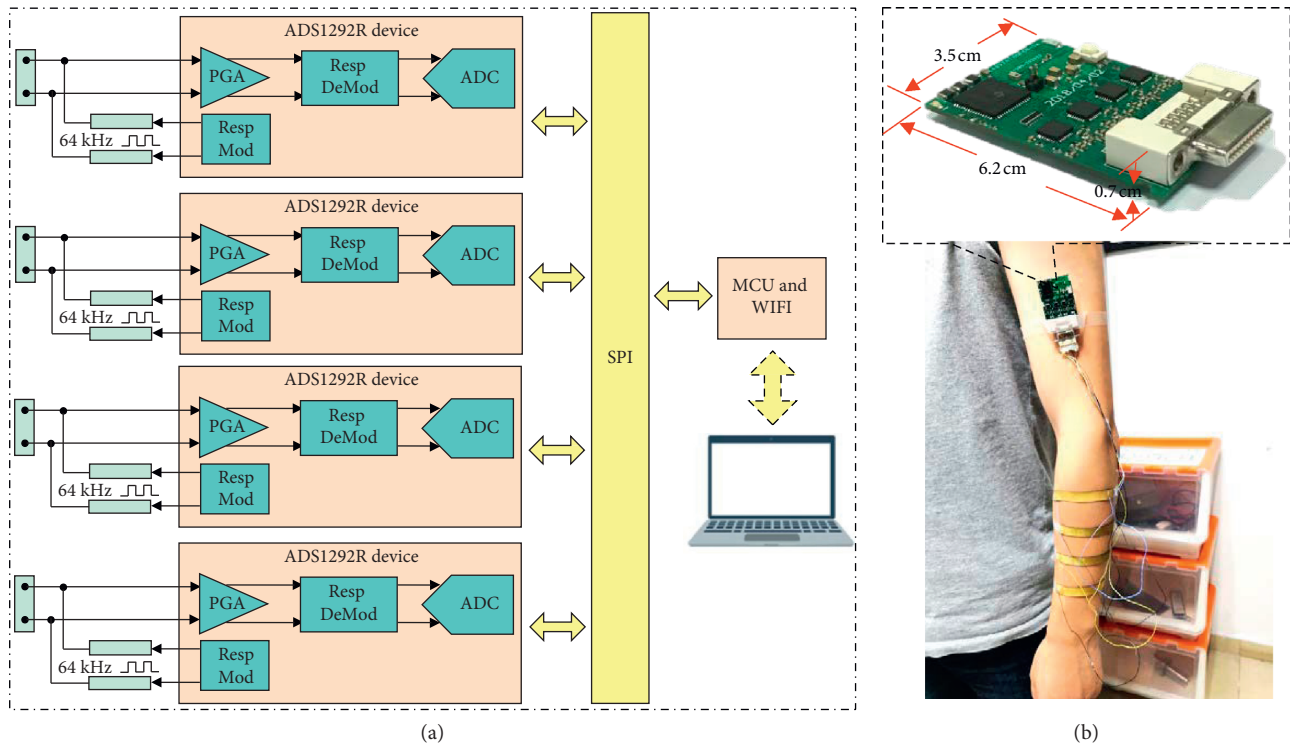


FIGURE 2: The data acquisition system. (a) The schematic diagram of the acquisition system. (b) The practical system.

2.4.1. Participants' Information. In this study, a total of nine able-bodied subjects including six males and three females (aged from 24 to 30, with an average of 26.3) were recruited. The protocol of this study was approved by the Institutional Review Board of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. All subjects gave written informed consent and provided permission for the publication of their photographs for scientific and educational purposes.

2.4.2. Setup of the Movements. The MSC signals were acquired at a sampling frequency of 1000 Hz using the four-channel data acquisition system described above. More precisely, two different kinds of stretchable-flexible sensors, large-sized sensors (length: 8 cm and width: 0.8 cm) and small-sized sensors (length: 3 cm and width: 5.0 mm), were designed for the MSC data collection with an attempt to see if the sensor size would affect the MSC recordings, as shown in Figure 3(a). During data collection sessions, each participant was instructed to perform seven classes of targeted upper-limb movements that were hand close (HC), hand open (HO), wrist pronation (WP), wrist supination (WS), wrist extension (WE), wrist flexion (WF), and one inactive limb movement known as the rest state (RS) as shown in Figure 3(b). Note that these classes of the upper-limb movement tasks have been considered in a number of previous related studies [40–42]. Prior to the data collection sessions, the subjects were properly instructed about the experimental procedure to guarantee high-quality recordings. Furthermore, each subject was allowed to perform several preexperimental trials to get

themselves familiar with the experimental protocol. Following these procedures, the subjects performed each movement based on a video prompt for 5 seconds, and each movement class was followed by a rest session of five seconds before observing the next active movement class. In training, the order of active movements is as follows: HC, HO, WP, WS, WE, and WF, and each subject was asked to repeat the process three times.

2.4.3. Locations of the Sensors. In order to examine the optimal location for the MSC sensor placement, 16 locations along the vertical plane were selected. This is because if placed along the longitudinal direction of the arm, the sensors would be folded and therefore capture lesser information since they will not be making absolute contact with forearm muscles. As shown in Figure 3(c), each column's sensors were equally distributed between the chelidon and the end (near the hand) of the brachioradialis muscle. For the large-sized sensors, different placement locations, namely, the radial side (column 1, sensors 1 to 4, named region 1), ulnar side (column 2, sensors 5 to 8, named region 2), posterior side (column 3, sensors 9 to 12, named region 3), and the anterior side (column 4, sensors 13 to 16, named region 4) of the forearm, were designed. For the small-sized sensors, four additional sensor placement strategies were used, but this time in a rowwise manner. The small-sized sensor placements are described as row 1 (sensors 1, 5, 9, and 13, named region 5), row 2 (sensors 2, 6, 10, and 14, named region 6), row 3 (sensors 3, 7, 11, and 15, named region 7), and row 4 (sensors 4, 8, 12, and 16, named region 8). It should be noted that the four

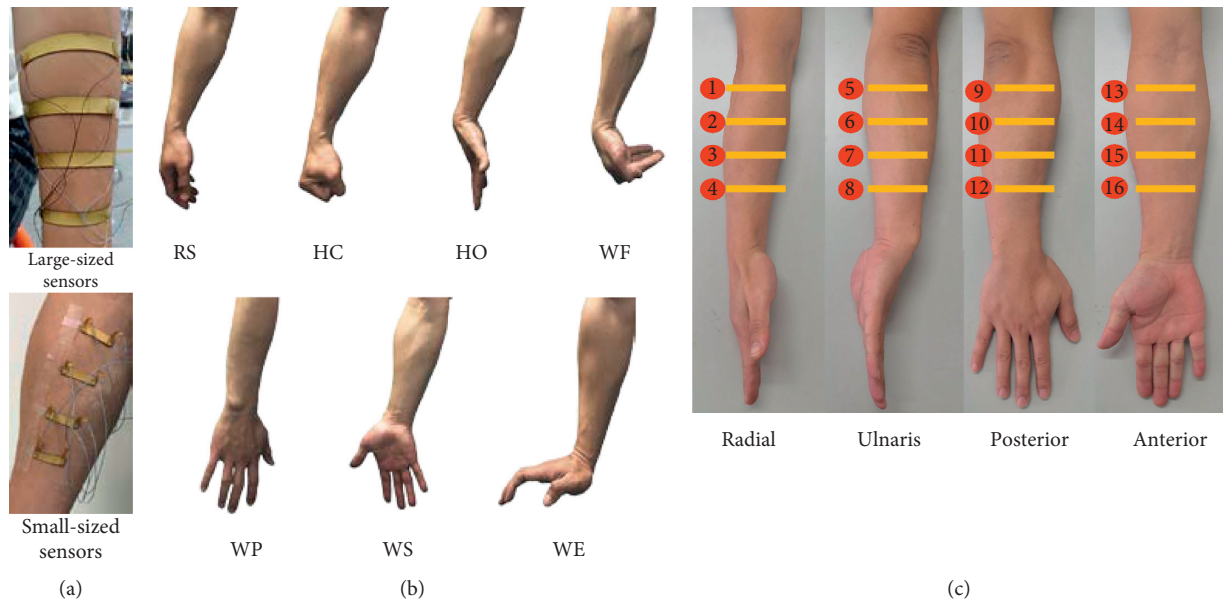


FIGURE 3: The protocol of the experiments. (a) Placements of two types of sensors on the forearm. (b) Seven targeted movements. (c) Sixteen locations.

sensors in a row assume a ring shape around the forearm as shown in Figure 3(c). Using the self-made four-channel acquisition system, the MSC signals could be measured in one row or one column on the forearm. For the large-sized sensors, the data from one experiment could be collected in four times, and the data from one experiment could be collected in 8 times for the small-sized sensors. During the experiment, each subject was asked to take a rest for two minutes between two acquisition sessions. Before being used, each sensor was prestretched and then adhered to the skin with medical adhesive tapes to ensure that they were firmly fixed to the skin during the experimental trials.

2.4.4. Data Preparation. After the MSC signals were acquired, a five-point moving average filter was applied to attenuate the inherent noise. Then, the filtered MSC data were downsampled from 1000 Hz to 500 Hz, 250 Hz, 100 Hz, 50 Hz, 40 Hz, and 20 Hz, respectively. Finally, to evaluate the effect of different window lengths on the accuracy, a series of windows of 20 ms, 50 ms, 100 ms, 200 ms, and 300 ms were used to segment the MSC data with the overlap length of half of their window lengths.

2.4.5. Feature Selection and Classification. For each of the windowed MSC signals, six features (the mathematical expressions are shown in Table 1), mean value (MVAL), root mean square (RMS), simple square integral (SSI), third moment (TM3), logarithm detector (LOGD), and standard deviation (STD), were extracted, which were used in some previous studies [42, 43]. With the feature sets, the principal component analysis technique was applied to remove redundant information, and then a five-fold cross validation was utilized to partition the feature vector into a training set and a testing set. A linear discriminant analysis (LDA)

classifier was built for each subject to predict the limb-movement intents [44, 45]. The major consideration to use the LDA classifier is its computational efficiency coupled with its wide usage for the human-machine interface. The detailed operational procedure of the LDA algorithm could be referred to as [46].

2.4.6. Statistical Analysis. To examine whether each of the five factors (feature, sensor size, sampling rate, location, and window length) of the sensor has an impact on the accuracy of movement classification, the one-way ANOVA with a post hoc analysis LSD was conducted in terms of mean classification accuracy, using the SPSS Statistical Modeling software (SPSS 22.0 IBM Corp., Chicago, IL). To perform one-way ANOVA, when designing experiments and grouping data, only one of the five factors was changed at a time while several other factors retained their typical values unchanged. A level of $p < 0.05$ was selected as the threshold for statistical significance with the null hypothesis that the classification accuracies achieved by one factor's changing (such as the frequency changing, 1000 Hz, 500 Hz) among the five factors are not significantly different from each other.

3. Results

3.1. Waveforms of MSC Signals. Figures 4(a) and 4(b) show two typical MSC recordings by the large-sized and small-sized sensors, respectively. The six classes of active movements and the inactive movement could be visually distinguished from the MSC recordings. The MSC signal values in different channels varied in a range of dozens of ohms during the movements and had a different baseline. The preapplied tension and the intrinsic resistance of the sensors were different from each other. At the rest state, MSC signal

TABLE 1: The mathematical expressions of the six features.

Serial number	Feature name	Abbreviation	Mathematical expression
1	Mean value	MVAL	$MVAL = (1/k) \sum_{n=1}^k x_n$
2	Root mean square	RMS	$RMS = \sqrt{(1/k) \sum_{n=1}^k (x_n)^2}$
3	Third moment	TM3	$TM3 = (1/k) \sum_{n=1}^k (x_n)^3$
4	Simple square integral	SSI	$SSI = \sum_{n=1}^k (x_n)^2$
5	Logarithm	LOGD	$LOGD = e^{(1/k) \sum_{n=1}^k (x_n)}$
6	Standard deviation	STD	$RMS = \sqrt{(1/k) \sum_{n=1}^k (x_n - MVAL)^2}$

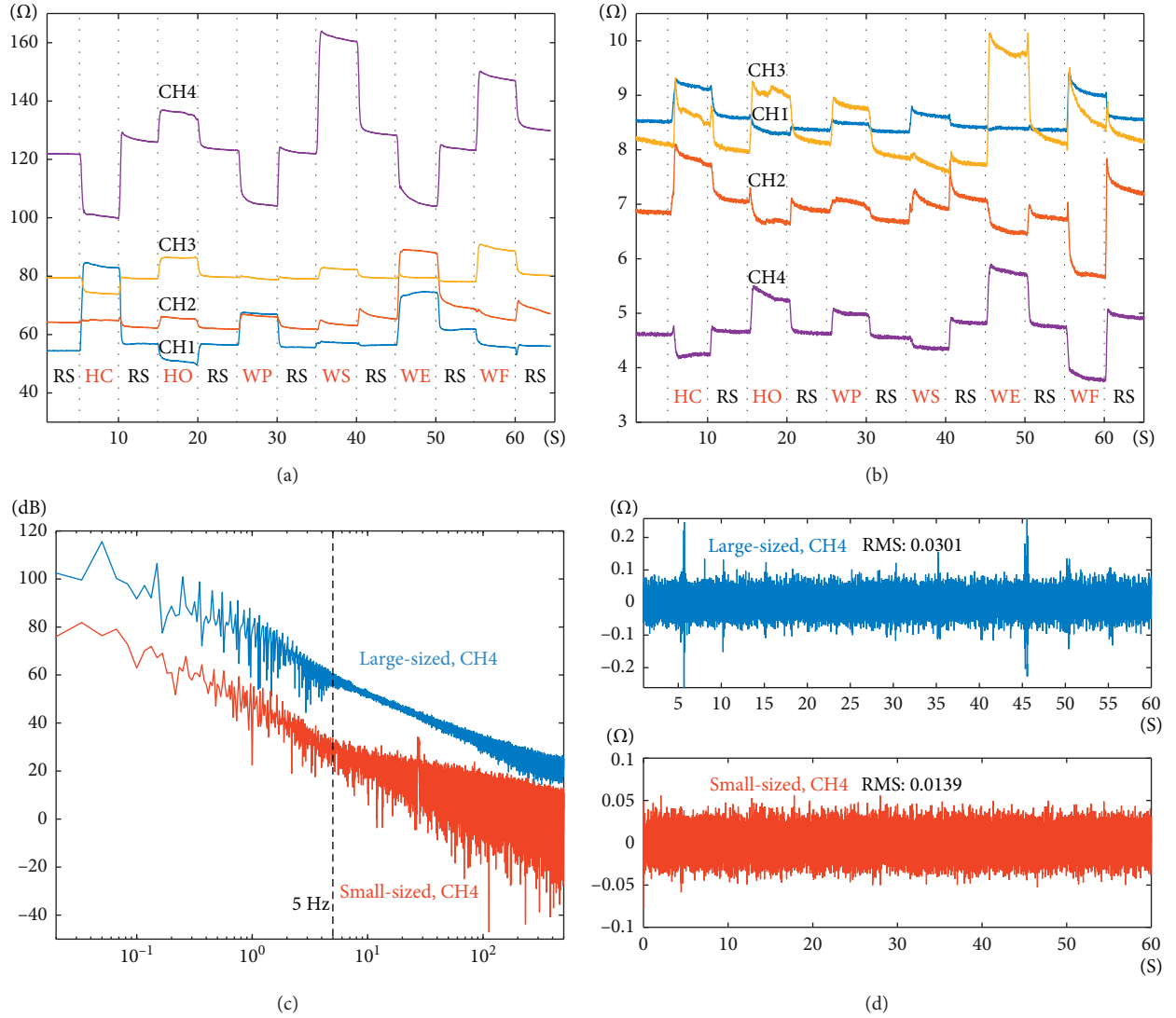


FIGURE 4: Typical waveforms of MSC signal recordings. (a) The 4-channel waveforms of the large-sized sensors. (b) The 4-channel waveforms of the small-sized sensors. (c) The signal spectrums of CH4 for both sensors. (d) The noises of CH4 for both sensors.

values seem different in different channels, varying with individual movement. Comparatively, the value of the large-sized sensors changed more than that of the small-sized sensors.

Figure 4(c) shows the spectrums of CH4 of both sensors. It could be known that the main components of the MSC signal are concentrated within 5 Hz. Therefore, the clean MSC signal S_{MSC} is obtained by using a low-pass filter (IIR,

Butterworth, the cutoff frequency is 5 Hz), with very small noise (for the band is only 5 Hz). The noise N_i is obtained by a high-pass filter (IIR, Butterworth, the cutoff frequency is 5 Hz). Figure 4(d) shows the noise of CH4 of both sensors, and the small-sized sensors have lower noise. Then, the SNR is calculated as equation (1), and the large-sized sensor and the small-sized sensor have SNRs of 49.10 ± 4.91 dB and 29.21 ± 1.97 dB, respectively. The SNR of the large-sized

sensor is about 20 dB higher than that of the small-sized sensor.

$$\text{SNR} = 10 \log_{10} \frac{P_{S_{\text{MSC}}}}{P_{N_i}}. \quad (1)$$

3.2. Effect of Features on the Classification Accuracy.

Figure 5(a) shows the relationship between the average classification accuracy vs. the six features across all the sensor locations in nine able-bodied subjects with a sampling frequency of 1000 Hz and window lengths of 100 ms. For the two groups of sensors, the classification accuracies when using the STD feature were significantly different from the other features ($p < 0.01$). For the small-sized sensors, except for the STD feature, the TM3 reflected a different characteristic in comparison to the MVAL, RMS, and LOGD ($p < 0.016$) features, respectively. For the large-sized sensors, MVAL, RMS, TM3, SSI, and LOGD almost had the same accuracy and no significant difference (between all groups, $p > 0.635$). It can be seen from Figure 5(a) that the accuracy of the large-sized sensors was about 3% higher than that of the small-sized sensors. For all the six features, the average accuracies of the movement classifications were about $81.81 \pm 21.29\%$ when using the large-sized sensors and $75.95 \pm 24.30\%$ when using the small-sized sensors. ANOVA shows that the average classification accuracies were significantly different for the different sensor sizes ($p = 0.005$).

Furthermore, using the MVAL feature as a basis, the other five features were added one-by-one in the sequence of RMS, STD, LOGD, TM3, and SSI, and then used to classify the movement intentions for each subject as shown in Figure 5(b). It can be seen from Figure 5(b) that the classification accuracies slightly increased as the number of features increases (from 1 to 6) for both the large-sized and small-sized sensor configurations. For the small-sized sensors, the first four features have no significant impact on accuracy (at, $p > 0.50$), while the addition of the fifth and sixth features led to a significant increase in accuracy (for five features, $p = 0.035$; six features, $p = 0.013$), from $87.57 \pm 8.20\%$ to $90.97 \pm 6.19\%$. For the large-sized sensors, there is no significant difference in accuracy ($p > 0.09$) even when all the six features were concatenated and used for classifying the limb movement intent of the subjects, where only a slight increase can be observed in accuracy (from $91.47 \pm 5.66\%$ to $93.73 \pm 4.90\%$). These results suggested that the large-sized sensors would achieve higher accuracy in comparison to the small-sized sensors ($p = 0.002$).

3.3. Effects of Sensor Locations on the Classification Accuracy.

With an attempt to look for the optimal sensor placements on the forearm, the effects of different sensor locations on the forearm of the subjects were investigated. Eight different regions for the four small-sized sensors' placement and four different regions for the large-sized sensors' placement, as shown in Figure 3(b), were examined. Six features were used along with the sampling frequency and window length adopted in the previous section. The average classification

accuracy over all the nine subjects was calculated with each sensor placement region and is presented in Figure 6. For the small-sized sensors, it can be seen from Figure 6(a) that the region 6 achieved the highest accuracy of $95.07 \pm 3.87\%$, while the region 4 had the lowest accuracy of $88.64 \pm 6.34\%$. Additionally, the ANOVA showed that there was significant difference between region 6 and region 4 ($p = 0.042$) and also between region 6 and region 7 ($p = 0.046$). Meanwhile, the total average accuracy (across all the subjects and all the regions) was about $90.97 \pm 6.19\%$. For the large-sized sensors, the ANOVA indicated that no significant difference between all the four regions was observed ($p > 0.15$). Figure 6(b) shows that region 3 had the highest average accuracy of $96.06 \pm 1.84\%$ among all the four regions, and the total average accuracy (across all the nine subjects and all the four regions) was about $93.73 \pm 4.90\%$.

3.4. Effects of Sampling Rates and Window Lengths on the Classification Accuracy.

To investigate the effect of different MSC signal sampling rates on the motion intention recognition accuracy, the acquired signal was downsampled from 1000 Hz to 500 Hz, 250 Hz, 100 Hz, 50 Hz, 40 Hz, and 20 Hz, respectively. The identical window length of 300 ms was used for the different sampling rates and the six features were extracted from each analysis window. The overall classification accuracy over all the regions and all the subjects was calculated for each sampling rate, as shown in Figure 7(a). It can be observed from Figure 7(a) that using a sampling rate from 100 Hz to 1000 Hz, both the large-sized and small-sized sensors showed a steady accuracy. If the sampling frequency was lower than 50 Hz, there was an obvious and significant decrease in the accuracy ($p < 0.001$).

In addition, the effect of window length on motion intention recognition accuracy was examined by using five different window lengths (20 ms, 50 ms, 100 ms, 200 ms, and 300 ms), respectively. The six features were extracted from each analysis window with each window length for the motion intention recognition. The overall classification accuracies over all the regions and the nine subjects are shown in Figure 7(b). We can see from Figure 7(b) that the movement classification accuracies of both the large-sized and small-sized sensors only had a slight increment (less than 1.2%) with a corresponding increase in the window length.

3.5. Classification Accuracies of Different Movements.

Following the above-described procedures, to evaluate the classification performance of different movements, we calculated the confusion matrices of classification across all subjects for optimal sensor locations (Figure 8, in counts of testing samples). It can be known from the figure that (a) all active movements almost have the same true positives; (b) the movements HO and RS have the greatest interaction; (c) the samples of the classes are imbalanced (each active movement has 2133 samples, while the RS movement has 12933 samples; the ratio is about 1 : 6).

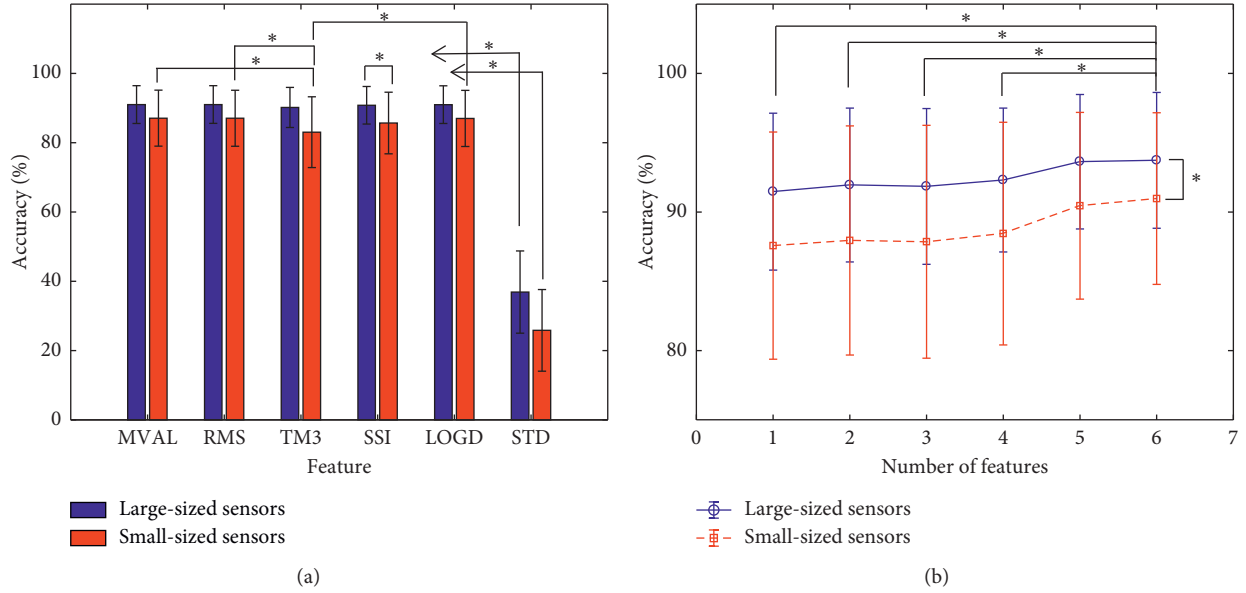


FIGURE 5: Effect of the features on movement classification accuracy. (a) Effect of different features on accuracy. (b) Effect of different numbers of features on accuracy.

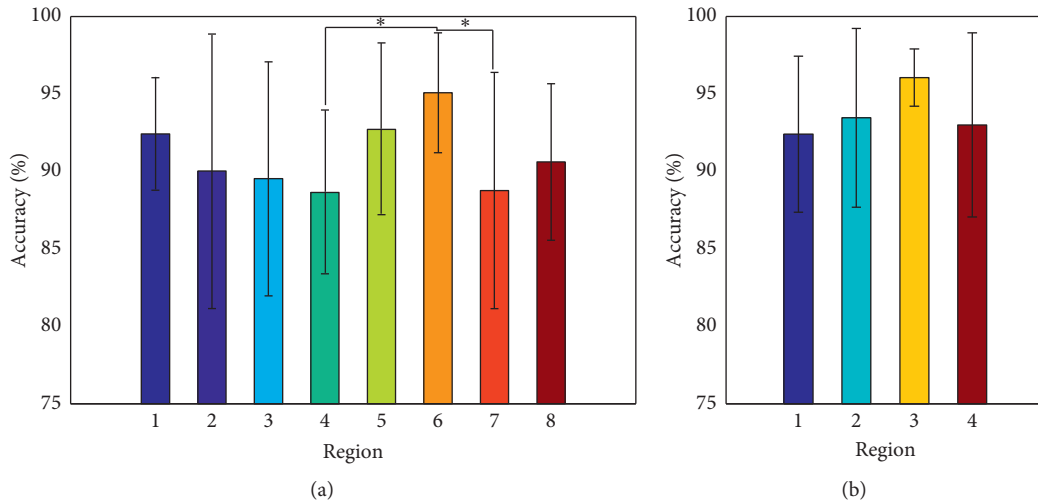


FIGURE 6: Accuracies vs. regions. (a) Accuracies of the small-sized sensors. (b) Accuracies of the large-sized sensors.

According to Sokolova and Lapalme, there are eight measures for multiclass classification, and the measures have some invariance properties (that is, they preserve their value under a change in the confusion matrix) [47]. For example, average accuracy is invariant to the exchange of positives and negatives of the confusion matrix, while recall is invariant to the change of true negative counts. This is beneficial for evaluating the performance of classification. Therefore, Precision (P_i), Recall (R_i), and F-score (F_i) of each class ($\beta=1$, Precision and Recall are considered equal) were calculated (shown in Table 2), and Macro-Precision (P_M), Macro-Recall (R_M), and Macro-F-score (F_M) were also calculated (shown in Table 3). For both sensors, it can be seen from Table 2 that (a) the RS movement had the lowest Precision among all the features, but has the highest Recall; (b) the HO movement had the lowest Recall. (c) When

considering both precision and recall, the HO movement had the worst performance. Additionally, all the three measures showed that the large-sized sensors could achieve relatively higher performance in comparison to the small-sized sensors (shown in Table 3).

4. Discussion

Adequate human motion intention recognition technique aids the realization of efficient human-robot interaction mechanisms required to provide intelligent control systems in the context of rehabilitation or service robots. Meanwhile, information extracted from a number of physiological signals such as EMG and EEG has been widely utilized for the decoding of human motion intention. However, such signals are often subjected to different interferences resulting from

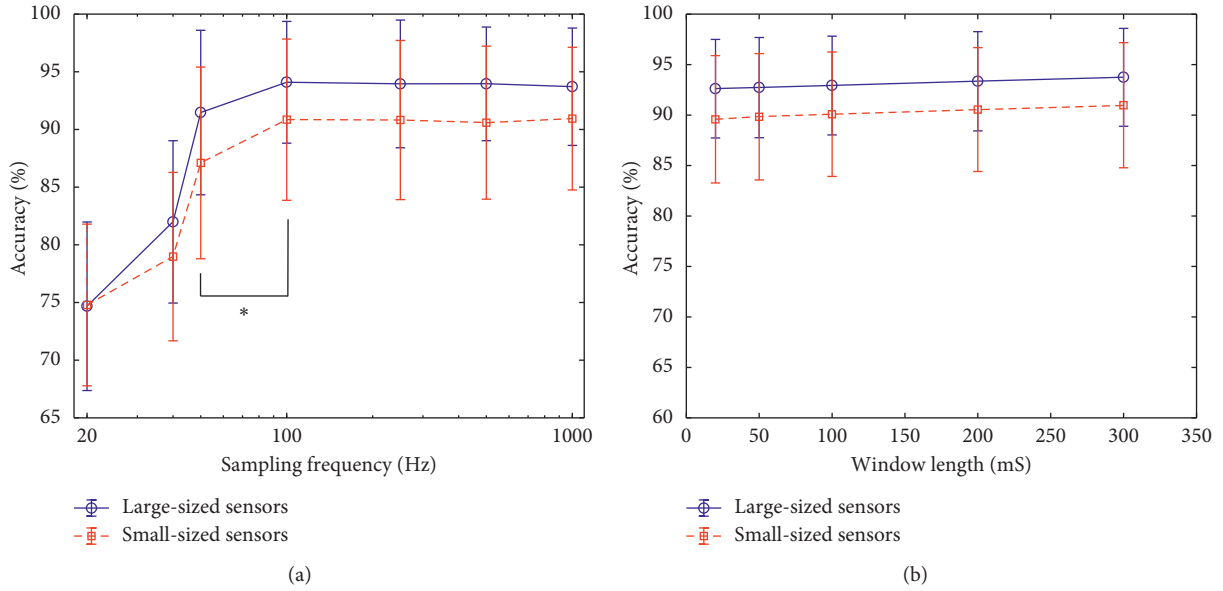


FIGURE 7: Effects of sampling rates and window lengths on movement classification accuracy. (a) Effect of sampling rates on accuracy (abscissa is logarithmic). (b) Effect of window lengths on accuracy.

	HC	HO	WP	WS	WE	WF	RS
HC	2114	0	1	4	0	1	92
HO	0	2070	18	1	38	23	483
WP	0	1	2097	0	0	3	76
WS	12	0	0	2109	10	2	145
WE	1	7	7	3	2044	5	401
WF	0	27	0	3	3	2079	112
RS	6	27	11	14	38	21	11623

(a)

	HC	HO	WP	WS	WE	WF	RS
HC	2067	3	2	10	0	1	195
HO	12	2047	35	5	15	14	740
WP	13	19	2034	1	10	9	277
WS	11	1	0	2075	1	11	350
WE	3	2	9	1	2067	1	249
WF	2	12	19	7	4	2089	62
RS	26	50	34	33	37	8	11060

(b)

FIGURE 8: Confusion matrix. (a) Fusion matrix of large-sized sensors. (b) Fusion matrix of small-sized sensors.

TABLE 2: Precision, Recall, and F-score of each movement (unit: %).

	Measure	HC	HO	WP	WS	WE	WF	RS
Large-sized sensors	P_i	99.10	97.05	98.32	98.86	95.81	97.45	89.87
	R_i	95.59	78.62	96.27	92.58	82.83	93.45	99.01
	F_i	97.31	86.87	97.28	95.62	88.85	95.41	94.22
Small-sized sensors	P_i	96.89	95.96	95.34	97.3	96.89	97.93	85.52
	R_i	90.73	71.41	86.11	84.69	88.63	95.22	98.32
	F_i	93.71	81.88	90.49	90.56	92.58	96.56	91.47

TABLE 3: Macro-Precision, Macro-Recall, and Macro-F-score of both sensors (unit: %).

	P_M	R_M	F_M
Large-sized sensors	96.64	91.19	93.65
Small-sized sensors	95.12	87.87	91.04

electromagnetic artifacts, touch resistance between skin and electrodes, and muscle fatigue among others. And these interferences have been well studied with reports revealing

their negative effects towards degrading real-time performances of motion intention decoding. Alternatively, non-physiological signals based on muscle geometric and/or morphology changes that can be measured by different techniques such as ultrasound [7], capacitance [9], muscle circumference [10], and muscle activation [11, 12] have been considered for motion intention recognition. In that regard, this study hypothesized that nonphysiological MSC signals should offer adequate information for limb movement intent

decoding. In this study, we systematically examined the feasibility of utilizing the MSC signals acquired through the newly developed nanogold flexible and stretchable sensor for upper-limb movement intent decoding in the currently evolving human-robot interaction systems.

Firstly, we extensively examined the characteristics of the MCS signals when different features, namely, linear (MVAL), nonlinear (RMS, SSI, TM3, and LOG), and statistical (STD) features, were extracted for the limb movement decoding task under multiple criteria. The outcome of the investigation reveals that except for the STD feature that recorded extremely low accuracies (less than 50%), the other examined features achieved high accuracies with somewhat similar performance for the movement intent decoding task. Also, by concatenating the features extracted from the MSC signals in an incremental manner, it was found that using more features would only result in slight increase in accuracy (that is, about a 3.4% increase for the small-sized sensors and 2.26% increase for the large-sized sensors), indicating that it might be unnecessary to utilize multiple features when adopting the MSC signals in practical applications. Importantly, we found that regardless of the feature used to predict the movement intent of the subjects, they still exhibited similar waveforms except for their amplitudes that appeared to be different. Also, one or two features would be sufficient to achieve acceptable accuracy, which may minimize computational complexity.

Secondly, sixteen different forearm locations were mapped out to determine the most appropriate regions on the forearm for the sensor placement while considering two distinct sensor sizes (small-sized sensors and large-sized sensors). The experimental results showed that the sensors placed on the locations with more muscles led to higher accuracy in comparison to locations with fewer forearm muscles. Also, the small-sized sensors placed around region 6 achieved the highest accuracy as against the sensors placed around region 4 which recorded the lowest accuracy (Figure 6). This analysis is supported by Figure 9, in which the sensors in region 4 (sensors 13 to 16 of Figure 4) were placed in the center of the extensor digitorum, located at the anterior side of the forearm. Compared to the posterior side where there are seven superficial muscles, there are only four muscles in the anterior region. Thus, the sensors in region 4 may acquire less limb motion information in comparison to the other regions [48, 49]. In region 6 (row 2, sensors 2, 6, 10, and 14), the center of these four sensors was placed right on the bulges of all the muscles' bellies that had obvious shape changes when doing different movements, so these sensors may pick up the maximum shape change of the muscles and obtain relatively higher information than the sensors in the other locations. Thus, higher motion recognition accuracy could be achieved when utilizing MSC signals from region 6. For the large-sized sensors, the placement locations were similar to those of the small-sized sensors. Due to the relatively large surface area of the large-sized sensors, they cover more muscles and could capture more MSC information related to muscle activities. For example, sensor 7 of region 3 covers

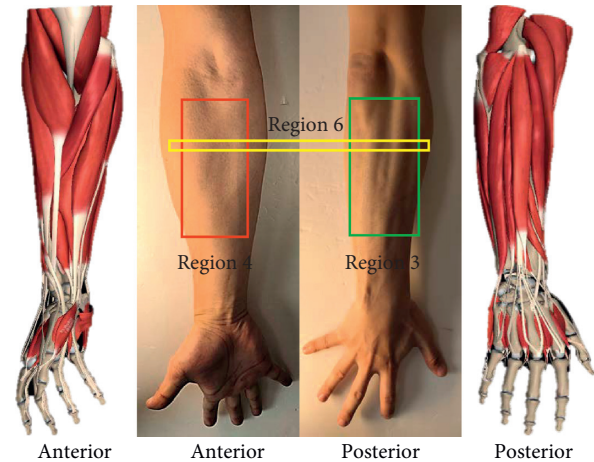


FIGURE 9: The muscles of the forearm.

the brachioradialis, part of the extensor digitorum and part of the flexor carpi ulnaris. This is equivalent to the information obtained from the three small-sized sensors (sensors 11, 7, and 15). On the contrary, this makes the large-sized sensors less sensitive to local muscle information than their small-sized counterparts. This could explain the results presented in Figure 6(b).

Thirdly, two important parameters, the sampling rate and window length associated with processing the MSC signals, which accounts for the computational complexity of the entire motion intention recognition task, were also investigated. In the real-time applications (particularly for embedded microcontrollers) of the motion intention recognitions, a high sampling rate would normally lead to large computation time while a long window length often results in large delay. Thus, it is preferred to develop a system that has a lower sampling rate and adopts a shorter window length for the data processing task. As shown in Figure 7(a), sampling frequencies that are less than 100 Hz could be seen to affect the accuracy of the motion intention recognition classifier because the highest frequency of the MSC signal is between 20 and 50 Hz [50]. According to the Nyquist sampling theorem which states that the sampling frequency of a signal should be at least twice the signal's bandwidth [51], a sampling frequency of 100 Hz would be sufficient to preserve all the relevant information of the MSC signals. Therefore, utilizing sampling frequencies from 100 Hz or above could help maintain the motion intention recognition accuracy of the newly proposed MSC sensors. Meanwhile, Figure 7(b) shows that the window length has little effect on the motion intention recognition accuracy. In addition, it can be observed in Figure 4 that the MSC signal exhibits fewer changes while features extracted from longer windows yielded almost the same performance in terms of motion recognition accuracy compared to those extracted from shorter windows. In other words, varying the window length would only result in a slight increment in the motion intention recognition rate. Thus, the calculation amount can be further reduced by reducing the sampling rate, and the response time of the system can be improved by reducing the window length.

TABLE 4: Running time per subject (unit: ms).

(Hz)	Time for feature extracting		Time for training		Time for classification	
	One feature	Six features	One feature	Six features	One feature	Six features
1000	23.49 ± 6.57	165.8 ± 11.30	23.10 ± 3.88	20.79 ± 1.77	0.98 ± 0.45	0.86 ± 0.26
100	7.45 ± 3.95	33.79 ± 7.80	23.59 ± 3.85	22.15 ± 2.96	0.98 ± 0.37	0.69 ± 0.16

In summary, using the newly proposed MSC signal, one may realize a computationally efficient motion intention recognition system by considering a sampling frequency of about 100 Hz, a window length of about 50 ms, and one feature. The computational complexity of the intention recognition system is estimated as follows. It mainly includes three parts: (a) feature extracting. From Table 1, it can be known that the computational complexity is $O(NM)$, where N is the total number of samples and M is the number of features (in this study, $M < N$); (b) training. According to [52], the computational complexity of LDA is $O(NM^2)$ when $M < N$; (c) classification. This step is the product of the coefficients and the data to be identified, so the computational complexity is $O(NM)$. In total, the computational complexity of the system is about $O(NM^2)$. When one feature is used, the computational complexity is around $O(N)$. In our system, N is a small number (about a few thousand when in training, and about a few hundred when in real-time application), so the system would be easily realized on a microcontroller and the system would be easily realized on a microcontroller. Table 4 shows some results of the running time on our system (computer: Intel i5, Windows 7, MATLAB 2016; the large-sized sensors in region 1 of the nine subjects were tested with a window length of 100 ms). Sampling rate and number of features dramatically affect the time for feature extracting, and lower sampling rate and less features would decrease the computational complexity.

Despite the interesting results obtained in the current study, some issues were observed while analyzing the MSC signals. For instance, compared to the EMG signal, the MSC signal exhibited a relatively simpler waveform characteristic suggesting that it might contain relatively less information. Hence, it may be a challenge to recognize more classes of limb movements with high accuracy when using the proposed MSC signals. According to the work of Li et al. [53], most of the useful information from the EMG recordings for motion intention recognition is contained in the frequency range of 60 Hz to 250 Hz. Meanwhile, the proposed MSC signals have a relatively lower frequency of 50 Hz. Hence, a combination of these two kinds of signals (EMG and MSC signals) may provide complementary information in different frequency bands that would be potential for the development of accurately robust motion intention recognition system in real-life applications particularly when several targeted limb movements are to be decoded.

On the other hand, the MSC signals were characterized by creep [54], which causes crosstalk between the active and nonactive (RS) motion recordings, thus attenuating the motion recognition accuracy. This situation is particularly severe in cases where the RS and HO movements of the subjects are being predicted (Figure 8). Importantly, when data corresponding to the RS were excluded, higher motion

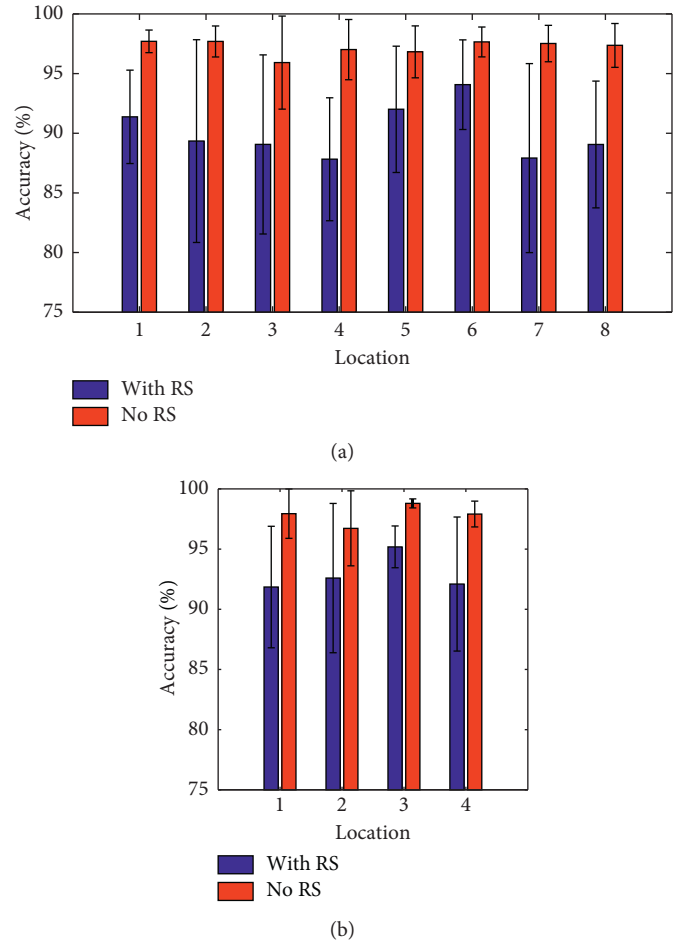


FIGURE 10: The RS was removed. (a) The accuracies of the small-sized sensors. (b) The accuracies of the large-sized sensors.

intention recognition accuracies were achieved for both the small-sized and large-sized sensors across all locations (Figures 10(a) and 10(b)). Because the occurrence of the creep can be modeled using some methods [55], one possible solution would be to develop a creep-sensitive algorithm to reconstruct the MSC signal patterns according to a pre-defined model. Another possible way to resolve this issue would be to consider using EMG signal for detecting the RS states. This is because when the limb assumes a rest state, the amplitude of the EMG signal drops to around the baseline, and afterward, there is an obvious rise in the signal's amplitude when a targeted limb movement is elicited. Therefore, using EMG signal as a switch for the RS state, that is, only using EMG signal to identify the RS state, and using a combination of EMG and MSC signals to identify the active

movement classes may lead to high and stable motion intention recognition in practical applications.

5. Conclusions

In summary, the proposed nanogold flexible and stretchable sensor was developed towards providing an alternative approach for motion intention recognition that may enhance the practical applications of pattern recognition systems. Based on the experimental results obtained in this study, accuracies of up to $95.07 \pm 3.87\%$ and $96.06 \pm 1.84\%$ were recorded for the small-sized and large-sized sensors, respectively, particularly when the sensors were placed at the optimal locations. Additionally, when using the proposed sensor for motion intention recognition, it often does not require a high sampling rate (just over 100 Hz) and a long window length (50 ms) for data processing. Interestingly, the newly proposed MSC sensor is not sensitive to feature set selection, indicating that simple feature methods could be applied to achieve an acceptable motion intention recognition accuracy in practical settings. Thus, this suggests that the proposed nanogold flexible and stretchable sensor would be feasible and effective in practical applications.

Despite the potential results obtained in the current study, there are still some shortcomings that need to be addressed in our future work. For instance, the issue of degradation in the performance of the proposed MSC sensors resulting from the creeping nature of the materials would hopefully be addressed through systematic investigation in our future work either through the development of intelligent signal processing algorithms or via a technique that would combine the MSC and EMG signals for motion intention recognition.

Data Availability

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

Conflicts of Interest

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

Authors' Contributions

Pingao Huang and Hui Wang contributed equally to this work.

Acknowledgments

This study was partly supported by the grants from the National Key R&D Program of China (grant no. 2018YFA0701400), National Natural Science Foundation of China (grant nos. U1613222, 81927804, and 61803361), Ministry of Science and Technology of the People's Republic of China (2016YFE0124100), Shenzhen Science and Technology Plan Project (grant no. JCYJ20160331174854880), and Guangdong Science and Technology Project (2019B090915002). The study was also funded by the

Shenzhen Institute of Artificial Intelligence and Robotics for Society.

References

- [1] S. Patel, H. Park, P. Bonato, L. Chan, M. Rodgers et al., "A review of wearable sensors and systems with application in rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 1, p. 21, 2012.
- [2] P. Rodgers, T. Kirstein, and G. Tröster, "Wearable systems for health care applications," *Methods of Information in Medicine*, vol. 43, no. 43, pp. 232–238, 2004.
- [3] Y. Long, Z.-j. Du, W.-d. Wang, and W. Dong, "Human motion intent learning based motion assistance control for a wearable exoskeleton," *Robotics and Computer-Integrated Manufacturing*, vol. 49, pp. 317–327, 2018.
- [4] H. Dong, S. Lee, S. Kanbe et al., "Power assist method for HAL-3 using EMG-based feedback controller," in *Proceedings of the 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, pp. 1648–1653, Washington, DC, USA, October 2003.
- [5] F. Taffoni, D. Rivera, A. La Camera, A. Nicolò, J.R. Velasco, and C. Massaroni, "A wearable system for real-time continuous monitoring of physical activity," *Journal of Healthcare Engineering*, vol. 2018, Article ID 1878354, 16 pages, 2018.
- [6] A. Pantelopoulou and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 1–12, 2010.
- [7] N. Akhlaghi, C. A. Baker, M. Lahlou et al., "Real-time classification of hand motions using ultrasound imaging of forearm muscles," *Ieee Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1687–1698, Aug 2016.
- [8] K. C. Kong and D. Jeon, "Design and control of an exoskeleton for the elderly and patients," *IEEE-ASME Transactions on Mechatronics*, vol. 11, pp. 428–432, 2006.
- [9] E. H. Zheng, L. Wang, Y. M. Luo et al., "Non-contact capacitance sensing for continuous locomotion mode recognition: design specifications and experiments with an amputee," in *Proceedings of the 2013 IEEE 13th International Conference on Rehabilitation Robotics (Icorr)*, Seattle, WA, USA, June 2013.
- [10] W. S. Kim, H. D. Lee, D. H. Lim et al., "Development of a muscle circumference sensor to estimate torque of the human elbow joint," *Sensors and Actuators A-Physical*, vol. 208, pp. 95–103, 2014.
- [11] H. Han and J. Kim, "Novel muscle activation sensors for estimating of upper limb motion intention," in *Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1–20, pp. 3767–3770, Minneapolis, MN, USA, September 2009.
- [12] R. B. Woodward, S. J. Shefelbine, and R. Vaidyanathan, "Pervasive monitoring of motion and muscle activation: inertial and mechanomyography fusion," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 5, pp. 2022–2033, 2017.
- [13] S. R. Larimi, H. Rezaei Nejad, M. Oyatsi, M. Hoorfar, A. O'Brien, and H. Najjaran, "Low-cost ultra-stretchable strain sensors for monitoring human motion and bio-signals," *Sensors and Actuators A: Physical*, vol. 271, pp. 182–191, 2018.
- [14] M. M. Rodgers, V. M. Pai, and R. S. Conroy, "Recent advances in wearable sensors for health monitoring," *IEEE Sensors Journal*, vol. 15, no. 6, pp. 3119–3126, 2015.

- [15] M. Amjadi, K. U. Kyung, I. Park et al., "Stretchable, skin-mountable, and wearable strain sensors and their potential applications: a review," *Advanced Functional Materials*, vol. 26, no. 11, pp. 1678–1698, 2016.
- [16] D. J. Lipomi, M. Vosgueritchian, B. C.-K. Tee et al., "Skin-like pressure and strain sensors based on transparent elastic films of carbon nanotubes," *Nature Nanotechnology*, vol. 6, no. 12, pp. 788–792, 2011.
- [17] X. Xiao, L. Y. Yuan, J. W. Zhong et al., "High-strain sensors based on ZnO nanowire/polystyrene hybridized flexible films," *Advanced Materials*, vol. 23, no. 45, p. 5440, 2011.
- [18] S. Luo and T. Liu, "SWCNT/Graphite nanoplatelet hybrid thin films for self-temperature-compensated, highly sensitive, and extensible piezoresistive sensors," *Advanced Materials*, vol. 25, no. 39, pp. 5650–5657, 2013.
- [19] M. Amjadi, A. Pichitpajongkit, S. Lee, I. Ryu, and I. Park, "Highly stretchable and sensitive strain sensor based on silver nanowire-elastomer nanocomposite," *Acs Nano*, vol. 8, no. 5, pp. 5154–5163, 2014.
- [20] S. Gong, W. Schwalb, Y. W. Wang et al., "A wearable and highly sensitive pressure sensor with ultrathin gold nanowires," *Nature Communications*, vol. 5, 2014.
- [21] U.-H. Shin, D.-W. Jeong, S.-M. Park, S.-H. Kim, H. W. Lee, and J.-M. Kim, "Highly stretchable conductors and piezo-capacitive strain gauges based on simple contact-transfer patterning of carbon nanotube forests," *Carbon*, vol. 80, pp. 396–404, 2014.
- [22] S. Gong, D. T. H. Lai, B. Su et al., "Highly stretchy black gold E-skin nanopatches as highly sensitive wearable biomedical sensors," *Advanced Electronic Materials*, vol. 1, 2015.
- [23] T. Teramae, D. Kushida, F. Takemori, and A. Kitamura, "Construction of an intelligent massage system based on human skin-muscle elasticity," *Electronics and Communications in Japan*, vol. 94, no. 10, pp. 26–33, 2011.
- [24] X. Z. Kitamura, H. L. Sun, X. Y. Yue et al., "A highly stretchable carbon nanotubes/thermoplastic polyurethane fiber-shaped strain sensor with porous structure for human motion monitoring," *Composites Science and Technology*, vol. 168, pp. 126–132, 2018.
- [25] P. K. Yang, L. Lin, F. Yi et al., "A flexible, stretchable and shape-adaptive approach for versatile energy conversion and self-powered biomedical monitoring," *Advanced Materials*, vol. 27, no. 25, pp. 3817–3824, 2015.
- [26] Y. Wang, L. Wang, T. T. Yang et al., "Wearable and highly sensitive graphene strain sensors for human motion monitoring," *Advanced Functional Materials*, vol. 24, no. 29, pp. 4666–4670, 2014.
- [27] C. S. Boland, U. Khan, C. Backes et al., "Sensitive, high-strain, high-rate bodily motion sensors based on graphene-rubber composites," *Acs Nano*, vol. 8, no. 9, pp. 8819–8830, 2014.
- [28] L. Cai, L. Song, P. S. Luan et al., "Super-stretchable, transparent carbon nanotube-based capacitive strain sensors for human motion detection," *Scientific Reports*, vol. 3, no. 1, 2013.
- [29] T. Yamada, Y. Hayamizu, Y. Yamamoto et al., "A stretchable carbon nanotube strain sensor for human-motion detection," *Nature Nanotechnology*, vol. 6, no. 5, pp. 296–301, 2011.
- [30] S. Zhao, J. Li, D. Cao et al., "Recent advancements in flexible and stretchable electrodes for electromechanical sensors: strategies, materials, and features," *ACS Applied Materials & Interfaces*, vol. 9, no. 14, pp. 12147–12164, 2017.
- [31] S. R. Larimi, H. R. Nejad, M. Hoorfar et al., "Control of artificial human finger using wearable device and adaptive network-based fuzzy inference system," in *Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (Smc)*, pp. 3754–3758, Budapest, Hungary, October 2016.
- [32] E. Roh, B.-U. Hwang, D. Kim, B.-Y. Kim, and N.-E. Lee, "Stretchable, transparent, ultrasensitive, and patchable strain sensor for human-machine interfaces comprising a nano-hybrid of carbon nanotubes and conductive elastomers," *Acs Nano*, vol. 9, no. 6, pp. 6252–6261, Jun 2015.
- [33] S. Lim, D. Son, J. Kim et al., "Transparent and stretchable interactive human machine interface based on patterned graphene heterostructures," *Advanced Functional Materials*, vol. 25, no. 3, pp. 375–383, 2015.
- [34] W. Song, B. Gan, T. Jiang et al., "Nanopillar arrayed triboelectric nanogenerator as a self-powered sensitive sensor for a sleep monitoring system," *Acs Nano*, vol. 10, no. 8, pp. 8097–8103, 2016.
- [35] G. F. Cai, J. X. Wang, K. Qian et al., "Extremely stretchable strain sensors based on conductive self-healing dynamic cross-links hydrogels for human-motion detection," *Advanced Science*, vol. 4, 2017.
- [36] T. Q. Trung, S. Ramasundaram, B. U. Hwang et al., "An all-elastomeric transparent and stretchable temperature sensor for body-attachable wearable electronics," *Advanced Materials*, vol. 28, no. 3, p. 502, 2016.
- [37] J. T. Muth, D. M. Vogt, R. L. Truby et al., "Embedded 3D printing of strain sensors within highly stretchable elastomers," *Advanced Materials*, vol. 26, no. 36, pp. 6307–6312, 2014.
- [38] J. Meyer, P. Lukowicz, and G. Troster, "Textile pressure sensor for muscle activity and motion detection," in *Proceedings of the IEEE International Symposium on Wearable Computers*, Montreux, Switzerland, October 2006.
- [39] Z. Y. Liu, X. T. Wang, D. P. Qi et al., "High-adhesion stretchable electrodes based on nanopile interlocking," *Advanced Materials*, vol. 29, no. 2, 2017.
- [40] T. Lorrain, N. Jiang, and D. Farina, "Influence of the training set on the accuracy of surface EMG classification in dynamic contractions for the control of multifunction prostheses," *Journal of Neuroengineering and Rehabilitation*, vol. 8, no. 1, 2011.
- [41] X. Li, S. Chen, H. Zhang, O. W. Samuel et al., "Towards reducing the impacts of unwanted movements on identification of motion intentions," *Journal of Electromyography and Kinesiology*, vol. 28, pp. 90–98, 2016.
- [42] O. W. Samuel, Z. Hui, X. Li et al., "Pattern recognition of electromyography signals based on novel time domain features for amputees' limb motion classification," *Computers & Electrical Engineering*, vol. 67, pp. 1–10, 2018.
- [43] O. W. Samuel, Y. Geng, X. Li et al., "Towards efficient decoding of multiple classes of motor imagery limb movements based on EEG spectral and time domain descriptors," *Journal of Medical Systems*, vol. 41, p. 194, 2017.
- [44] H. Huang, F. Zhang, L. J. Hargrove et al., "Continuous locomotion-mode identification for prosthetic legs based on neuromuscular-mechanical fusion," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2867–2875, 2011.
- [45] T. A. Kuiken, G. L. Li, B. A. Lock et al., "Targeted muscle reinnervation for real-time myoelectric control of multi-function artificial arms," *Jama-Journal of the American Medical Association*, vol. 301, no. 6, pp. 619–628, 2009.
- [46] H. Huang, P. Zhou, G. Li et al., "An analysis of EMG electrode configuration for targeted muscle reinnervation based neural machine interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 1, pp. 37–45, 2008.

- [47] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
- [48] G. G. Davis, *Applied Anatomy: The Construction of the Human Body Considered in Relation to its Functions; Diseases and Injuries*, J. B. Lippincott & Co., Philadelphia, PA, USA, 1918.
- [49] R. O'Rahilly and F. Müller, *Basic Human Anatomy: A Regional Study of Human Structure*, WB Saunders Company, Philadelphia, PA, USA, 1983.
- [50] T. W. Beck, "Applications of Mechanomyography for examining muscle function," in *Technical Aspects of Surface Mechanomyography*, pp. 95–107, Transworld Research Network, Trivandrum, India, 2010.
- [51] H. J. Landau, "Sampling, data transmission, and the Nyquist rate," *Proceedings of the IEEE*, vol. 55, no. 10, pp. 1701–1706, 1967.
- [52] T. G. Alaa Tharwat, A. Ibrahim, and A. E. Hassanien, "linear discriminant analysis: a detailed tutorial," *AI Communications*, vol. 30, no. 2, p. 22, 2018.
- [53] G. Li, Y. Li, L. Yu, and Y. Geng, "Conditioning and sampling issues of EMG signals in motion recognition of multifunctional myoelectric prostheses," *Annals of Biomedical Engineering*, vol. 39, no. 6, pp. 1779–1787, Jun 2011.
- [54] M. R. Vegas and J. L. Martin del Yerro, "Stiffness, compliance, resilience, and creep deformation: understanding implant-soft tissue dynamics in the augmented breast: fundamentals based on materials science," *Aesthetic Plastic Surgery*, vol. 37, no. 5, pp. 922–930, Oct 2013.
- [55] K. Le Phan, "Methods to correct for creep in elastomer-based sensors," in *Proceedings of the SENSORS*, pp. 1119–1122, IEEE, Lecce, Italy, October 2008.

Research Article

Using Eye Aspect Ratio to Enhance Fast and Objective Assessment of Facial Paralysis

Jialing Feng ¹, Zhexiao Guo ^{1,2}, Jun Wang ³ and Guo Dan ^{1,2}

¹School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China

²Shenzhen Institute of Neuroscience, Shenzhen, China

³Shenzhen Hospital, Southern Medical University, Shenzhen, China

Correspondence should be addressed to Guo Dan; danguo@szu.edu.cn

Received 25 November 2019; Accepted 27 December 2019; Published 29 January 2020

Guest Editor: Mirian C. D. Pinheiro

Copyright © 2020 Jialing Feng 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.

A rapid and objective assessment of the severity of facial paralysis allows rehabilitation physicians to choose the optimal rehabilitation treatment regimen for their patients. In this study, patients with facial paralysis were enrolled as study objects, and the eye aspect ratio (EAR) index was proposed for the eye region. The correlation between EAR and the facial nerve grading system 2.0 (FNGS 2.0) score was analyzed to verify the ability of EAR to enhance FNGS 2.0 for the rapid and objective assessment of the severity of the facial paralysis. Firstly, in order to accurately calculate the EAR, we constructed a landmark detection model based on the face images of facial paralysis patients (FP-FLDM). Evaluation results showed that the error rate of facial feature point detection in patients with facial paralysis of FP-FLDM is 17.1%, which was significantly superior to the landmark detection model based on normal face images (NF-FLDM). Secondly, in this study, the Fréchet distance was used to calculate the difference in bilateral EAR of facial paralysis patients and to verify the correlation between this difference and the corresponding FNGS 2.0 score. The results showed that the higher the FNGS 2.0 score, the greater the difference in bilateral EAR. The correlation coefficient between the bilateral EAR difference and the corresponding FNGS 2.0 score was 0.9673, indicating a high correlation. Finally, through a 10-fold crossvalidation, we can know that the accuracy of scoring the eyes of patients with facial paralysis using EAR was 85.7%, which can be used to enhance the objective and rapid assessment of the severity of facial paralysis by FNGS 2.0.

1. Introduction

Facial paralysis can lead to the loss of autonomic motor function of the unilateral mimetic muscles of the face. Accurate and objective evaluation of the degree of facial nerve motor function damage is key for the treatment and rehabilitation of facial paralysis patients. Clinically, the House–Brackmann grading system (HBGS) [1] is used to evaluate and grade facial nerve motor dysfunction in patients with facial paralysis. Because the description of symptoms of adjacent facial paralysis grades in HBGS is ambiguous, the evaluation is prone to subjective differences. Moreover, HBGS cannot reflect motor dysfunction in the local regions of the eyebrow, eye, mouth, and nose, which affects the design of follow-up treatment and rehabilitation regimen for facial paralysis patients. Therefore, in 2009, the

American Academy of Otorhinolaryngology-Head and Neck Surgery Facial Neuropathy proposed the facial nerve grading scale 2.0 (FNGS 2.0) [2]. FNGS 2.0 assesses the severity of facial paralysis by manually measuring facial features in images, which makes it difficult to rapidly and objectively determine the severity of facial paralysis. The development of an objective and quantitative automated method to rapidly assess the severity of facial paralysis is extremely important for the design of subsequent rehabilitation regimens.

To objectively assess the severity of facial paralysis, Tomat and Manktelow [3] developed a measurement system to assess the asymmetry of local facial regions targeted at the displacement and angle changes of the landmarks of the still mouth and smiling mouth in 2005. In 2010, Liu et al. [4] captured images of facial paralysis patients with different

facial expressions, compared the shape of the region and area change of the eye, forehead, nose, and mouth, and calculated these changes to evaluate the degree of facial nerve injury. In 2016, Wang et al. [5] proposed a facial symmetry quantitative evaluation method based on the principle of local mirror symmetry. The method objectively and quantitatively assesses the severity of facial paralysis through the 3 steps of local region localization, extraction of asymmetric features, and quantification of asymmetry of the bilateral face. All three of them quantified the features of local facial regions to evaluate the symmetry of facial structures. However, Tomat's work measured the symmetry of the mouth of normal people without considering other local facial areas. Liu's work could not achieve real-time evaluation; Wang's work was not applied to studies related to facial paralysis. Therefore, the present study refers to the research ideas of Wang et al. to propose quantitative indicators for local regions of the face in real time to objectively and quantitatively assess the severity of the corresponding regions of facial paralysis. How to accurately obtain 2 d or 3 d facial information of patients with facial paralysis and evaluate the severity of facial paralysis based on this characteristic information is the main challenge. Based on the coordinates of feature points to extract facial information, the accurate detection of feature points and quantitative evaluation of facial structure and motor function symmetry are the main problems to be solved. Therefore, this study trained a new feature point detection model to improve the accuracy of feature point detection and proposed a regional index to quantify the motion symmetry of facial regions.

In FNGS 2.0, facial paralysis patients need to complete the facial expressions of the corresponding regions for evaluation. The corresponding regions include the eyebrows, eyes, nose, and mouth, and common facial expressions include opening eyes, closing eyes, baring teeth, and bulging mouth. The movements of the eyes and mouth better manifest the asymmetry of facial expressions caused by facial nerve injury of the affected side in facial paralysis patients than those of the eyebrows and nose, and the amplitude of movements on both sides of the eye is easier to observe and compare than those in the mouth region (Figure 1). In addition to damaging the facial features, facial paralysis also has a great impact on the patient's eyes. Facial paralysis may cause enlarged ocular fissures and an inability to close the eyes or blink. In particular, prolonged exposure of the cornea during sleep can cause dryness of the cornea, which can easily lead to eye infection and damaged vision in severe cases [6]. Studies of the objective and rapid assessment of eye motor function of facial paralysis patients are needed to develop optimal treatment regimens for facial paralysis patients to help them recover as quickly as possible. Therefore, this study proposes the eye aspect ratio (EAR) index, which targets the eye region of facial paralysis patients and experimentally analyzes the feasibility of applying the difference in bilateral EAR in the objective and rapid assessment of the severity of facial paralysis.

Common facial landmark detection models are constructed based on normal facial databases, such as 300-W [7] and AFLW [8], and are thus not suitable for face images of

facial paralysis patients. The detection results of such models for closed-eye images are shown in Figure 2. Facial paralysis patients exhibit strange facial expressions due to facial nerve injury and cannot open and close eyes, bare teeth and bulge mouth normally on the affected side. Therefore, existing facial landmark detection models based on normal face images are not suitable for the calculation of EAR.

As shown in Figure 3, in this study, to verify the feasibility of EAR in enhancing FNGS 2.0 to rapidly and objectively assess the eye motor function of facial paralysis patients, we first constructed a dataset of facial landmark images of facial paralysis patients. A facial landmark detection model for facial paralysis patients (FP-FLDM) was constructed using this dataset. The landmark detection model was used to acquire information on eye landmarks in the movement image series to calculate the EAR index during the eye movement process of facial paralysis patients. The Fréchet distance of the EAR index curve was calculated to represent the difference in bilateral EAR. Finally, the correlation between the difference in bilateral EAR and the corresponding FNGS 2.0 score was analyzed by experiments to prove the feasibility of EAR in enhancing FNGS 2.0 to objectively and rapidly assess eye motor function.

2. Methods

2.1. Facial Nerve Grading Scale 2.0(FNGS2.0). To address the inability of HBGS to reflect local motor dysfunction of the eyebrow, eye, mouth, and nose and the overlapping and inaccurate evaluation rules of adjacent facial paralysis grades, the American Academy of Otorhinolaryngology-Head and Neck Surgery Facial Neuropathy proposed FNGS 2.0 in 2009. FNGS 2.0 can assess changes in the eyebrow, eye, nose, and mouth regions on a scale of 1–6 according to the degree of motor dysfunction and can evaluate the severity of complications of facial paralysis on a scale of 0–3. The sum of these two scores is used to evaluate the severity of facial paralysis. The specific scoring rules of FNGS 2.0 are shown in Table 1.

2.2. Facial Landmark Detection Model. The cascaded regression model is a mapping function that learns directly from the facial appearance to the facial shape (or the parameters of the facial shape model) and then establishes the correspondence from the appearance to the shape. This method does not require complex facial shapes, and appearances for modeling is simple and efficient and achieves good positioning effects in controllable scenarios (human faces collected under laboratory conditions) and noncontrollable scenarios (face images from the Internet) [9]. The basis of the tree algorithm is the decision tree. The decision tree is widely used in statistics, data mining, and machine learning because it is easy to understand, easy to construct, and rapid [10]. Therefore, this study employed the ensemble of regression tree (ERT) algorithm, a regression tree method based on gradient boosting learning. The residual regression tree (gradient boosting decision tree, GBDT) algorithm is used to construct each level of regressors, where each

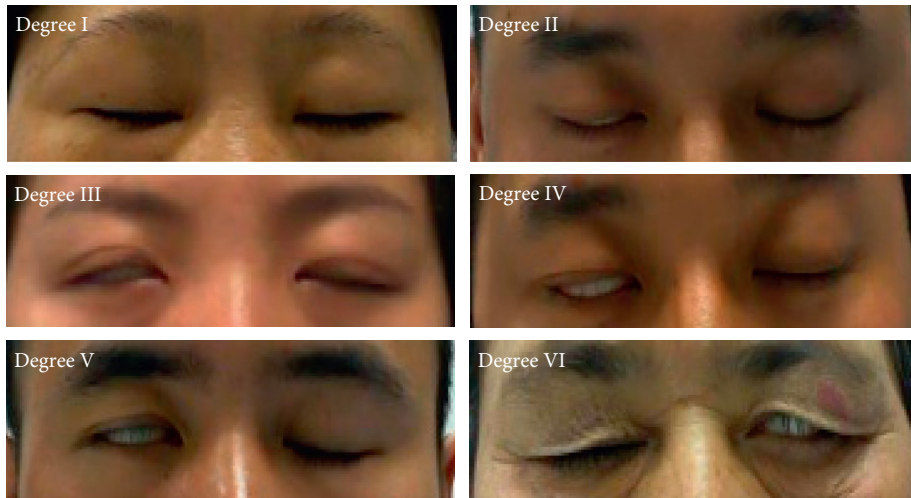


FIGURE 1: Eye region graded by severity in FNGS 2.0.

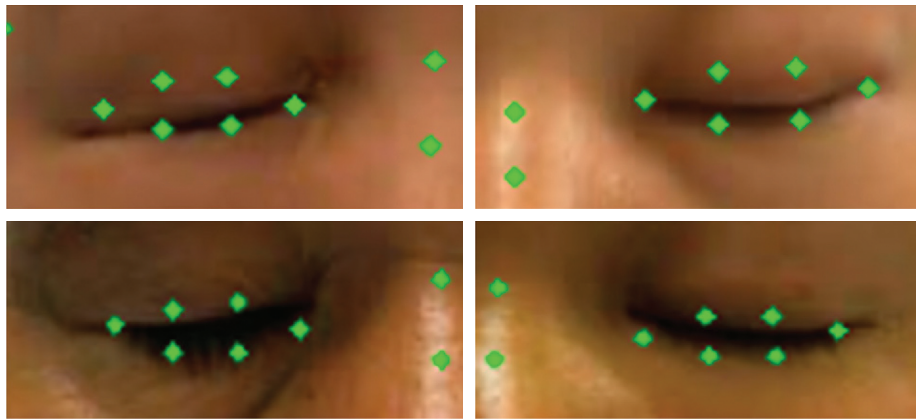


FIGURE 2: Detection results of closed-eye images in patients with facial paralysis by the landmark detection model based on normal face images: (a) Left eye (b) Right eye.

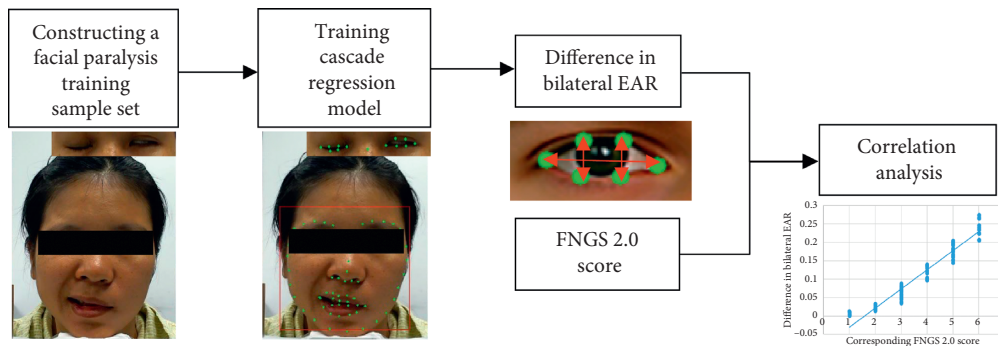


FIGURE 3: Study design and flowchart.

regressor is composed of multiple decision trees, and the parameters of each decision tree are obtained according to the residual between the current landmark distribution and the actual landmark distribution and through training of randomly selected pixel pairs. The actual landmark distribution in this paper is shown in Figure 4.

The construction process of the model mainly includes 3 main modules: generating the training sample set, training the optimal weak regressor, and updating the training sample set. The marked face images of facial paralysis patients were used as a training set, and the initial positions of the facial landmarks in each face image were randomly

TABLE 1: Facial nerve grading scale 2.0.

Score	Region			
	Brow	Eye	NLF	Oral
1	Normal	Normal	Normal	Normal
2	Slight weakness >75% of normal	Slight weakness >75% of normal, Complete closure with mild effort	Slight weakness >75% of normal	Slight weakness >75% of normal
3	Obvious weakness >50% of normal, Resting symmetry	Obvious weakness >50% of normal, Complete closure with maximal effort	Obvious weakness >50% of normal, Resting symmetry	Obvious weakness >50% of normal, Resting symmetry
4	Asymmetry at rest <50% of normal, Cannot close completely	Asymmetry at rest <50% of normal	Asymmetry at rest <50% of normal	Asymmetry at rest <50% of normal
5	Trace movement	Trace movement	Trace movement	Trace movement
6	No movement	No movement	No movement	No movement
Secondary movement (global assessment)				
Score	Degree of movement			
0	None			
1	Slight synkinesis; minimal contracture			
2	Obvious synkinesis; mild to moderate contracture			
3	Disfiguring synkinesis; severe contracture			
Reporting: sum scores for each region and secondary movement				
Grade	Total score			
I	4			
II	5–9			
III	10–14			
IV	15–19			
V	20–23			
VI	24			

NLF, nasolabial fold.

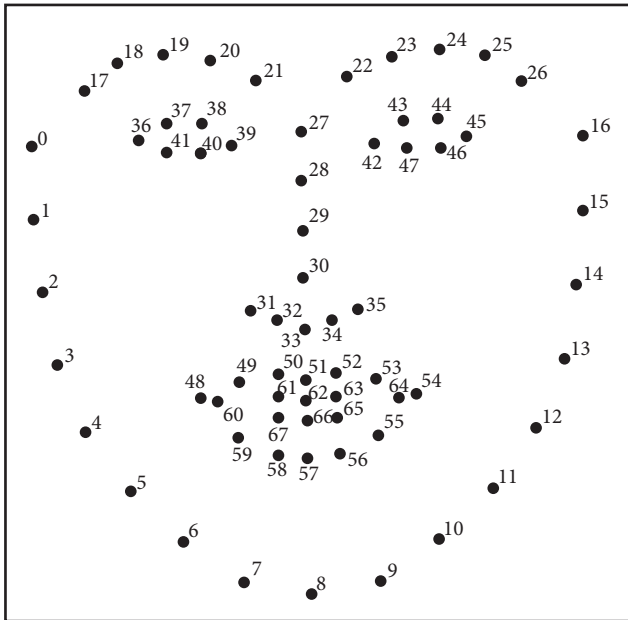


FIGURE 4: The 68 points mark-up used for actual landmark distribution [7].

generated as the training sample T . The GBDT algorithm was used to reduce the normalized mean error (NME) between the current landmark distribution coordinates and the actual landmark distribution coordinates [11], the least squares method was used to minimize errors, the cascade regression factor of each level was obtained, and the optimal weak

regressor was obtained through continuous training. The NME can be written as follows:

$$\text{NME} = \frac{1}{N} \sum_{i=1}^N \frac{\|z_i - s_i\|^2}{d_i}, \quad (1)$$

where z_i denotes the current landmark distribution coordinates of the landmarks of the face, s_i is the actual landmark distribution coordinates and d_i is the interocular distance.

The core formula is shown below, where $\hat{S}^{(t)}$ represents the shape of the t -th regressor and r_t represents the update of the t -th regressor.

$$\hat{S}^{(t+1)} = \hat{S}^{(t)} + r_t(I, \hat{S}^{(t)}). \quad (2)$$

The detection process of the model was to first initialize the input eye landmarks, extract eye features according to the current landmark distribution, input the extracted features into the established weak regressor, and update the current landmark distribution according to the residuals of the landmark distribution, which contains the eye landmark information required for calculating the EAR index.

2.3. Calculation of the Difference in Bilateral EAR. EAR refers to the aspect ratio of the eye region, which is often used to calculate the temporal consistency and speed of left and right eye blinks [12] and in fatigue detection [11,13–15]. There has been no report of the application of EAR to the objective assessment of facial paralysis. In this study, EAR was used to characterize the displacement changes of the eye landmarks

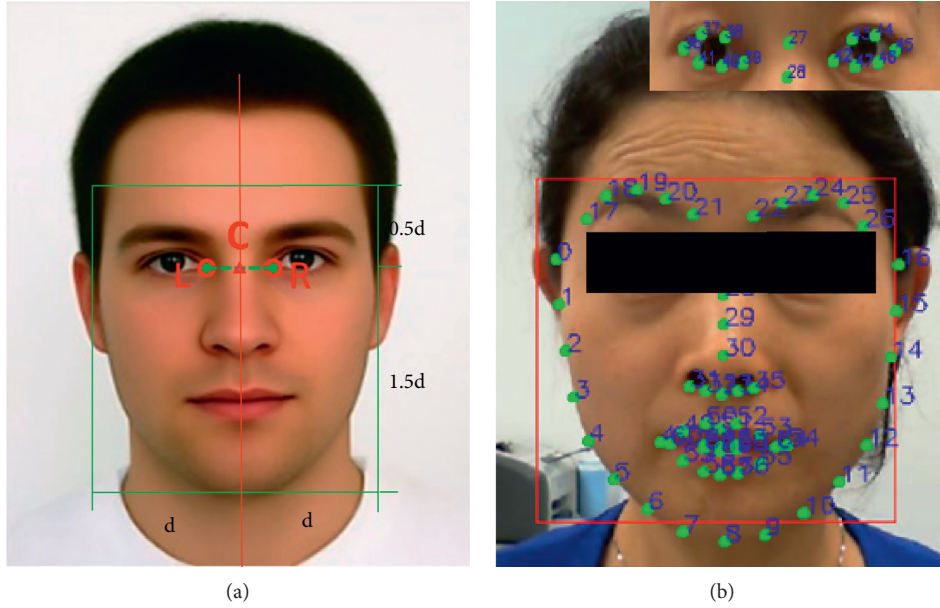


FIGURE 5: (a) Normalized geometric model of the face; (b) landmark detection model index numbers.

of facial paralysis patients in the movement image series, and then the difference in bilateral eye movements of facial paralysis patients was evaluated according to the changes in EAR.

In both normal face images and facial paralysis face images, the landmarks of the nasal root and the nose tip on the face changed little during the movements (displacement changes of less than 2 pixels). To increase the difference in bilateral eye movements to improve the performance of the difference in EAR characteristics and based on the observation that the movements of the bilateral eyebrows were also inconsistent during the process of eye opening and closing in facial paralysis patients, the Euclidean distance from the centre of the eyebrows to the tip of the nose divided by the length of the nose was used as an amplification factor for EAR [16]. The formula for calculating the one-side EAR is as follows: in the t -frame image, the EAR of the left eye is EAR_{Lt} , the EAR of the right eye is EAR_{Rt} , and LM_n is the landmark corresponding to the index n :

$$EAR_{Lt} = \frac{\|LM_{37} - LM_{41}\| + \|LM_{38} - LM_{40}\|}{2 * \|LM_{36} - LM_{39}\|} * \frac{\|LM_{19} - LM_{33}\|}{\|LM_{27} - LM_{33}\|},$$

$$EAR_{Rt} = \frac{\|LM_{43} - LM_{47}\| + \|LM_{44} - LM_{46}\|}{2 * \|LM_{42} - LM_{45}\|} * \frac{\|LM_{24} - LM_{33}\|}{\|LM_{27} - LM_{33}\|}. \quad (3)$$

In addition, when facial paralysis patients perform facial expressions, their heads often swing. Direct calculation of the landmark displacement of each frame in the movement image series will result in a large error, and thus it is necessary to perform tilt correction on the faces in the images. Moreover, because the size of the face area of each subject is different, geometric normalization of the faces is required [17]. According to the coordinate values of the two eye corners, the face was rotated to calibrate the tilt. The distance

between the two eyes was set as d , and the midpoint was C . In the rectangular feature area of the face, with point C as the reference, the distance to each side was set as d , and distances of $0.5d$ and $1.5d$ on the vertical direction were set. The normalized geometric models of the face and facial landmarks are shown in Figure 5.

To establish the relationship between EAR and the FNGS 2.0 score, the Fréchet distance was used to represent the difference in bilateral EAR. In 1994, Eiter and Mannila proposed the definition of the discrete Fréchet distance [18]: by calculating the Euclidean distance between the pair of sequential points of two trajectory curves, the maximum value of the sequence is selected as the spatial similarity of the two curves. The core formula is as follows:

$$\max \text{distance}_{(P,Q)} = \max \left[\sqrt{\sum_{i=1}^n (p_i - q_i)^2} \right], \quad (4)$$

where p_i denotes the i -th EAR difference P on left side and q_i is the i -th EAR difference Q on right side.

In this study, the Fréchet distance was used to represent the bilateral EAR difference of facial paralysis patients. The larger the Fréchet distance, the smaller the similarity of the EAR, that is, the greater the difference in the movements of the bilateral eye regions.

3. Experiments

3.1. Data and the Construction of the Facial Landmark Detection Model (FLDM). The facial image data of facial paralysis patients in this study were derived from the literature [19]. Using FNGS 2.0, rehabilitation doctors of the collaborating hospitals objectively evaluated the eye motor function of 105 patients with facial paralysis with scores of 1–6 corresponding from normal to severe based on the

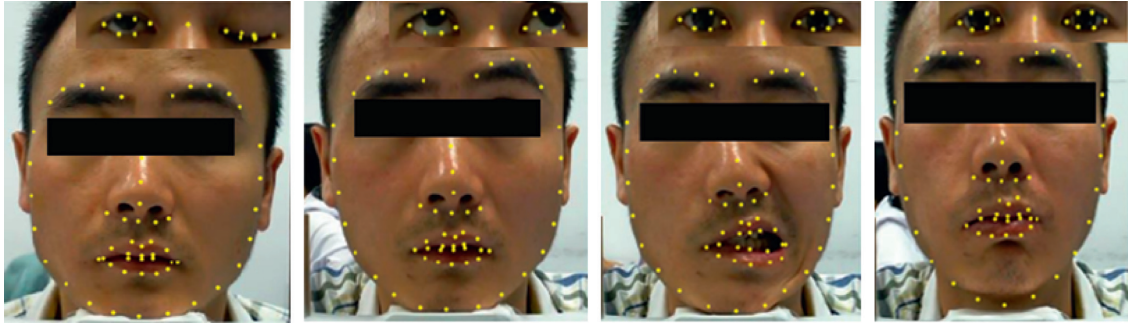


FIGURE 6: Marked data sets for closed eye, open eye, baring teeth, and bulging mouth from left to right.

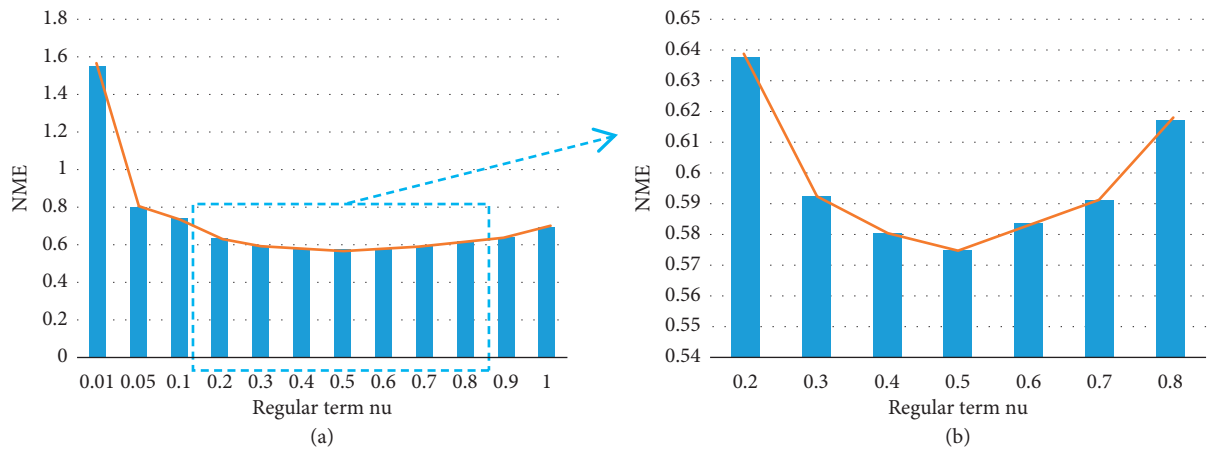


FIGURE 7: (a) Model error column graph with a tree depth of 5 and regular term parameters from 0.01 to 1.0; (b) model error column graph with a tree depth of 5 and regular term parameters from 0.2 to 0.8.

severity of paralyzed eyes: 5 patients received scores of 1 point, 21 patients 2 points, 31 patients 3 points, 13 patients 4 points, 23 patients 5 points, and 12 patients 6 points. The major facial muscles involved in facial paralysis are the occipital frontal muscle, frontal abdomen, orbicularis oculi muscle, zygomatic muscle, oral muscle, and orbicularis oculi muscle. Opening the eyebrows, closing the eyes, baring the teeth, and bulging the mouth help to train these major muscles and help to restore normal motor function of the entire facial expression muscle [20]. As a result, to construct the FLDM, we extracted 4 images each of open eye, closed eye, baring teeth, and bulging mouth from the image series of 105 facial paralysis patients and marked 68 facial landmarks on a total of 420 facial paralysis images. The marked results are shown in Figure 6.

To evaluate the accuracy of the model detection results, the marked data sets were divided into a training set and a testing set containing 336 images and 84 images, respectively. Due to the small number of samples in the training set, to prevent overfitting of the model, data amplification of the training set was necessary. Kazemi and Sullivan [21] trained and tested the cascading regression tree model on the Helen [22] dataset, with 2000 training sets and 330 test sets. We have 336 training data and 84 test data. In order to keep the data volume basically consistent with the reference model, our study amplified the random deformation of the dataset by 10 times. The amplified training set was input into

the cascade regressor to perform weak regressor training to generate the model required for subsequent study. We set the decision tree depth parameters as 2, 4, 5, and 10 and the regular term parameters as 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. Paired combinations of the decision tree depth parameters and regular term parameters were used to construct the FLDM, and finally the model with the highest accuracy was selected for subsequent study.

3.2. Evaluation of the FLDM. FLDMs commonly use the NME to evaluate model accuracy [23]. In this study, the NMEs between the model detection values and the standard values of the eye landmark coordinates were calculated as the evaluation standard of the model training effect. Figure 7(a) shows the model NME output when the tree depth was 5 and the regular term coefficient distribution was adjusted from 0.01 to 1. Figure 7(b) shows a partial enlargement of Figure 7(a). Figure 7 shows that the accuracy of the model is highest when the tree depth is 5 and the regular term parameter is 0.5.

In 2014, Kazemi et al. constructed a cascaded regression tree landmark detection model using normal face data (NH-FLDM). The representative algorithm based on cascading regression is the cascading regression tree model, which is used to learn the mapping function from face appearance to face shape. Our model also uses regression model to predict facial feature points, so it is meaningful to compare with the

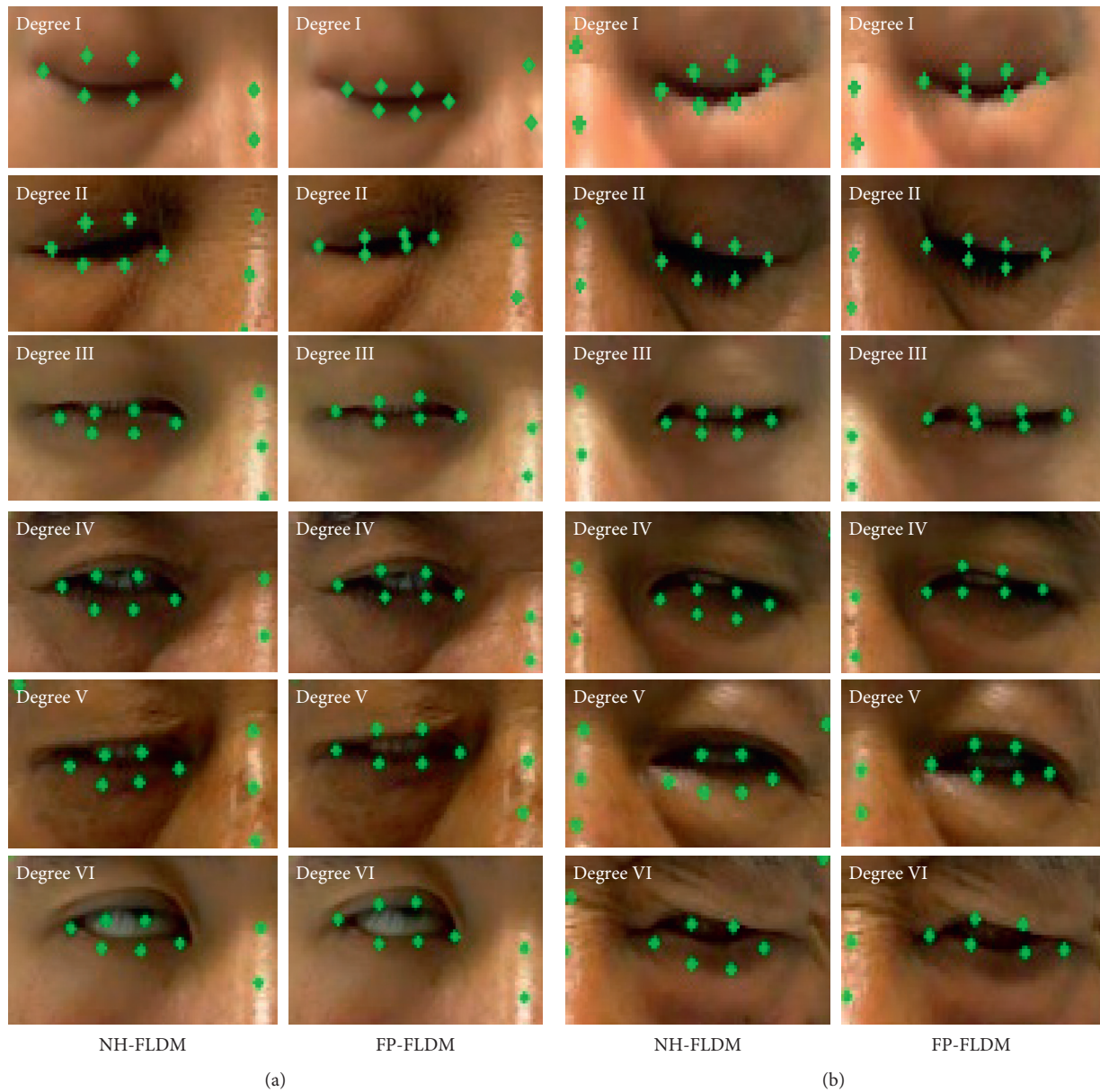


FIGURE 8: Comparison of NH-FLDM and FP-FLDM in the detection of the closed-eye landmarks of patients with facial paralysis: (a) left eye is the affected side and (b) right eye is the affected side.

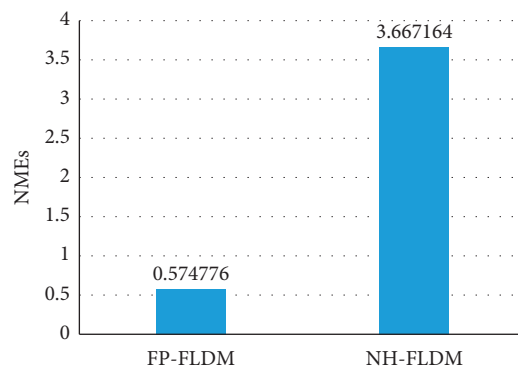


FIGURE 9: Comparison of average NMEs of NH-FLDM and FP-FLDM in eye detection in the movement image series.

TABLE 2: Statistics of incorrect detection of eye landmarks in the movement image series by the two models.

FNGS2.0 score	1	2	3	4	5	6	Total
NH-FLDM (%)	1.9	4.8	5.7	2.9	5.7	6.7	27.6
FP-FLDM (%)	0	2.9	5.7	1.9	4.8	1.9	17.1

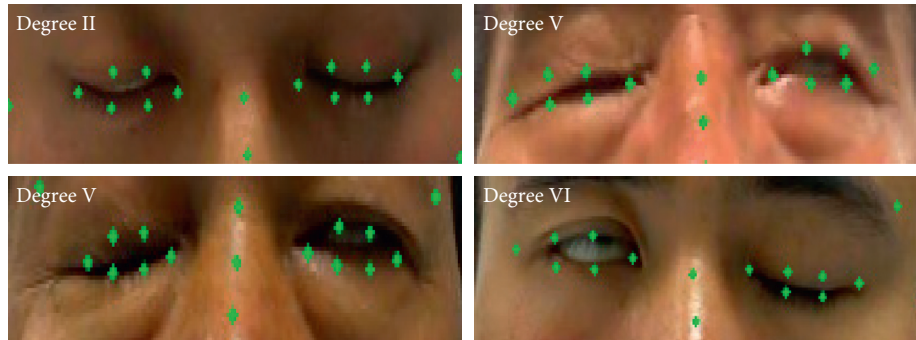


FIGURE 10: Sample representation of poor detection results of closed eye by FP-FLDM.

TABLE 3: Quantity distribution of the original samples and the experimental samples with different FNGS 2.0 scores.

FNGS2.0 score	1	2	3	4	5	6
Original sample	5	21	31	13	23	12
Experimental sample	5	18	25	11	18	10

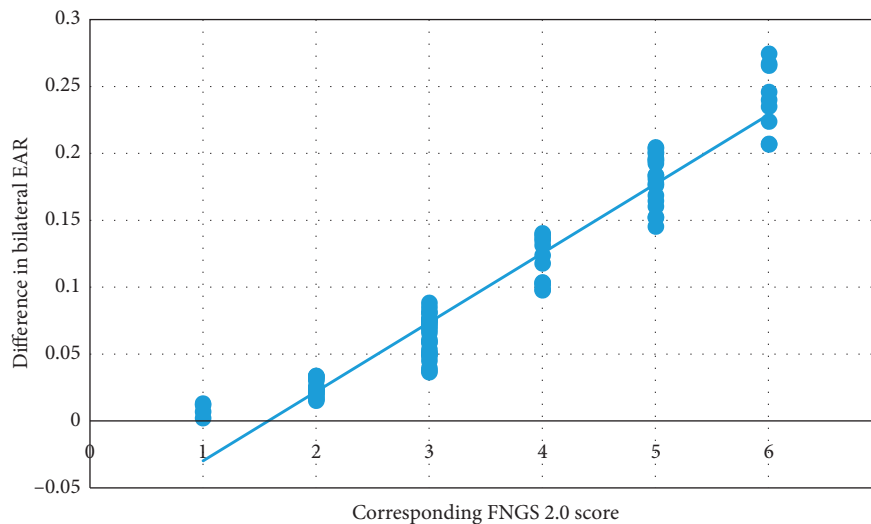


FIGURE 11: Scatter plot of the difference in bilateral EAR of the 87 samples with the corresponding FNGS 2.0 scores.

mainstream regression model. The detection accuracy of NH-FLDM was compared with that of our constructed model (FP-FLDM) on the eye landmarks of facial paralysis patients in the test set. Figure 8(a) and Figure 8(b) show the results of eye landmark detection of the affected side of facial paralysis using NH-FLDM and FP-FLDM, respectively. Figure 9 shows the comparison of the NMEs of NH-FLDM and FP-FLDM in the detection of eye landmarks of facial paralysis patients. Table 2 shows the statistical analysis of the eye landmark detection effects of NH-FLDM and FP-FLDM on movement image series in the test set.

3.3. Calculation and Analysis of the Bilateral EAR Difference. To analyze the correlation between EAR and the FNGS 2.0 score, we excluded 18 samples incorrectly detected by FP-FLDM. The exclusion criteria are shown in Figure 10. Table 3 shows the quantity distribution of different FNGS 2.0 scores of the original samples and the remaining 87 samples. Figure 11 shows the scatter plot and fitted curve of the bilateral EAR difference of the 87 samples versus the corresponding FNGS 2.0 scores. Figure 12 shows the box plot of the bilateral EAR difference of the 87 samples with the corresponding FNGS2.0 scores. Figure 13 shows a more

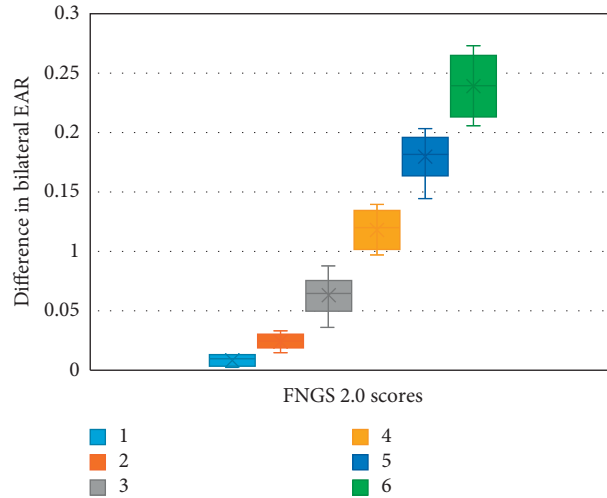


FIGURE 12: Box plot of the distribution of bilateral EAR differences corresponding to different FNGS 2.0 scores.

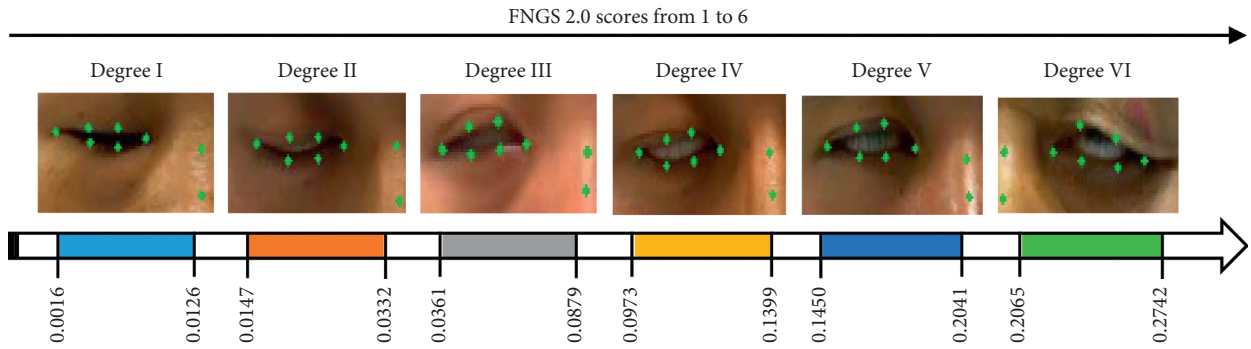


FIGURE 13: Differences in bilateral EAR and closed-eye images corresponding to different FNGS 2.0 scores.

intuitive representation of the distribution of the bilateral EAR difference of facial paralysis patients according to different FNGS 2.0 scores. Table 4 shows the results of correlation analysis between the FNGS 2.0 scores and the corresponding bilateral EAR differences.

4. Discussion

To verify the feasibility of EAR in enhancing FNGS 2.0 to rapidly and objectively assess the severity of the corresponding region of facial paralysis, the correlation between the bilateral EAR difference and the FNGS 2.0 score was calculated in the final part of the experiment. Table 4 shows that the correlation coefficient between the two was 0.9673, indicating that the bilateral EAR difference is highly correlated with the FNGS 2.0 score. As shown in Figures 12 and 13, as the FNGS 2.0 score increases, that is, the degree of eye movement dysfunction becomes more severe, the differences in bilateral EAR gradually increase. The interval of bilateral EAR differences corresponding to different FNGS 2.0 scores was significant, indicating that the bilateral EAR difference can easily distinguish the corresponding FNGS 2.0 score. Moreover, this study was performed on a 3.4 GHz PC based on the Python platform with a processing speed of 30

frames per second, indicating that these indicators can rapidly assess the severity of the corresponding region of facial paralysis. Therefore, in summary, the bilateral EAR difference can objectively and rapidly assess the severity of facial paralysis.

In addition, this study analyzed the evaluation results of 105 cases of facial paralysis with reference to the research method of Lee et al. [24]. The comparison of regional scores and facial paralysis grading showed that both HBGS and FNGS 2.0 had the highest consistency on eye movement function scores and facial paralysis grades, followed by the mouth (Table 5), suggesting that the examination of movement differences in the eye region is meaningful for the objective classification of facial paralysis.

Furthermore, Table 2 shows that in the detection of eye landmarks in the movement image series, the incorrect detection rate of NH-FLDM was 27.6% and the incorrect detection rate of FP-FLDM was 17.1%. In addition, patients with severe facial paralysis have extremely abnormal facial features and are considered difficult samples for landmark detection models. When the NH-FLDM detected the eye landmarks of facial paralysis patients with scores of 5 and 6, the incorrect detection rates were 5.7% and 6.7%, respectively, while those of FP-FLDM were only 4.8% and 1.9%,

TABLE 4: Correlation analysis between FNGS 2.0 scores and bilateral EAR differences.

FNGS2.0 score	1	2	3	4	5	6
Median of bilateral EAR difference	0.0071	0.0240	0.0620	0.1186	0.1746	0.2404
Correlation coefficient	0.9673					

TABLE 5: Consistency of facial paralysis grading and HBGS, and FNGS 2.0 scores in different regions ($n = 105$).

	Eyebrow	Eye	Mouth	Nasolabial groove
HBGS	47(44.8%)	53(50.5%)	52(49.5%)	42(40.0%)
FNGS 2.0	52(49.5%)	60(57.1%)	55(52.4%)	43(41.0%)

respectively, indicating a significantly better effect of FP-FLDM compared with NH-FLDM. These results imply that it is essential to construct a facial landmark detection model for facial paralysis patients and that the eye landmark detection model constructed in this study is suitable for patients with facial paralysis.

As an exploratory study, several experiments await completion in the future. Table 5 shows that the severity of the eye region of facial paralysis patients cannot fully represent the final score of the facial paralysis patient. Therefore, we still need to explore the corresponding indicators of the eyebrow, nose, and mouth to enhance FNGS 2.0 scoring as a fast and accurate objective assessment for facial paralysis based on a small number of indicators. In addition, although the performance of FP-FLDM was superior to NH-FLDM in facial paralysis patients, Table 1 shows that 17.1% of patient samples were still not accurately detected. Therefore, the performance of the eye landmark detection model constructed in this paper awaits further improvement.

In this study, a 10-fold cross-validation method was used to randomly divide 87 samples into 10 samples. One percent of the samples was selected for each test, and the other 10% was used to construct the EAR feature dataset. The eigenvalues and FNGS2.0 scores were divided into six categories. The process was repeated 10 times, and the average accuracy was 85.2%.

5. Conclusion

In this study, face image data of facial paralysis patients were used as a training set, and a facial landmark detection model for facial paralysis patients was constructed to calculate the EAR. The incorrect detection rate of this model for the eye region of facial paralysis patients was 17.1%, an improvement over currently available detection models. Detection and calculation using the constructed detection model yielded the difference in bilateral EAR of 87 facial paralysis patients. Subsequently, the relationship between the FNGS 2.0 score and the corresponding bilateral EAR difference was analyzed, and the correlation analysis results showed that the two were highly correlated, with a correlation coefficient of 0.9673. By implementing the proposed method and related experiments, this study proves that the bilateral EAR difference can be applied to the objective and rapid assessment of the severity of facial paralysis.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study was supported by Shenzhen Science and Technology Innovation Council (JCYJ20180507182040213, JCYJ20170818141853626), Guangdong Key Basic Research grant (2018B030332001), and Guangdong Pearl River Talents Plan (2016ZT06S220).

References

- [1] T. S. Kang, J. T. Vrabec, N. Giddings, and D. J. Terris, "Facial nerve grading systems (1985–2002): beyond the house-brackmann scale," *Otology & Neurotology*, vol. 23, no. 5, pp. 767–771, 2002.
- [2] J. T. Vrabec, D. D. Backous, H. R. Djalilian et al., "Facial nerve grading system 2.0," *Otolaryngology Head and Neck Surgery*, vol. 140, no. 4, pp. 445–450, 2009.
- [3] L. R. Tomat and R. T. Manktelow, "Evaluation of a new measurement tool for facial paralysis reconstruction," *Plastic and Reconstructive Surgery*, vol. 115, no. 3, pp. 696–704, 2005.
- [4] L. Liu, G. Cheng, J. Dong, S. Wang, and H. Qu, "Evaluation of facial paralysis degree based on regions," in *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, pp. 514–517, IEEE Computer Society, Washington, DC, USA, January 2010.
- [5] T. Wang, S. Zhang, J. Dong, L. Liu, and H. Yu, "Automatic evaluation of the degree of facial nerve paralysis," *Multimedia Tools and Applications*, vol. 75, no. 19, pp. 11893–11908, 2016.
- [6] M. Sohrab, U. Abugo, M. Grant, and S. Merbs, "Management of the eye in facial paralysis," *Facial Plastic Surgery*, vol. 31, no. 2, pp. 140–144, 2015.
- [7] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-Wild challenge: the first facial landmark localization challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 397–403, Sydney, Australia, December 2013.
- [8] M. Koestinger, P. Wohlhart, P. M. Roth et al., "Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 2144–2151, Barcelona, Spain, November 2011.
- [9] G. Tzimiropoulos, "Project-out cascaded regression with an application to face alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3659–3667, Boston, MA, USA, June 2015.

- [10] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems Man and Cybernetics*, vol. 21, no. 3, pp. 660–674, 2002.
- [11] H. Jian, Z. Yan, Z. Cheng, M. Zhou, and Y. Han, "A non-invasive real-time solution for driving fatigue detection based on left prefrontal eeg and eye blink," in *Brain Informatics and Health*, pp. 325–335, Springer, Berlin, Germany, 2016.
- [12] T. Soukupova and J. Cech, "Eye blink detection using facial landmarks," in *Proceedings of the 21st Computer Vision Winter Workshop*, Rimske Toplice, Slovenia, February 2016.
- [13] P. C. Anil, B. Aravind, S. Saritha, and V. George, "Development of a headband for acquisition and analysis of forehead EOG signal for driver fatigue detection based on eye-blink patterns," *International Journal of Computer Applications*, vol. 96, no. 23, pp. 42–46, 2014.
- [14] M. H. Abdullah, K. J. Raman, A. Azman et al., "Driver fatigue detection," in *Information Science and Applications (ICISA) 2016*, pp. 269–278, Springer, Berlin, Germany, 2016.
- [15] Y. Qiao, Z. Kai, L. Xu, and X. Yin, "A smartphone-based driver fatigue detection using fusion of multiple real-time facial features," in *Proceedings of the 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pp. 230–235, Las Vegas, NV, USA, January 2016.
- [16] P. Bakker, *Real-time face and landmark localization for eye-blink-response detection: a heterogenous CPU-GPU approach*, Master dissertation, Delft University of Technology, Delft, Netherlands, 2017.
- [17] H. Sadeghi, A. A. Raie, and M. R. Mohammadi, "Facial expression recognition using geometric normalization and appearance representation," in *Proceedings of the 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, pp. 159–163, Tehran, Iran, December 2013.
- [18] T. Eiter and H. Mannila, "Computing discrete Fréchet distance," Tech. Rep. CD-TR 94/64, Information Systems Department, Technical University of Vienna, Vienna, Austria, 1994.
- [19] Z. X. Guo, G. Dan, J. Xiang et al., "An unobtrusive computerized assessment framework for unilateral peripheral facial paralysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 835–841, 2017.
- [20] A. Belal, "Structure of human muscle in facial paralysis," *The Journal of Laryngology & Otology*, vol. 96, no. 4, pp. 325–34, 1982.
- [21] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874, Columbus, OH, USA, June 2014.
- [22] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Proceedings of the European Conference on Computer Vision*, pp. 679–692, Florence, Italy, October 2012.
- [23] B. Chattopadhyay and K. Kelley, "Estimating the standardized mean difference with minimum risk: maximizing accuracy and minimizing cost with sequential estimation," *Psychological Methods*, vol. 22, no. 1, pp. 94–113, 2017.
- [24] H. Y. Lee, M. S. Park, J. Y. Byun, J. H. Chung, S. Y. Na, and S. G. Yeo, "Agreement between the facial nerve grading system 2.0 and the house-Brackmann grading system in patients with bell palsy," *Clinical and Experimental Otorhinolaryngology*, vol. 6, no. 3, pp. 135–139, 2013.

Research Article

DBT Masses Automatic Segmentation Using U-Net Neural Networks

Xiaobo Lai ¹, Weiji Yang ², and Ruipeng Li ³

¹College of Medical Technology, Zhejiang Chinese Medical University, Hangzhou 310053, China

²College of Life Science, Zhejiang Chinese Medical University, Hangzhou 310053, China

³Hangzhou Third People's Hospital, Hangzhou 310009, China

Correspondence should be addressed to Ruipeng Li; 705213095@qq.com

Received 12 November 2019; Revised 17 December 2019; Accepted 18 December 2019; Published 28 January 2020

Guest Editor: Plácido R. Pinheiro

Copyright © 2020 Xiaobo Lai 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.

To improve the automatic segmentation accuracy of breast masses in digital breast tomosynthesis (DBT) images, we propose a DBT mass automatic segmentation algorithm by using a U-Net architecture. Firstly, to suppress the background tissue noise and enhance the contrast of the mass candidate regions, after the top-hat transform of DBT images, a constraint matrix is constructed and multiplied with the DBT image. Secondly, an efficient U-Net neural network is built and image patches are extracted before data augmentation to establish the training dataset to train the U-Net model. And then the presegmentation of the DBT tumors is implemented, which initially classifies per pixel into two different types of labels. Finally, all regions smaller than 50 voxels considered as false positives are removed, and the median filter smoothes the mass boundaries to obtain the final segmentation results. The proposed method can effectively improve the performance in the automatic segmentation of the masses in DBT images. Using the detection Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), and area under the curve (AUC) as evaluation indexes, the Acc, Sen, Spe, and AUC for DBT mass segmentation in the entire experimental dataset is 0.871, 0.869, 0.882, and 0.859, respectively. Our proposed U-Net-based DBT mass automatic segmentation system obtains promising results, which is superior to some classical architectures, and may be expected to have clinical application prospects.

1. Introduction

Breast cancer is the most common malignant tumor of breast epithelial tissue in women, which seriously threatens the physical and mental health of patients [1]. In recent years, breast cancer has become a significant public health problem in today's society with a rising incidence and the younger incidence groups. Early diagnosis and treatment can effectively reduce the mortality and improve the quality of life of patients [2, 3]. Digital breast tomosynthesis (DBT) is a new 3D tomography method for breast cancer screening. It can reconstruct a small amount of low-dose mammographic images from a limited angle into three-dimensional mammographic images and can better detect some small hidden lesions [4]. Compared with traditional mammography, although DBT has improved the sensitivity of detecting breast masses, it has also significantly increased the

amount of interpretation data for radiologists, which is time consuming and poor repeatability [5–7]. If the computer can automatically detect breast masses in DBT images, it can not only reduce the review time for radiologists but also effectively reduce the misjudgment caused by excessive fatigue. Therefore, automatic segmentation of breast masses in DBT images is of great clinical value in assistant screening, early diagnosis, and preoperative localization of breast cancer [8–10].

Automatic segmentation of breast masses in DBT images is a challenging task because the signal-to-noise ratio of two-dimensional slice images of DBT is not high and the significance of breast masses is not strong [11]. Although it is difficult to achieve precise automatic segmentation of DBT tumors, many scientists have made fruitful explorations on automatic segmentation of DBT tumors in recent years because of its great significance in assistant diagnosis and

treatment [12]. These studies can be roughly divided into two categories [13]. One is breast mass detection based on DBT reconstructed slice images. Reiser et al. [14] proposed a breast mass detection computer-aided diagnosis (CAD) system, which uses a radial gradient index to detect and segment suspicious lesions in DBT reconstructed images. The experimental results show that incorporating vertical direction information does not improve the performance of gradient-based classifiers, but it can improve the performance of shape-based classifiers. In another early study, Chan et al. [15] also proposed a method for automatic detection of breast masses, including DBT reconstruction image interpolation to obtain cubic pixels, gradient field analysis to determine suspicious regions, three-dimensional region growth segmentation, and feature analysis, with the sensitivity of 85%. Then, Chan et al. [16] studied the relationship between the number of projected images and the dose used to obtain DBT images and the performance of the algorithm and used a set of 21 two-dimensional projected images or 11 reconstructed slice images to test. Another is breast mass detection based on the two-dimensional projection image of DBT. van Schie et al. [17] proposed an automatic detection method for breast masses, using a mammography image library to train classifiers. To optimize and make the technique suitable for DBT images, tomographic images were generated from reconstructed volume images for analysis. Palma et al. [18] constructed a system of automatic detection of breast masses in DBT reconstruction images by using fuzzy theory and antagonistic reasoning method. Kim et al. [19] studied the influence of the saliency of reconstructed slice images on the detection performance of breast masses in DBT reconstructed images and proposed an automatic detection method of breast masses based on the saliency of reconstructed slice images by DBT. In addition, some researchers fused the information of projection images and reconstructed images to detect DBT tumors. A hybrid method of two-dimensional and three-dimensional images was used to segment DBT masses with projection image and reconstructed image information [20].

Over the last few years, convolutional neural network (CNN) based on deep learning has become a research hotspot in the field of computer vision because of its strong ability to express image features [21, 22]. It has achieved fruitful results in image recognition and classification, target detection, and other fields. Elboushaki et al. validated that the CNN model can recognize fine mammographic features [23]. Viguera-Guillen et al. first proposed a full-convolution network for semantic segmentation, replacing the conventional full-connection layer in the CNN with the convolution layer to obtain a rough label graph and then using the deconvolution layer to sample the rough label graph to achieve the classification results of each pixel [24]. Cirean et al. used patches of 101×101 pixels to train a CNN for mitosis detection in breast cancer histology images, who won the ICPR 2012 Mitosis Detection Contest with F1-score of 0.782 [25]. Zhang et al. proposed a new FCN-like structure, U-net, for Bio-cell image segmentation [26]. This

method has attracted considerable attention in the medical field because the U-net architecture supports a small amount of the data training model and fast image segmentation with the trained model can be achieved. At present, this method has been applied to many different tasks and also achieved excellent results, such as image segmentation and image conversion. [27, 28].

In this work, we propose an efficient DBT masses automatic segmentation algorithm by using a U-Net architecture, which works with only weakly human-annotated mass masks. To suppress the background tissue noise and enhance the contrast of the mass candidate regions, we construct a constraint matrix, which is multiplied with the DBT image after the top-hat transform. A U-Net architecture is built, and image patches are extracted before data augmentation. Then, the presegmentation of the breast tumors in DBT images is implemented. All regions smaller than 50 voxels considered as false positives are removed and the final segmentation results are obtained after the median filter smoothes the mass boundaries. The proposed method can effectively improve the performance in the automatic segmentation of the masses in DBT images. The architecture is developed and evaluated with the DBT images database prepared by a neuroradiologist in our research team. Experimental results tested on the DBT database indicate that the presented DBT mass CAD architecture achieves the high level of segmentation. To our knowledge, this is the first DBT study to employ the U-Net framework to segment the masses in DBT images automatically.

The remaining sections are organized as follows. In Section 2, the proposed method is presented. The database used for evaluation is detailed and results are presented and discussed in Section 3. Finally, the main conclusions are presented in Section 4.

2. Methods

Our proposed approach consists of six main stages: DBT image preprocessing, patch extraction, data augmentation, voting scheme fusion, segmentation via the U-Net architecture, and postprocessing. An overview of our presented architecture is illustrated in Figure 1.

2.1. DBT Image Preprocessing. Usually, the random distribution of X-ray photons in mammography or DBT images will seriously affect the quality of breast images. However, as the average photon number (X-ray dose) increases, the noise will gradually decrease. For a typical DBT system, radiation exposure is an important factor to avoid the risk of radiation-induced cancer. Therefore, the low radiation dose is often used when creating the tomosynthesis images and the total radiation dose of DBT is slightly higher than that of standard mammography. In theory, the typical DBT images usually contain Poisson distribution noise. To address this issue, a top-hat transform is applied to enhance the contrast between candidate tumor location regions and background tissues, which is defined as

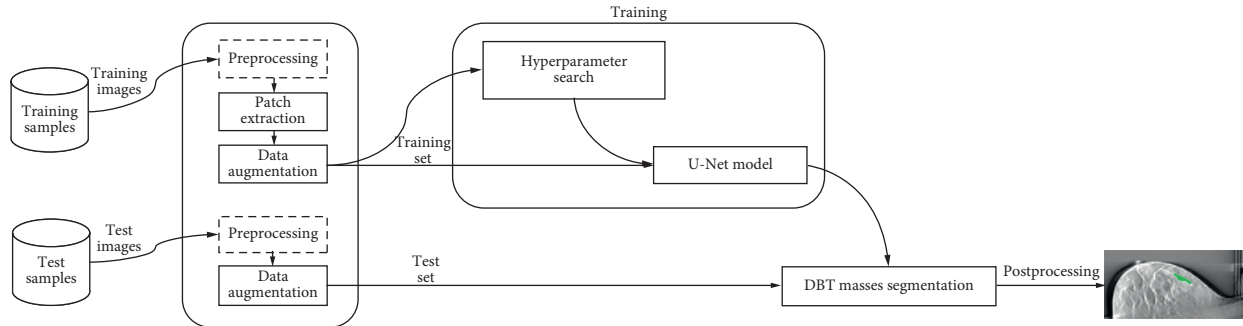


FIGURE 1: Overview of the presented method.

$$I_{\text{top-hat}} = I - \min((I \cdot s_c) \circ s_o; I), \quad (1)$$

where I denotes the preprocessed image, \cdot represents the morphological “closed” operation, \circ denotes the morphological “open” operation, and s_c and s_o are the disc structure elements. Also, to enhance candidate location regions and suppress the background tissues, a constraint matrix we constructed is multiplied with the image matrix. The constraint matrix is generated by an isotropic radial basis function centered on the candidate location region with a variance σ^2 (σ is 5 mm). It can be detailed that randomly selected one mark location x_r from all locations x in one image view, remove x_c from x , where

$$x_c = \{x_i \mid \|x_i - x_r\| \leq 5, \quad x_i \in x\}. \quad (2)$$

Figure 2 shows the DBT image preprocessing effects, where Figure 2(a) is the original image and Figure 2(b) is the preprocessed image.

2.2. Image Patch Generation. Generally, it is straightforward to train the proposed U-Net directly by using the image patches extracted from the DBT mass regions because we have location information of the masses in the training and testing image sets. However, the available dataset has a small number of examples as compared to other U-net classification problems, and the direct use of whole images would most likely result in overfitting. This can be addressed by splitting images into patches which increases dataset complexity and dimension. In fact, nonmass regions can also provide some useful information for the breast mass segmentation task. In our model, the input to the U-Net architecture is a two-dimensional array, of shape height \times width, since it consists of a two-dimensional patch of width \times height voxels. The two-dimensional patches are taken along the x - y axis, also called the axial plane in anatomy. To avoid overfitting, we can extract the image patches from mass and nonmass regions to augment the training data. That is, we use the image patches extracted from the nonmass regions as additional negative samples for the U-Net architecture training, to help the proposed model to distinguish confounding regions from DBT masses.

In addition, the training data has to be balanced; that is, the same number of examples for each class should be included in the training data, which is to ensure that the U-Net

model can generalize well. However, the number of pixels in the mass regions is significantly less than that in the nonmass regions, which leads to a severe class imbalance problem. To address this issue, we randomly resample at each epoch the same number of patches for each class from all possible patches for that class.

2.3. U-Net Architecture. In this part, we will briefly introduce the architecture of the proposed typical U-Net model and its application to our DBT mass CAD system. We perform an end-to-end pixel-wise segmentation via a U-Net model. As shown in Figure 3, we illustrate the framework of our model. The proposed U-Net-based DBT mass segmentation architecture is mainly composed of a contracting path in the left side and an expansive path in the right side. The contraction path in the model follows the typical structure of the convolution network, including two 3×3 convolutions applied repeatedly, each convolution is followed by a rectified linear unit (ReLU) and a 2×2 max-pooling operation with stride 2, which is used for down sampling. In each of the down sampling step, we double the number of feature channels. Each step in the expansive path includes the up sampling of the feature map, followed by a 2×2 convolution (“up convolution”), which halves the number of feature channels, cascades them with the corresponding cropped feature map in the contraction path, and two 3×3 convolutions, each of which is followed by a ReLU. Because every convolution will lose the boundary pixels of the image, the image must be cropped. In the last layer of the network, a 1×1 convolution is used to map the characteristic vector of each 64 components to the required class number. In total, there are 23 convolutional layers in our proposed U-Net architecture.

2.4. Training Procedure. During the training process, we have balanced the training data by randomly resampling at each epoch the same number of patches for each class from all possible patches for that class. However, it is worth noting that the mass detection task still has a class-imbalance problem, where the number of positive samples (i.e., pixels in mass regions) is much less than the number of negative samples (i.e., pixels in nonmass regions). Hence, in our study, the proposed U-Net model uses the F_β -measure as the cost function, rather than the cross-entropy-based or the

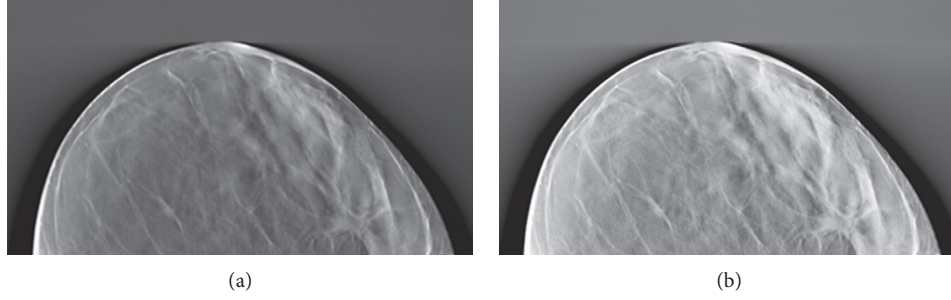


FIGURE 2: Preprocessing effects. (a) Original image and (b) preprocessed image.

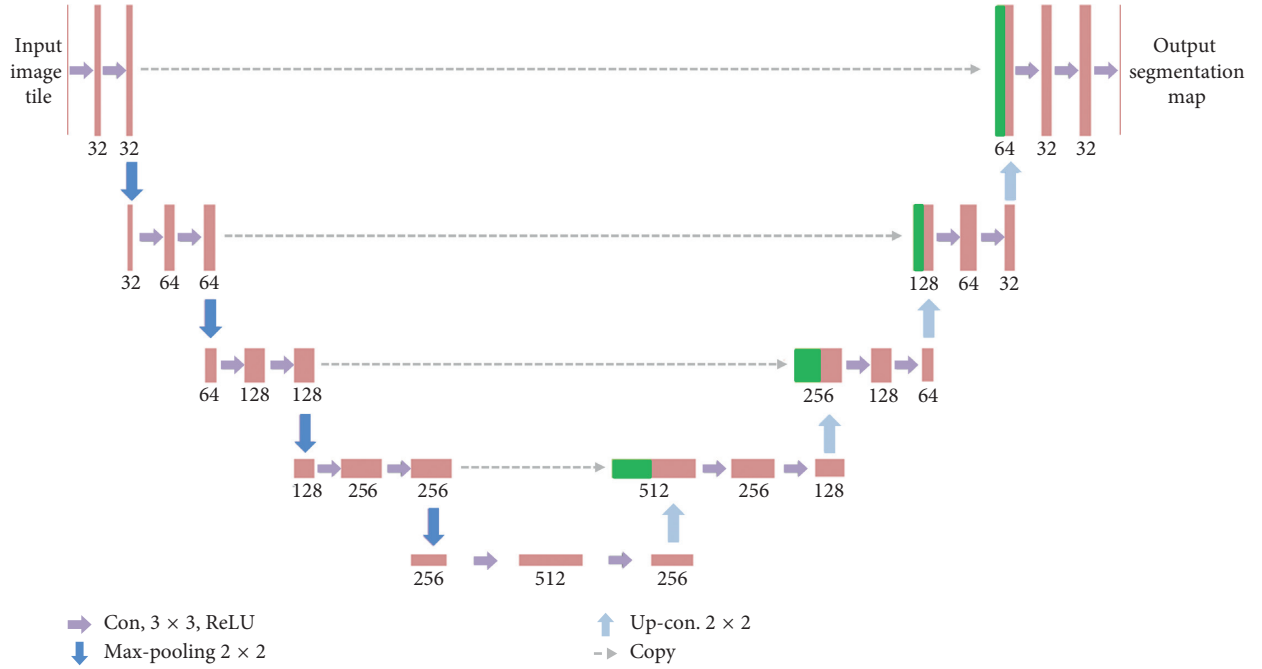


FIGURE 3: Proposed DBT mass segmentation U-Net architecture.

quadratic cost function. Denote S and T as the predicted heatmap and the ground truth heatmap, respectively. Let M represent the number of elements (pixels) in S and T , and the F_β -measure based loss function is defined as

$$F_\beta(S, T) = \frac{(1 + \beta^2) \sum_{i=1}^M s_i t_i}{\sum_{i=1}^M s_i + \sum_{i=1}^M t_i}, \quad (3)$$

where s_i is the i th element of the predicted heatmap and t_i is the i th element of the ground truth heatmap. In our study, we set $\beta = 1$.

In the training procedure, the input images and their corresponding segmentation heatmaps are used to train the U-Net model with the stochastic gradient descent. Besides that, we applied the Adaptive Moment Estimation Method (Adam) [29] which is a stochastic gradient descent method that computes adaptive learning rates for each parameter to minimize the F_β -measure-based loss function. The Adam optimizer parameters in our proposed U-Net architecture are set as learning rate = 0.0002 and the maximum number of epochs = 150. We adopted a Xavier normal heuristic [30]

to initialize kernel weights in our study, which allowed us to maintain the gradients in controlled levels and thus prevent gradient vanishing during back-propagation. The biases are all initialized to 0. Because of the unpadded convolutions, the size of the output image is smaller than that of the input image by a constant border width. In order to minimize the overhead and maximize the use of GPU memory, we tend to make large input tiles over a large batch size, and therefore reduce the batch to a single image. Therefore, we apply a high momentum to make a large number of the previously seen training samples to determine the update in the current optimization step.

2.5. DBT Data Augmentation. To improve the performance of U-Net, we need to extend the data to generate more training data from the original data. In typical applications of the U-Net neural network for image processing and computer vision tasks, translations and rotations are used. In this study, the data consists entirely of two-dimensional patches. Thus, translation cannot be used as it would result

in a different patch, with a possibly different label. However, using rotations of the patches might give some performance improvements. Therefore, we perform the rotations by using angles multiple of 90° .

2.6. Voting Scheme. Every test DBT image is first split into a set of patches, and for each patch a probabilistic prediction $p_i \in [0, 1]$ is made using the U-Net model. These predictions are then fused into the final image label using one of the following three voting schemes. The first three (Majority voting, Maximum probability, and Sum of probabilities) are also used and compared [31], whereas the other one (Connectivity) is proposed by us. Our motivation behind this voting schemes is to reinforce spatial consistency between votes of patches because in reality the true image label is likely to be assigned based on the structure of a particular connected region of the tissue rather than on many disconnected bits. In particular, if patches that vote for the same label are adjacent to each other, then this should be a more significant vote than if they are separated. The precise definitions follow.

2.6.1. Majority Voting. Let us define the number of patches that vote for the class label k as

$$v_k = \sum_{i \in P} \prod (l_i = k), \quad (4)$$

where l_i is the class label of patch i . The image label \tilde{k} is then selected as the most common patch label by

$$\tilde{k} = \arg \max_{k \in K} (v_k). \quad (5)$$

2.6.2. Maximum Probability. The patch with the highest-class probability decides the image class label as

$$\tilde{k} = \arg \max_{k \in K} \left(\max_{i \in P} (p_i[k]) \right), \quad (6)$$

where $p_i[k] = p$ (patch $i \in$ class k).

2.6.3. Sum of Probabilities. The patch class probabilities are summed and the class with the largest sum is chosen as

$$\tilde{k} = \arg \max_{k \in K} \left(\sum_{i \in P} p_i[k] \right). \quad (7)$$

2.6.4. Connectivity. This method is based on counting the number c_k of connections for each class k , where the connection means that two adjacent patches have the same class label k . These counts are calculated as

$$c_k = \sum_{i \in P} \sum_{j \in P_i} \prod (l_i = l_j), \quad (8)$$

where P_i is the set of patches adjacent to patch i , including the patches along diagonals. The obtained counts are then used to weight the class votes v_k as

$$\tilde{k} = \arg \max_{k \in K} \left(\frac{c_k v_k}{\sum_{k' \in K} c_{k'}} \right) = \arg \max_{k \in K} (c_k v_k). \quad (9)$$

2.7. Segmentation Postprocessing. Some small clusters may be mistakenly classified as the DBT masses. To deal with this issue, we impose volumetric constrains by removing clusters in the segmentation obtained by the U-Net that are less than 50 voxels in volumes.

3. Experimental Details

3.1. Materials. The benchmarking clinical DBT images used are collected at Zhejiang Chinese Medical University Affiliated Guangxing Hospital and Zhejiang Provincial Hospital of Traditional Chinese Medicine (TCM) with Institutional Review Board (IRB) approval. Every DBT image is produced by low dose exposure, where the total shot dose should be within the range of a regular mammogram dose. DBT cases are acquired in mediolateral oblique (MLO) and craniocaudal (CC) views (Siemens Mammomat Inspiration DBT system) using a total tomographic angular range of 60° with a 5° increment of rotation and 12 projection views. The DBTs are reconstructed to the images with 1 mm slice spacing by using the simultaneous algebraic reconstruction technique (SART). We convert the images into TIFF stack/slices and used data in JPEG format. Depending on the thickness of the breast, each DBT volume provides between 50 and 80 2D slices with a resolution of 1200 ± 901 pixels, which are saved in the JPEG format.

The database consists of 87 DBT volumes and 3960 2D X-ray images slices. Among these cases of breast cancer patients, 29 are malignant and 42 are benign (absolute healthy). The noncancerous DBT volumes are collected from the left and right breasts of 23 patients without early signs of breast cancer. The cohort of cancerous cases is annotated by two experienced radiologists with a 2D bounding box for DBT.

3.2. Experiments Design. Our purpose is to evaluate three scenarios that reflect common practices in research and evaluation of the DBT mass segmentation with the U-Net model:

- (i) The data used for model training and for model testing are from the same hospital
- (ii) The data used for model training and for model testing are from a different hospital
- (iii) The data used for model training are from the same institution as the data for model testing are enriched by additional data coming from a different hospital resulting in an increased size of the training dataset

We use the following way to simulate three scenarios with 5-fold crossvalidation. First, we number the DBT slice images serially for each hospital. Within a fold, the serial numbers for training and test set are obtained. To automatically segment the DBT mass in the test set, we develop

three U-Net models as follows. (i) The first U-Net model is trained on the DBT images from the same hospital using the serial numbers in the training set. (ii) The second U-Net model is trained on the DBT images using the training set serial numbers from the other dataset. In this way, we use the same number of DBT images used for training the U-Net model. (iii) The third U-Net model is built using all the DBT images used in (i) and (ii) such that DBT images from both hospitals are used. Hence, we use six types of train-test combinations: (a) train on hospital 1 (denoted as H1), test on H1, (b) train on hospital 2 (denoted as H2), test on H2, (c) train on H1, test on H2, (d) train on H2, test on H1, (e) train on (H1 + H2), test on H1, and (f) train on (H1 + H2), test on H2. Please note that the DBT images segmented by a U-Net model are never present in the training set for that U-Net model. For each DBT image under test, we classify each voxel into one of the two classes (nonmass region and mass region).

3.3. Evaluation Metrics. To enable comparison with other state-of-the-art works, we used three metrics commonly found in the literature: Accuracy (Acc), Sensitivity (Sen), and Specificity (Spe) as evaluation of classification results. Acc refers to the ratio of the number of pixels correctly segmented to the number of total pixels in the image, Sen refers to the probability of a positive test among the subjects with the condition, and Spe refers to the probability of a negative test among the subjects without the condition. The DBT masses segmentation evaluation metrics are defined as follows:

$$\begin{aligned} \text{Sen} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{Spe} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \\ \text{Acc} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \end{aligned} \quad (10)$$

where TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively.

Additionally, to evaluate the robustness of our proposed U-Net model, the receiver operating characteristic (ROC) curve and the average area under the curve (AUC) are calculated and compared.

The proposed approach is implemented in Python using the machine learning library Keras. The training and test experiments are performed using the cloud computing service PAI-DSW provided by Ali. Specifically, we use the runtime platform processor of Intel (R) Core (TM) i7-6800K CPU @ 3.40 GHz, 32 GB RAM, Nvidia GeForce RTX 2080 Ti, 64-bit Windows 10. The presented figures are produced using the plotting library matplotlib. All the parameters are set according to our preexperimental study, and it takes about 56 minutes to learn the parameters. An example code is shown below:

```
import numpy as np
import configparser
```

```
import matplotlib as plt
import os
from keras.models import Model
from keras.layers import Input, concatenate,
Conv2D, MaxPooling2D, UpSampling2D, Reshape,
core, Dropout
from keras.optimizers import Adam
from keras.callbacks import ModelCheckpoint,
LearningRateScheduler
from keras import backend as K
from keras.optimizers import SGD
from lib.help_functions import *
```

3.4. Results and Analysis. To assess the segmentation performance of our proposed method based on U-Net, we evaluated the overlap between the proposed DBT mass labels and the ground truth (GT). Figure 4 showcases example results of the DBT mass automatic segmentation with the U-Net model trained and tested on the images (patient #1, patient #2, and patient #3) from the same hospital. The 1st row shows DBT original images for patient #1, patient #2, and patient #3, the 2nd row shows respective images segmented with our proposed U-Net architecture, and the 3rd row shows the same images segmented manually. The results indicate that our proposed U-Net model has high agreement between the generated results and the provided labels, and it takes about 132 seconds to label a sample.

Table 1 presents the performance of various voting schemes to assign final image labels. We can see that when the U-Net model is trained on the DBT images comes from the same hospital and tested on the DBT images comes from the same hospital, the maximum probability achieves best accuracy, which highlights the importance of a good voting scheme, namely, maximum probability. Contrary to my expectations, the method connectivity, whose aim is to reinforce consensus between spatially close patches, does not provide better results and scored the same as the majority voting. This may imply that having such spatial constraints is not relevant to this segmentation problem.

Table 2 shows the average Sensitivity (Sen) values of the DBT mass automatic segmentation by training-testing within the same hospital, across hospitals, AND with both hospitals using the maximum probability voting scheme. It can be found that the performance of the DBT mass automatic segmentation based on the proposed U-Net model significantly decreases when the proposed U-Net model is trained on data comes from a different hospital (Sen = 0.83 ± 0.015 for H1 and Sen = 0.85 ± 0.013 for H2) as compared with when it is trained with the data comes from the same hospital (Sen = 0.88 ± 0.009 for H1 and Sen = 0.89 ± 0.021 for H2).

Similar to Sensitivity (Sen), Specificity (Spe) shows that training the U-Net model on different hospital DBT images decreases the performance (Spe = 0.86 ± 0.009 for H1 and Spe = 0.87 ± 0.013 for H2) compared with training on the same hospital dataset (Spe = 0.89 ± 0.011 for H1 and

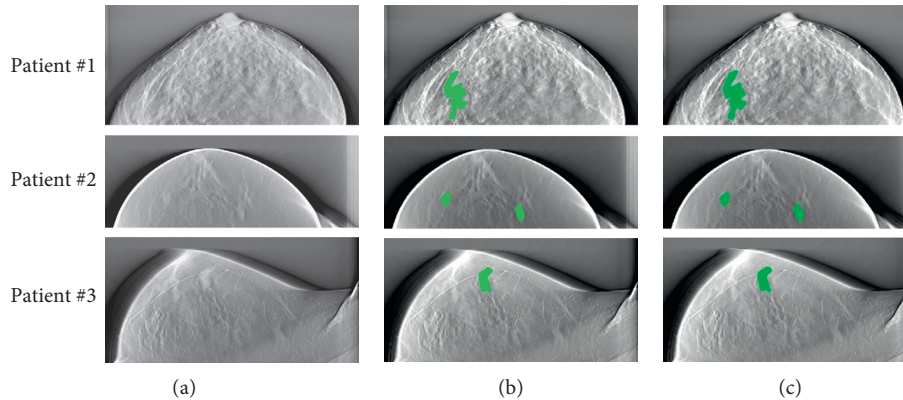


FIGURE 4: Illustration of predicted DBT mass of three patients. (a) Original image. (b) Segmented image. (c) Ground truth.

TABLE 1: Accuracy obtained using various voting schemes.

Voting scheme	Train on T1, test on T1	Train on T2, test on T2
Majority voting	0.85 ± 0.021	0.86 ± 0.018
Maximum probability	0.87 ± 0.017	0.89 ± 0.009
Sum of probability	0.86 ± 0.013	0.87 ± 0.012
Connectivity	0.85 ± 0.011	0.86 ± 0.017

TABLE 2: Sensitivity (Sen) of different types of train-test combinations using 5-fold crossvalidation.

Train	Test on H1	Test on H2
Same hospital	0.88 ± 0.009	0.89 ± 0.021
Different hospitals	0.78 ± 0.015	0.86 ± 0.013
Both hospitals	0.87 ± 0.011	0.88 ± 0.015

Spe = 0.89 ± 0.017 for H2). All results using Spe are presented in Table 3. Accuracy (Acc) has been used as another metric of evaluation, as shown in Table 4. We find that when the proposed U-Net model is trained on different hospital DBT images, the Acc values are less than 0.87. However, when the proposed U-Net model is trained on the same or both hospitals DBT images, the Acc values are always greater than 0.87. Specifically, low Sen and high Acc can be obtained (Sen = 0.78 ± 0.015 and Acc = 0.86 ± 0.009) while training on H2 and testing on H1, which indicates that many FN are present, and high Sen and low Acc are obtained (Sen = 0.86 ± 0.013 and Acc = 0.79 ± 0.015) while training on H1 and testing on H2, which indicates many FN are present. For training and testing on the same hospital DBT images, both Sen and Acc are greater than 0.88 (Sen = 0.88 ± 0.009 and Acc = 0.88 ± 0.011 for H1) and (Sen = 0.89 ± 0.021 and Acc = 0.89 ± 0.017 for H2). Similar to training/testing on the same hospital, high Sen and Acc are obtained while training on both hospitals DBT images (Sen = 0.87 ± 0.011 and Acc = 0.86 ± 0.021 for H1) and (Sen = 0.88 ± 0.015 and Acc = 0.87 ± 0.019 for H2).

3.5. Discussion. To validate our U-Net CAD framework, we combine the two DBT image datasets into a bigger dataset,

TABLE 3: Specificity (Spe) of different types of train-test combinations using 5-fold crossvalidation.

Train	Test on H1	Test on H2
Same hospital	0.89 ± 0.011	0.89 ± 0.017
Different hospitals	0.86 ± 0.009	0.87 ± 0.013
Both hospitals	0.88 ± 0.013	0.88 ± 0.021

TABLE 4: Accuracy (Acc) of different types of train-test combinations using 5-fold crossvalidation.

Train	Test on H1	Test on H2
Same hospital	0.88 ± 0.011	0.89 ± 0.009
Different hospitals	0.85 ± 0.019	0.79 ± 0.015
Both hospitals	0.86 ± 0.021	0.87 ± 0.019

and the combination is denoted as the entire dataset. We compare the performance of various methods on automated DBT mass detection at aspects of the classifier used, DBT dataset size, Sen, Acc, and AUC in Table 5, our network achieve quite a competitive result than some of them. Among these models, we will discuss in detail the research works by Kim et al. [31], Fotin et al. [32], and Samala et al. [33], which applied deep learning methods to DBT mass detection and segmentation. Their works evaluated the DBT mass automatic segmentation CAD frameworks, which are based on both hand-crafted feature and deep convolutional neural network (DCNN)-based models. Samala et al. [33] proposed a DCNN architecture consists of four convolutional layers and three fully connected layers. Firstly, the DCNN model is trained on large-scale 2D mammography dataset, then the first three convolutional layers weights are frozen, and the rest of which is trained. The results of the DCNN model have shown the AUC of over 80% and the 80% Sen. Fotin et al. [32] have developed a CAD framework of the DBT mass detection using a DCNN that is trained on the generated candidate region of interest (ROIs), which contains 1864 breast lesions in the mammography and 339 breast lesions from the DBT images data. It is reported that their model achieved an Acc of 86.40% and 89% Sen. The latent bilateral feature representations of masses in reconstructed DBT volumes o are classified with the DCNN model proposed by Kim et al. [31], in which low-level features are

TABLE 5: Comparisons of selected studies in the detection of masses in the DBT images.

Method	Classifier	DBT dataset size	Sen	Acc	AUC
Kim et al. [31]	LDA	36	0.90	—	—
Shamsolmoali et al. [28]	SVM	160	—	—	0.847
Sajjad et al. [29]	DCNN	344	0.89	0.864	—
Glorot and Bengio et al. [30]	DCNN	324	0.80	—	0.80
Palma et al. [17]	SVM	101	0.90	—	—
Chan et al. [16]	Neural network	752	0.80	—	—
Reiser et al. [14]	LDA	100	0.80	—	—
Proposed	U-net	87	0.869	0.871	0.859

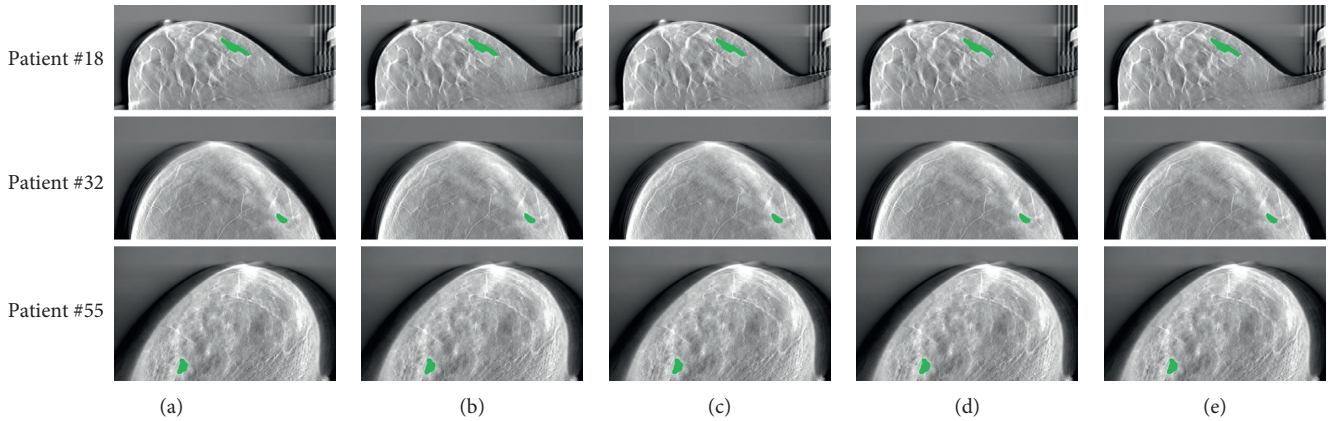


FIGURE 5: Examples of DBT masses segmented by our U-Net architecture and other classical CAD frameworks. (a) Proposed U-Net. (b) Reiser et al. (c) Kim et al. (d) Fotin et al. (e) Samala et al.

extracted from the ROIs and the corresponding ROIs through the convolutional layers separately. To represent the high-level bilateral features of the DBT masses, the low-level features are combined in the fully connected layer. The AUC of 0.847 for the latent bilateral feature representation model is reported. Concerning our method, we obtained 0.871 Acc, 0.869 Sen, 0.882 Spe, and an AUC of 0.859 for the testing dataset with 87 DBT volumes.

In other models, not based on DCNN, we select the works of Chan et al. [15], Palma et al. [18], and Schie et al. [17]. Chan et al. [15] introduce three methods based on 2D and 3D, and the hybrid that combines 2D and 3D. For the hybrid method, they report 80% Sen with 1.23 FPs per volume for the dataset of 100 DBT images containing 69 malignant patient cases. Palma et al. [18] have developed the two-channel DBT masses detection CAD framework, in which every channel classifies one type of DBT lesions. They combine the findings from channels with the disjunctive fusion method. Their results show 90% Sen for 101 DBT volumes containing 53 lesions. van Schie et al. [17] present a two-stage method. The ROIs in 2D slice images are detected in the first step, and then extracted regions from 2D slice images are combined to localize 3D ROIs on DBT volumes in the second step. Obtained results on the DBT images data from 192 patients with 49 patients having one or more malignancies show 80% Sen with 3 FPs per volume. Reiser et al. [34] introduce the approach that detects DBT masses in 2D projection views then using the visibility angular range of

findings combines the detections, and 90% Sen for 36 DBT volumes is reported.

Figure 5 shows examples of DBT masses segmented by our U-Net architecture and other classical CAD frameworks. However, it is not a feasible to make a fair comparison between our CAD model with other models on DBT images because other models are trained and tested on different private datasets that are not public. Although the proposed DBT mass automatic segmentation CAD model could not achieve the best overall segmentation performance, our U-Net architecture achieves the 87.1% Acc and 86.9% Sen with an AUC of 0.859. The experimental results show that our approach achieves promising results given the fact they are obtained on DBT images data, and the U-Net model is trained on 2D slice images from DBT volumes and not on 2D mammography dataset. Although the proposed CAD framework has achieved promising results in the automatic segmentation of DBT masses, it can be further improved when more DBT images data are available. The main limitation of this work is the lack of sufficient DBT image data. To achieve satisfactory general performance, the proposed CAD framework requires diverse data and more structural distortion samples. We intend to identify all early signs of breast lesions in DBT images based on the method we used in future; by automatically detecting lesions in the DBT image, physicians can make diagnosis more accurately and quickly and surgeons can rely on it to discuss the procedure with colleagues.

4. Conclusions

Our study presents a novel U-Net architecture-based CAD model for the automatic detection of masses in DBT slice images, and we compare this model with other classical CAD frameworks. The advantages of the proposed U-Net architecture are that the U-Net model shortcuts among different layers can provide both global and local structural information of input images for breast mass detection. We demonstrate that our proposed U-Net CAD framework achieves promising results in the automatic segmentation of the DBT masses and exhibits outperformance compared with other classical CAD frameworks using the metrics of accuracy, AUC, specificity, and sensitivity. The future research work will focus on the combination of 3D reconstruction image information and 2D efficient data information. This combination is expected to improve the accuracy of other early signs of breast cancer detection procedures, which will be especially valuable when more clinical cases are available.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was funded in part by the National Science Foundation of China (Grant no. 61602419) and also supported by the Natural Science Foundation of Zhejiang Province of China (Grant nos. LY16F010008 and LQ16F020003).

References

- [1] A. Kumar, S. K. Singh, S. Saxena et al., "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," *Information Sciences*, vol. 508, pp. 405–421, 2020.
- [2] A. S. Lakshminarayanan, S. Radhakrishnan, G. M. Pandiasankar, and S. Ramu, "Diagnosis of cancer using hybrid clustering and convolution neural network from breast thermal image," *Journal of Testing and Evaluation*, vol. 47, p. 3975, 2019.
- [3] W. Wu, S. Pirbhulal, and G. Li, "Adaptive computing based biometric security for intelligent medical applications," *Neural Computing and Applications*, vol. 12, p. 1, 2018.
- [4] J. Krammer, S. Zolotarev, I. Hillman et al., "Evaluation of a new image reconstruction method for digital breast tomosynthesis: effects on the visibility of breast lesions and breast density," *The British Journal of Radiology*, vol. 92, p. 1103, 2019.
- [5] K. J. Geras, R. M. Mann, and L. Moy, "Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives," *Radiology*, vol. 293, no. 2, pp. 246–259, 2019.
- [6] A. H. Mark, "Digital mammography imaging: breast tomosynthesis and advanced applications," *Radiologic Clinics of North America*, vol. 48, p. 917, 2010.
- [7] K. Johnson, S. Zackrisson, A. Rosso et al., "Tumor characteristics and molecular subtypes in breast cancer screening with digital breast tomosynthesis: the malmö breast tomosynthesis screening trial," *Radiology*, vol. 293, no. 2, pp. 273–281, 2019.
- [8] S. P. Poplack, T. D. Tosteson, C. A. Kogel, and H. M. Nagy, "Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography," *American Journal of Roentgenology*, vol. 189, p. 616, 2012.
- [9] Y.-J. Kim, I. Sechopoulos, M. S. Newell et al., "Comparison of two-view digital breast tomosynthesis to three-view digital mammography in a simulated screening setting," *Acta Radiologica*, vol. 60, no. 9, pp. 1094–1101, 2019.
- [10] L. R. Borges, B. Barufaldi, R. F. Caron et al., "Technical note: noise models for virtual clinical trials of digital breast tomosynthesis," *Medical Physics*, vol. 46, p. 2683, 2019.
- [11] K. Mendel, H. Li, D. Sheth, and M. Giger, "Transfer learning from convolutional neural networks for computer-aided diagnosis: a comparison of digital breast tomosynthesis and full-field digital mammography," *Academic Radiology*, vol. 26, no. 6, pp. 735–743, 2019.
- [12] J. M. Lee, S. C. Partridge, G. J. Liao et al., "Double reading of automated breast ultrasound with digital mammography or digital breast tomosynthesis for breast cancer screening," *Clinical Imaging*, vol. 55, pp. 119–125, 2019.
- [13] R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, "Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets," *IEEE Transactions on Medical Imaging*, vol. 38, no. 3, pp. 686–696, 2019.
- [14] I. Reiser, R. M. Nishikawa, M. L. Giger et al., "Computerized detection of mass lesions in digital breast tomosynthesis images using two- and three dimensional radial gradient index segmentation," *Technology in Cancer Research & Treatment*, vol. 3, no. 5, pp. 437–441, 2004.
- [15] H.-P. Chan, J. Wei, B. Sahiner et al., "Computer-aided detection system for breast masses on digital tomosynthesis mammograms: preliminary experience," *Radiology*, vol. 237, no. 3, pp. 1075–1080, 2005.
- [16] H. P. Chan, J. Wei, Y. Zhang, S. Berkman, H. Lubomir, and A. H. Mark, "Detection of masses in digital breast tomosynthesis mammography: effects of the number of projection views and dose," in *Lecture Notes in Computer Science*, vol. 5116, Springer, Berlin, Germany, 2008.
- [17] G. van Schie, M. G. Wallis, K. Leifland, M. Danielsson, and N. Karssemeijer, "Mass detection in reconstructed digital breast tomosynthesis volumes with a computer-aided detection system trained on 2D mammograms," *Medical Physics*, vol. 40, p. 283, 2013.
- [18] G. Palma, I. Bloch, and S. Muller, "Detection of masses and architectural distortions in digital breast tomosynthesis images using fuzzy and a contrario approaches," *Pattern Recognition*, vol. 47, no. 7, pp. 2467–2480, 2014.
- [19] S. T. Kim, D. H. Kim, and Y. M. Ro, "Breast mass detection using slice conspicuity in 3D reconstructed digital breast volumes," *Physics in Medicine and Biology*, vol. 59, no. 17, pp. 5003–5023, 2014.
- [20] H. P. Chan, J. Wei, Y. Zhang et al., "Computer-aided detection of masses in digital tomosynthesis mammography: combination of 3D and 2D detection information," in *Proceedings of*

- the SPIE Medical Imaging Conference*, vol. 6514, p. 521, San Diego, CA, USA, March 2007.
- [21] B. Fang, X. Zeng, and M. Zhang, "NestDNN: resource-aware multi-tenant on-device deep learning for continuous mobile vision," in *Proceedings of the ACM International Conference on Mobile Computing and Networking*, vol. 109, p. 115, New Delhi, India, October 2018.
- [22] M. Penmetcha, S. Luo, A. Samantaray, J. E. Dietz, B. Yang, and B. C. Min, "Computer vision-based algae removal planner for multi-robot teams," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 16, p. 1575, Bari, Italy, October 2019.
- [23] A. Elboushaki, R. Hannane, K. Afdel, and L. Koutti, "MultiD-CNN: a multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences," *Expert Systems with Applications*, vol. 139, p. 313, 2020.
- [24] J. P. Viguera-Guillen, B. Sari, S. F. Goes et al., "Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation," *BMC Biomedical Engineering*, vol. 1, p. 4, 2019.
- [25] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 8150, p. 411, Nagoya, Japan, September 2013.
- [26] Y. Zhang, J.-H. Chen, K.-T. Chang et al., "Automatic breast and fibroglandular tissue segmentation in breast mri using deep learning by a fully-convolutional residual neural network U-Net," *Academic Radiology*, vol. 26, no. 11, pp. 1526–1535, 2019.
- [27] L. C. Li and T. Jia, "Optical coherence tomography vulnerable plaque segmentation based on deep residual U-Net," *Reviews in Cardiovascular Medicine*, vol. 20, p. 171, 2019.
- [28] P. Shamsolmoali, M. Zareapoor, R. Wang, H. Zhou, and J. Yang, "A novel deep structure u-net for sea-land segmentation in remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3219–3232, 2019.
- [29] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *Journal of Computational Science*, vol. 30, pp. 174–182, 2019.
- [30] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, p. 249, Sardinia, Italy, May 2010.
- [31] D. H. Kim, S. T. Kim, and Y. M. Ro, "Latent feature representation with 3-D multi-view deep convolutional neural network for bilateral analysis in digital breast tomosynthesis," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 927, Shanghai, China, March 2016.
- [32] S. V. Fotin, Y. Yin, H. Haldankar, J. W. Hoffmeister, and S. Periaswamy, "Detection of soft tissue densities from digital breast tomosynthesis: comparison of conventional and deep learning approaches," in *Proceedings of the SPIE Medical Imaging*, vol. 9785, p. 495, San Diego, CA, USA, March 2016.
- [33] R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, J. Wei, and K. Cha, "Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography," *Medical Physics*, vol. 43, no. 12, pp. 6654–6666, 2016.
- [34] I. Reiser, R. M. Nishikawa, M. L. Giger et al., "Computerized mass detection for digital breast tomosynthesis directly from the projection images," *Medical Physics*, vol. 33, no. 2, pp. 482–491, 2006.