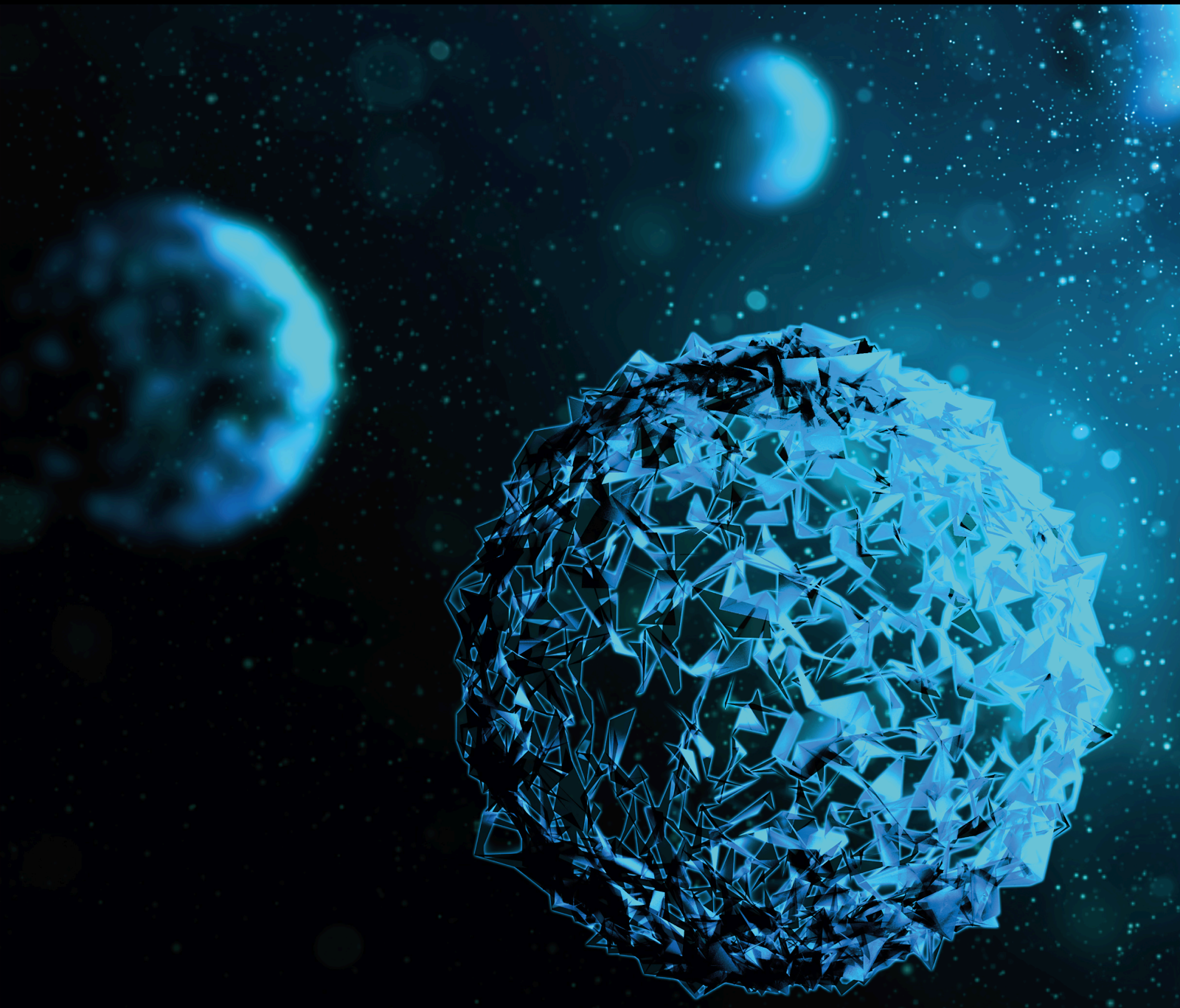


Machine Learning and Modelling for Biomedical Information Analysis

Lead Guest Editor: Fei Yang

Guest Editors: Lei Zhang, Qiushi Zhao, and Yue Zhang





Machine Learning and Modelling for Biomedical Information Analysis

BioMed Research International

Machine Learning and Modelling for Biomedical Information Analysis

Lead Guest Editor: Fei Yang

Guest Editors: Lei Zhang, Qiushi Zhao, and Yue
Zhang



Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in "BioMed Research International." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Section Editors





Penny A. Asbell, USA
David Bernardo , Spain
Gerald Brandacher, USA
Kim Bridle , Australia
Laura Chronopoulou , Italy
Gerald A. Colvin , USA
Aaron S. Dumont, USA
Pierfrancesco Franco , Italy
Raj P. Kandpal , USA
Fabrizio Montecucco , Italy
Mangesh S. Pednekar , India
Letterio S. Politi , USA
Jinsong Ren , China
William B. Rodgers, USA
Harry W. Schroeder , USA
Andrea Scribante , Italy
Germán Vicente-Rodríguez , Spain
Momiao Xiong , USA
Hui Zhang , China

Academic Editors











Bioinformatics

Contents




EEG-Based Multiword Imagined Speech Classification for Persian Words

M. R. Asghari Bejestani , Gh. R. Mohammad Khani , V. R. Nafisi , and F. Darakeh 
Research Article (20 pages), Article ID 8333084, Volume 2022 (2022)



A U-Net Approach to Apical Lesion Segmentation on Panoramic Radiographs

Ibrahim S. Bayrakdar , Kaan Orhan , Özer Çelik , Elif Bilgir , Hande Sağlam , Fatma Akkoca Kaplan , Sinem Atay Görür , Alper Odabaş , Ahmet Faruk Aslan , and Ingrid Różyło-Kalinowska 
Research Article (7 pages), Article ID 7035367, Volume 2022 (2022)





An Expert System for COVID-19 Infection Tracking in Lungs Using Image Processing and Deep Learning Techniques

Umashankar Subramaniam , M. Monica Subashini, Dhafer Almahles, Alagar Karthick , and S. Manoharan 
Research Article (17 pages), Article ID 1896762, Volume 2021 (2021)

Combined Feedback Feedforward Control of a 3-Link Musculoskeletal System Based on the Iterative Training Method

Amin Valizadeh  and Ali Akbar Akbari 
Research Article (8 pages), Article ID 8701869, Volume 2021 (2021)

A Review of Methods of Diagnosis and Complexity Analysis of Alzheimer's Disease Using EEG Signals

Mahshad Ouchani , Shahriar Gharibzadeh , Mahdiah Jamshidi , and Morteza Amini 
Review Article (15 pages), Article ID 5425569, Volume 2021 (2021)



CST: A Multitask Learning Framework for Colorectal Cancer Region Mining Based on Transformer

Dong Sui , Kang Zhang, Weifeng Liu, Jing Chen, Xiaoxuan Ma, and Zhaofeng Tian 
Research Article (8 pages), Article ID 6207964, Volume 2021 (2021)


Privacy Protection and Secondary Use of Health Data: Strategies and Methods

Dingyi Xiang  and Wei Cai 
Review Article (11 pages), Article ID 6967166, Volume 2021 (2021)



Identification of Specific Cell Subpopulations and Marker Genes in Ovarian Cancer Using Single-Cell RNA Sequencing

Yan Li , Juan Wang, Fang Wang, Chengzhen Gao, Yuanyuan Cao, and Jianhua Wang 
Research Article (27 pages), Article ID 1005793, Volume 2021 (2021)



A Two-Phase Mitosis Detection Approach Based on U-Shaped Network

Wenjing Lu 
Research Article (8 pages), Article ID 1722652, Volume 2021 (2021)

A Pyramid Architecture-Based Deep Learning Framework for Breast Cancer Detection

Dong Sui , Weifeng Liu, Jing Chen, Chunxiao Zhao, Xiaoxuan Ma, Maozu Guo, and Zhaofeng Tian 
Research Article (10 pages), Article ID 2567202, Volume 2021 (2021)





Review of Breast Cancer Pathological Image Processing

Ya-nan Zhang , Ke-rui XIA , Chang-yi LI, Ben-li WEI, and Bing Zhang
Review Article (7 pages), Article ID 1994764, Volume 2021 (2021)




Machine Learning Models for Survival and Neurological Outcome Prediction of Out-of-Hospital Cardiac Arrest Patients

Chi-Yung Cheng , I-Min Chiu , Wun-Huei Zeng, Chih-Min Tsai, and Chun-Hung Richard Lin 
Research Article (8 pages), Article ID 9590131, Volume 2021 (2021)

Identification of Novel Choroidal Neovascularization-Related Genes Using Laplacian Heat Diffusion Algorithm

Minjie Sheng , Haiying Cai, Ming Cheng, Jing Li , Jian Zhang , and Lihua Liu 
Research Article (10 pages), Article ID 2295412, Volume 2021 (2021)




A New Method for Syndrome Classification of Non-Small-Cell Lung Cancer Based on Data of Tongue and Pulse with Machine Learning

Yu-lin Shi , Jia-yi Liu, Xiao-juan Hu, Li-ping Tu, Ji Cui, Jun Li, Zi-juan Bi, Jia-cai Li, Ling Xu , and Jia-tuo Xu 
Research Article (14 pages), Article ID 1337558, Volume 2021 (2021)

Doctor Recommendation Model Based on Ontology Characteristics and Disease Text Mining Perspective

Chunhua Ju , and Shuangzhu Zhang 
Research Article (12 pages), Article ID 7431199, Volume 2021 (2021)

Whole Volume Apparent Diffusion Coefficient (ADC) Histogram as a Quantitative Imaging Biomarker to Differentiate Breast Lesions: Correlation with the Ki-67 Proliferation Index

Yuan Guo, Qing-cong Kong, Li-qi Li, Wen-jie Tang, Wan-li Zhang, Guan-yuan Ning, Jun Xue, Qian-wei Zhou, Ying-ying Liang , Mei Wu , and Xin-qing Jiang 
Research Article (9 pages), Article ID 4970265, Volume 2021 (2021)

Research Article

EEG-Based Multiword Imagined Speech Classification for Persian Words

M. R. Asghari Bejestani , Gh. R. Mohammad Khani , V. R. Nafisi , and F. Darakeh 

Electrical & IT Department, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran

Correspondence should be addressed to Gh. R. Mohammad Khani; mhmdreza46@yahoo.com

Received 30 April 2021; Revised 27 October 2021; Accepted 28 November 2021; Published 19 January 2022

Academic Editor: Yue Zhang

Copyright © 2022 M. R. Asghari Bejestani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study focuses on providing a simple, extensible, and multiclass classifier for imagined words using EEG signals. Six Persian words, along with the silence (or idle state), were selected as input classes. The words can be used to control a mouse/robot movement or fill a simple computer form. The data set of this study was 10 recordings of five participants collected in five sessions. Each record had 20 repetitions of all words and the silence. Feature sets consist of normalized, 1 Hz resolution frequency spectrum of 19 EEG channels in 1 to 32 Hz bands. Majority rule on a bank of binary SVM classifiers was used to determine the corresponding class of a feature set. Mean accuracy and confusion matrix of the classifiers were estimated by Monte-Carlo cross-validation. According to recording the time difference of inter- and intraclass samples, three classification modes were defined. In the long-time mode, where all instances of a word in the whole database are involved, average accuracies were about 58% for Word-Silence, 60% for Word-Word, 40% for Word-Word-Silence, and 32% for the seven-class classification (6 Words+Silence). For the short-time mode, when only instances of the same record are used, the accuracies were 96, 75, 79, and 55%, respectively. Finally, in the mixed-time classification, where samples of every class are taken from a different record, the highest performance achieved with average accuracies was about 97, 97, 92, and 62%. These results, even in the worst case of the long-time mode, are meaningfully better than random and are comparable with the best reported results of previously conducted studies in this area.

1. Introduction

Brain-Computer Interface (BCI) may be defined as a system that translates brain signals into other kinds of outputs [1]. Electroencephalography (EEG) signals are widely used in the development of the BCI system as well as other investigations regarding information extraction from the brain [2–5].

BCIs are usually concentrated on motor imagery, speech imagination, and image perception tasks. In motor imagery, the imagination of the movement of the hands, feet, eyes, tongue, or other muscles is examined. Usually, there are no physical movements due to disabilities or even absence of the whole body part. EEG signals are used to detect these types of imagination and perform the suitable actions like controlling a wheelchair or moving a robotic arm [6–10].

In speech imagination, also known as Silent-Talk and Silent-Speech, the participants imagine the pronunciation of a particular vowel [2, 11–13], syllable [14–17], or word [18–22] in some defined time intervals. EEG signal during these intervals is processed to determine the imagined word [17, 19, 23–25].

For image perception tasks, the subjects are watching some displayed pictures, for example, simple geometric shapes (rectangle, circle, triangle, etc.), real pictures (persons, animals, planets, objects, etc.), or even written words and letters. The BCI output resembles the type of picture (e.g., the picture is an animal or a planet), kind of the shape (circle or triangle), the person's name or characteristics (known or unknown, male or female, friend or foe, etc.), or whether the shown letter is what the subject has in his/her mind or not. P300 Event Related Potential (ERP) is a

well-known and useful feature of EEG signals for these kinds of BCIs [26, 27].

In this study, we examined imagined word recognition for six Persian words and the silence/rest state. The aim was to achieve a suitable classifier structure, which can be used with different number of classes (words), using a minimum processing power for real-time applications. The research process and a summary of some interesting results obtained during the experiment are demonstrated in the succeeding sections.

2. Material and Methods

2.1. Data Set

2.1.1. Subjects and Instruments. Five male subjects participated in the experiment. They were all healthy, right-handed, and aged between 25 and 45 years. Although the number of subjects is not so high, it is not necessary to incorporate more subjects because all recordings, processing, classifier training, and performance evaluations are done individually for each subject. Interpersonal measures and comparisons are used to assure consistency of the results between different subjects.

To record EEG signals, we used an EEG-3840 EEG set from Negar Andishgan LTD, Tehran, Iran. The system has 21 EEG channels along with 8 external, two ECG, and one EMG channel. Sampling frequency (F_s) was set to 500 samples per second. EEG electrodes were positioned on subject's head and kept in place by a string headset, based on the 10-20 standard (Figure 1).

2.1.2. Test Procedure. In order to acquire EEG signals during silent talk (imagined word repetition), a test procedure was designed. To implement the tests, we developed a special software, called *Test Generating Application (TGA)*. Figure 2 illustrates time sequence of tests.

Each "test" consists of several (N_t) "trials" of some (N_c) chosen "words" (or classes) in a pseudorandom order. For example, assume three words $W = \{\text{red, blue, black}\}$ with 5 trials ($N_c = 3, N_t = 5$). A single trial has exactly one instance of each word in a random order, e.g., (blue, red, black) or (red, black, blue) and so on. A test (T) is a concatenation of 5 trials (a string of 15 words) like:

$T = (\text{blue, red, black, red, black, blue, blue, black, red, black, red, blue, red, black, blue}).$

The TGA program generates these random sets and presents them to the subjects by vocal and/or visual stimulators. In the vocal mode, according to each word instance in the test, a sound or voice is played via earphones. In the visual mode, a shape or word will be displayed on the monitor. The time interval between presenting two consecutive words in a test is called *instance time* (T_i) and is constant during each test. In visual mode, the time of display of each signal is also a constant parameter (T_d).

In this experiment, we used vocal stimulation to record EEG signals of imagined words. Subjects were asked to sit on a chair, close their eyes, and listen to the words played on his earphone. When each word played completely, the

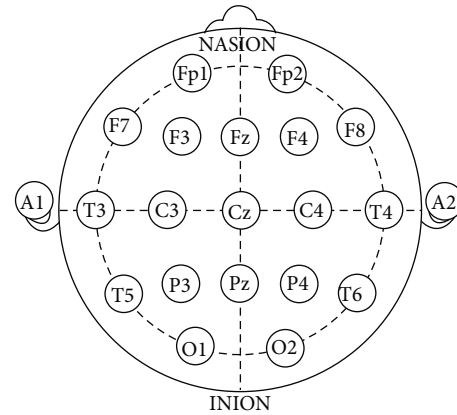


FIGURE 1: 21 electrode locations of the international 10-20 system for EEG recording [28].

subject should repeat the word silently for at least 3 times. Closing the eyes helps the subject to concentrate on test procedure. It also eliminates eye blinking, which is the most major EEG artefact. To minimize other kinds of artefacts, the subjects were also asked to be comfort but make no movement by tongue, lips, eyes, or any other organs or muscles. Also, an upper limit was considered for total test time, because elongated test times leads the subjects to become tired and lose their attention.

Subject's EEG signal during every test were continuously recorded and saved to a single file. Also, TGA generates some synchronization signals, which indicates the type and start times of each word instance. These signals were merged into EEG data via external inputs of EEG-3840.

2.1.3. Data Set Structure. We selected six Persian words for this experiment: {بالا، پایین، چپ، راست، بله، خیر}. They are pronounced as {(bala), (pāyin), (chæp), (rast), (bæle), (kheir)} and are equivalents of {Up, Down, Left, Right, Yes, No}, respectively. These words have been selected for several reasons, such as the following:

- (i) They are complete and meaningful Persian words
- (ii) They can be used for navigating a mouse/wheelchair or filling a simple questionnaire form
- (iii) Half of them (chæp, rast, and kheir) have only one syllable, while the others have two
- (iv) We can divide them into three pairs with opposite meanings, i.e., {Up, Down}, {Left, Right}, and {Yes, No}
- (v) The pairs also have all three possible combinations of syllable counts, (two, two), (one, one), and (one, two)
- (vi) The same pairing combines the two words, which are usually used together

Another special item, the Silence, was also added to the above word list to indicate the "rest" or "no word processing" state of the brain.

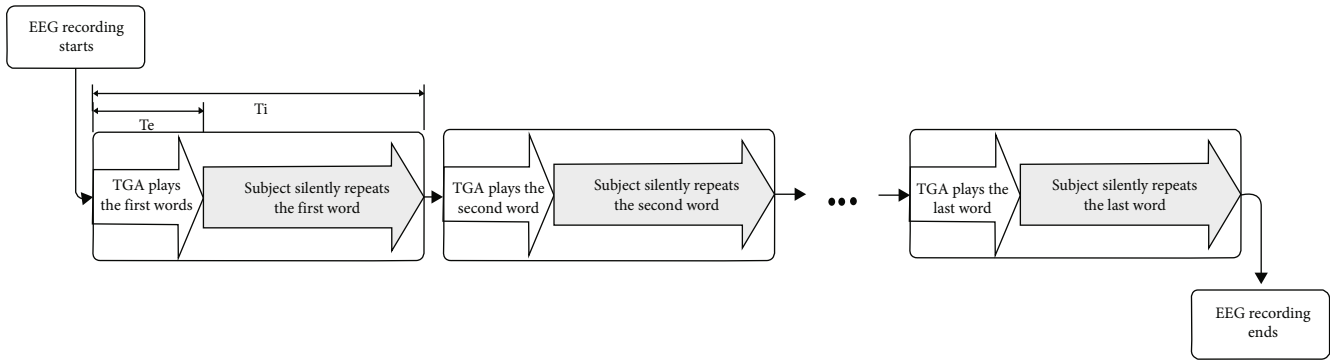


FIGURE 2: Time sequence of a test's record.

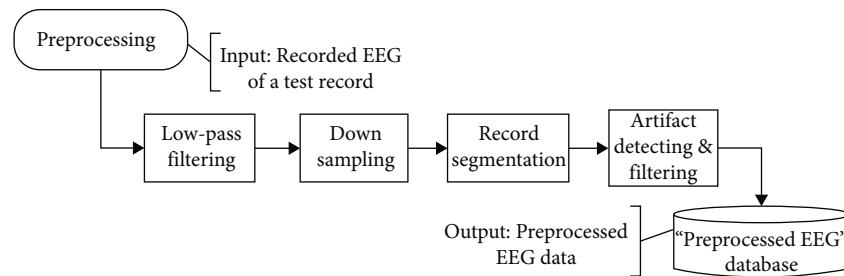


FIGURE 3: Preprocessing steps.

Playing (or repeating) time for each of selected words with normal speed is about 500 to 1000 mSec, so we considered instance time to be 4 seconds ($T_i = 4000$). If we limit the total test time to be about 4 minutes, then each test can contain 60 word instances. Thus, the six words were divided into three tests, each with two words and the Silence ($N_c = 3$) and set $N_t = 20$ trials. Specifically, there were three tests:

(T1) $W = \{\text{Up, Down, Silence}\}$; $N_t = 20$; $T_i = 4000$

(T2) $W = \{\text{Left, Right, Silence}\}$; $N_t = 20$; $T_i = 4000$

(T3) $W = \{\text{Yes, No, Silence}\}$; $N_t = 20$; $T_i = 4000$

Each subject incorporated in five “sessions,” each in a single day with at least a one-week interval. Each session consisted of two “parts” separated by a short break. In any part, all three tests were recorded so that all of the selected words were used in every part.

Finally, the raw EEG data set consists of 30 records (5 sessions \times 2 parts \times 3 records) for every subject, containing 200 instances (5 sessions \times 2 parts \times 20 trials) for each of the six selected words and 600 instances (200 \times 3 tests) of “Silence.” Each record had 19 EEG channels, with two additional sync channels sampled at 500 Hz with a 16-bit accuracy.

2.2. Preprocessing. All the recorded data were preprocessed in offline and EEG data for each word instance was saved as a separate record in *Pre-processed EEG Database*. Preprocessing consists of following steps (see Figure 3).

2.2.1. Low-Pass Filtering. Frequency bands of EEG signals spans from DC to Gamma band (above 30 Hz). Sampling frequency (F_s) was set to 500 Hz, thus the recorded EEG signals had frequency components up to 250 Hz, which is

far beyond the useful EEG frequency bands. In order to suppress higher frequencies and specially the power line noise (50 Hz), we used a 0-32 Hz low-pass filter (LPF).

2.2.2. Subsampling. After removing high frequency components, we down sampled the EEG signals from 500 Hz to 100 Hz by replacing each five consecutive samples with their average. This reduced the size of the recorded data to 1/5 of its original size and improved the speed of subsequent processes.

2.2.3. Record Segmentation. As mentioned above, each continuous record of EEG signals, consisted of $N_t = 20$ instances of two words and the Silence. In this step, EEG data of word instances were separated from each other. This was done by using the “Sync Data”, which was generated by TGA and saved in separate channels of EEG signal.

Although all word instances were assumed to have the same time length ($T_i = 4$ sec), EEG signals for each segment were cropped (or padded with zeros) to have exactly the same number of samples.

2.2.4. Artifact Detection and Filtering. EEG signals may always be corrupted by various kinds of artefacts such as blinking, eye movement, wanted or unwanted movements, and so on. In our experiment, due to closed eyes during EEG recording, the major artefacts of blinking and eye movement have been minimized. Moreover, subjects had been asked to control and avoid any additional movement or muscle activity during recordings.

Besides all of these, average energy of EEG signals during an instance has been considered as a measure of signal

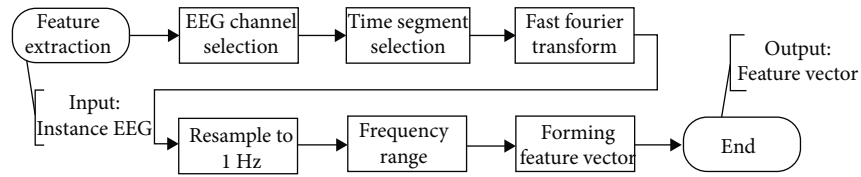


FIGURE 4: Feature extraction.

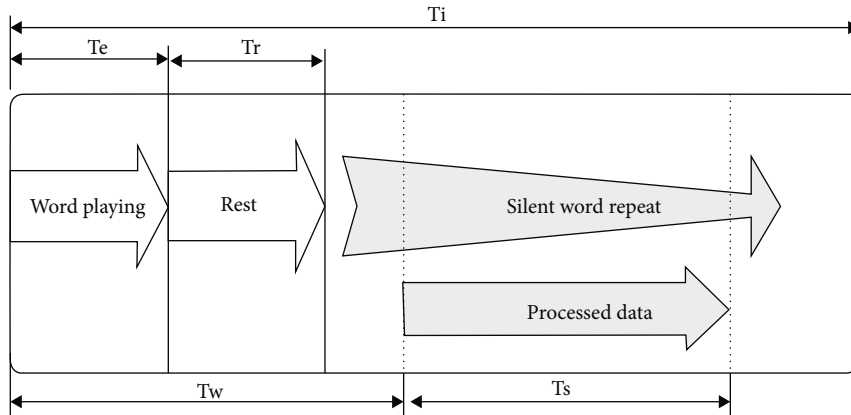


FIGURE 5: Process time selection.

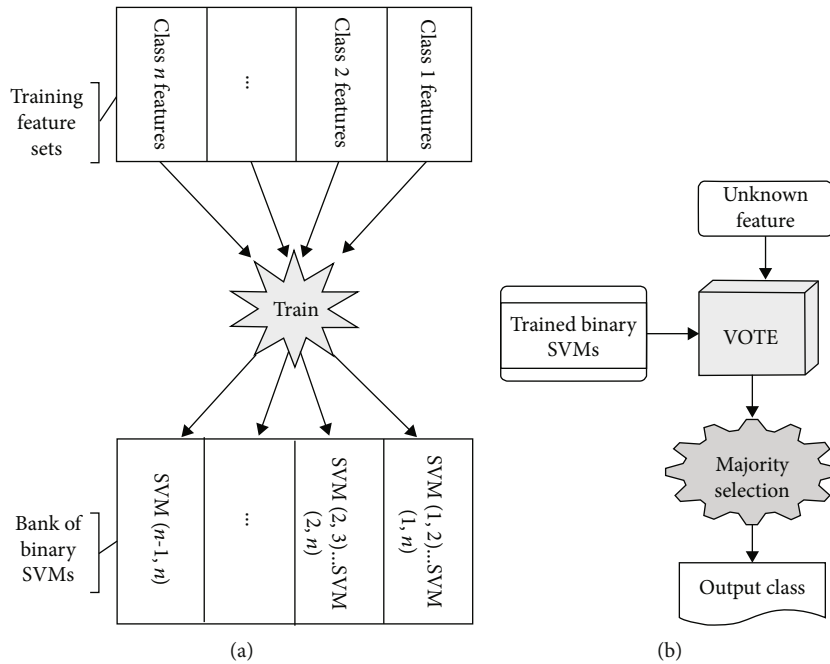


FIGURE 6: (a) Training of binary SVM machines and (b) classification algorithm.

quality. Any instance with an energy much above or below the overall average energy of whole recorded signal, has been filtered out. Total number of rejected instances was less than 2% of all recorded EEG data.

2.3. Feature Extraction. The feature vector used in this research was simply the amplitude of each single frequency in selected EEG channels. The advantage of frequency-

related parameters is that they are less susceptible to signal quality variations, which may be present due to electrode placement or the physical characteristics of subjects [5]. Figure 4 shows the feature extraction steps.

2.3.1. EEG Channel Selection. Depending on the application and features needed, only a selected set of EEG channels are used for processing, which may be one, all, or any other

TABLE 1: Class words and their class number.

Class word	ba λ	b \ae le	ch \ae p	kheir	pa γ in	ra σ t	Silence
English equivalent	Up	Yes	Left	No	Down	Right	Silence
Abbreviation	U	Y	L	N	D	R	S
Class no (whole data set)	1	2	3	4	5	6	7
Class no (single records)	10...19	20...29	30...39	40...49	50...59	60...69	70...79

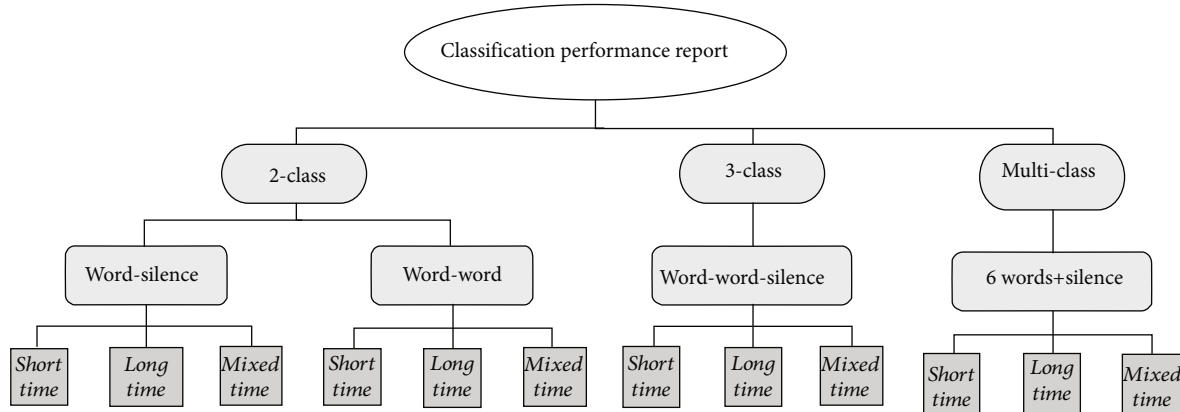


FIGURE 7: Organization of evaluation results.

combination of the channels. For the results presented in this article, we have used 19 EEG channels from 21 available EEG channels in 10-20 standard (see Figure 1). Channels A1 and A2 were not used because they are not actually scalp EEG channels. Instead, they are often used for contralateral referencing of all other EEG electrodes [28].

2.3.2. Data Time Segment Selection. The total time duration for signals of each instance of a word was T_i (4 seconds). The first portion of this time (T_e , up to 500mSec) corresponds to excitation (hearing the played word). Then, after a small rest time (T_r), the subject repeats the word silently for a few times. As illustrated in Figure 5, we only processed the EEG samples for an interval T_s , starting after T_w from beginning of the instance. Obviously, $T_w \geq T_e + T_r$ and $T_s \leq (T_i - T_w)$. Experimentally, we selected $T_w = 1$ and $T_s = 2.5$ seconds.

2.3.3. Fast Fourier Transform (FFT). To compute frequency components, we used absolute value of FFT of EEG signals. Furthermore, the amplitudes of these absolute FFTs were normalized by dividing them to their maximum value.

2.3.4. Resample to 1Hz Resolution. Time duration of the signals is T_s , therefore, the initial resolution of their spectrum is $1/T_s$. To normalize the length of feature vectors against sampling frequency and duration of the signals, we resampled FFTs to one Hertz resolution. This typically reduces the length of feature vector. Besides, since the signals were already filtered with a 32 Hz low-pass filter, only first 32 values of this resampled FFTs had valid amplitudes.

TABLE 2: Feature extraction parameters.

Parameter	Value
EEG channels	19 channels (all except A1 and A2)
Frequency range	1 to 32 Hz
T_w	1 second
T_s	2.5 seconds

TABLE 3: Word-Silence classification accuracies in all the three modes.

Classification mode	Classification accuracy (mean \pm standard deviation)
Short-time	95.7 \pm 8.3
Long-time	57.5 \pm 7.2
Mixed-time	97.1 \pm 5.1

2.3.5. Frequency Range Selection. Usually, not all of the signal frequency range are used. For example, one may filter out EEG Delta and Gamma bands by omitting first four and last two values of resampled FFT, corresponding to frequencies below 4 Hz (0 to 3) and above 30 Hz (30, 31). In this experiment, all frequency components were used, except the first (DC) value.

2.3.6. Forming Feature Vector. Finally, the selected FFT values (of all selected channels) were concatenated to form instance feature vector. If N_{ch} channels and N_f

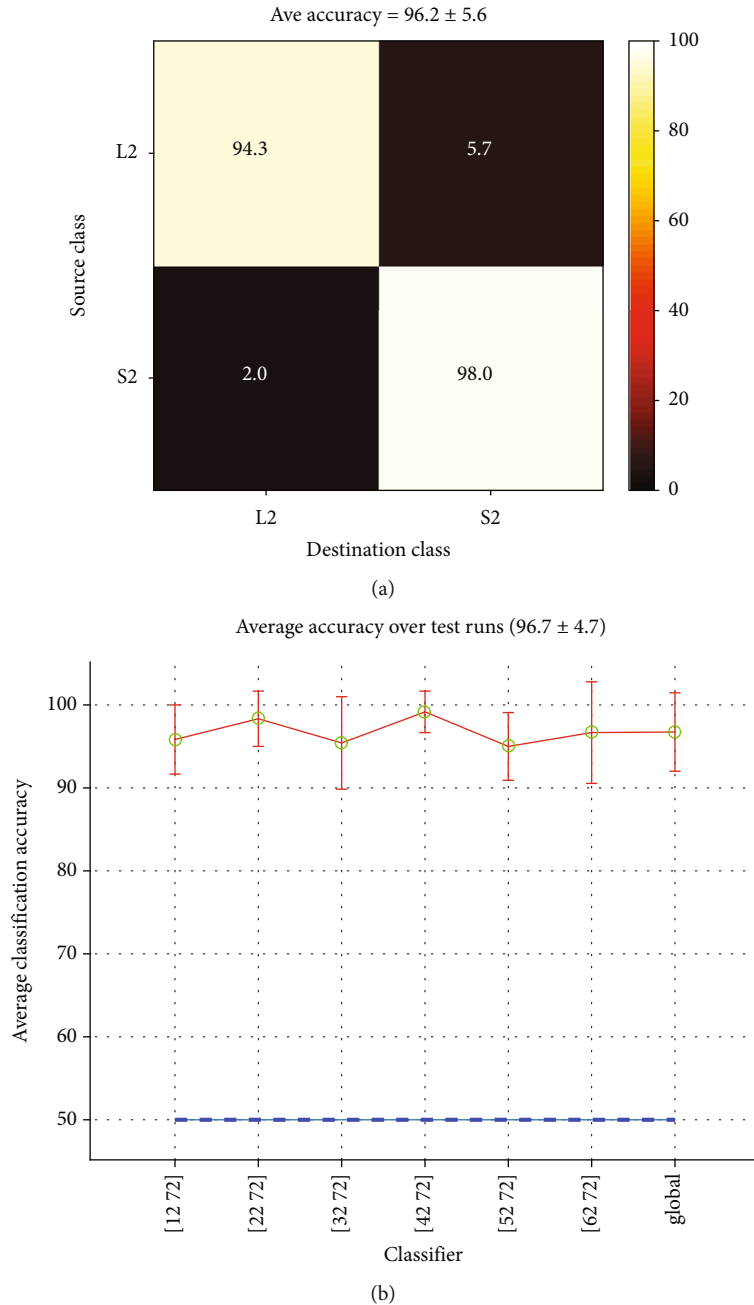


FIGURE 8: (a) Confusion matrix for [32 72] classification and (b) estimated classification accuracies for each Word-Silence pair in record No 2. The last value is the global performance.

frequency values were used, feature vector length would be $L_f = N_{ch} * N_f$.

2.4. Classifier Structure. We used binary (2-class) support vector machines (SVM) as the basis of our classification method. One SVM was trained to classify between 2 specific classes consisting of a pair of selected words or a word and the Silence. Let SVM (i, j) be the machine trained for two classes i and j . Obviously, if an arbitrary feature set is presented to such a machine, it would be classified to one of the trained classes (i or j), even though it may belong to neither.

For n -class classification, there would be $n(n-1)/2$ of such classifiers. *Majority Rule* was used for winner selection. Besides training, classification consists of the following steps (Figure 6):

- (i) The unknown feature set is presented to all binary classifiers
- (ii) If a relative majority of the classifiers vote for a single class k , then class k wins and the features are assigned to this class
- (iii) If two classes, namely k and l , had the same maximum vote counts, then the class voted by SVM (k, l) wins

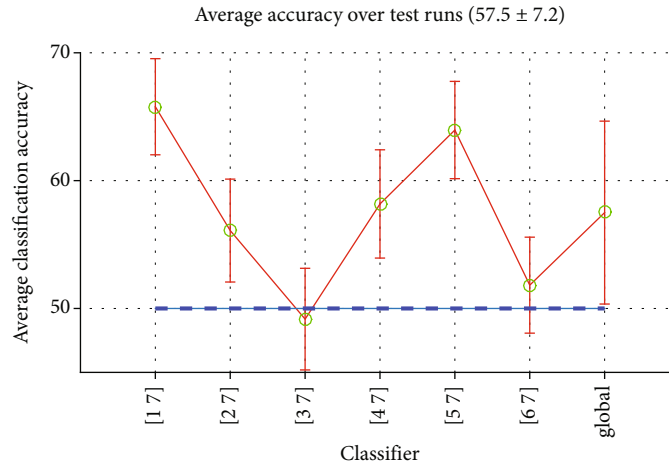


FIGURE 9: Performances of Word-Silence classifiers in long-time mode with their global average.

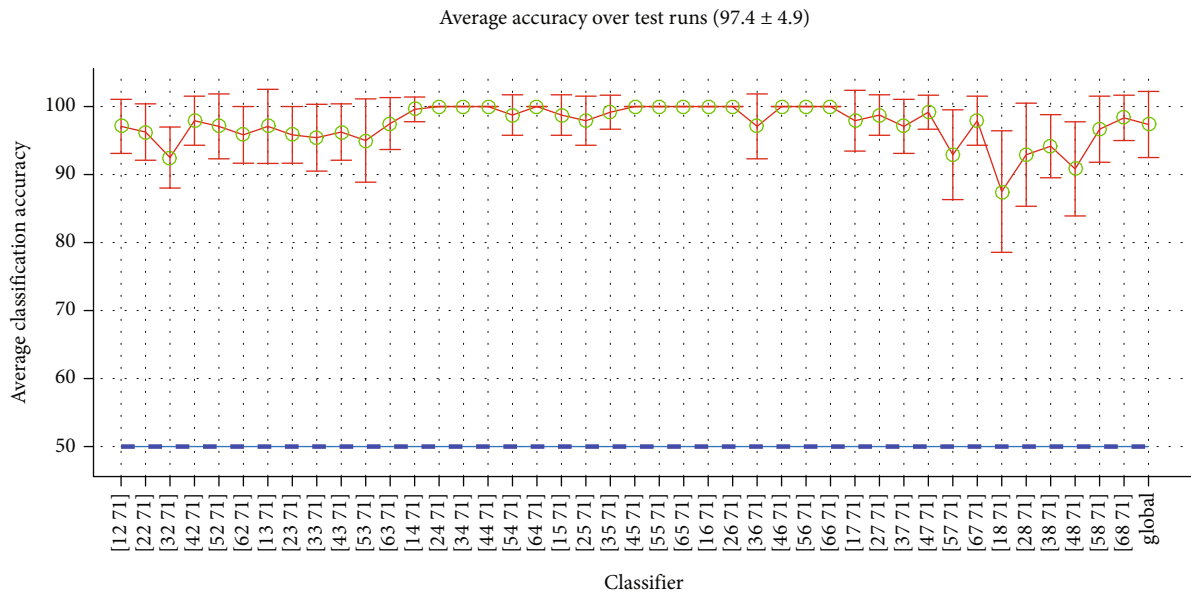


FIGURE 10: Performances of some Word-Silence classifiers in mixed-time mode and their average.

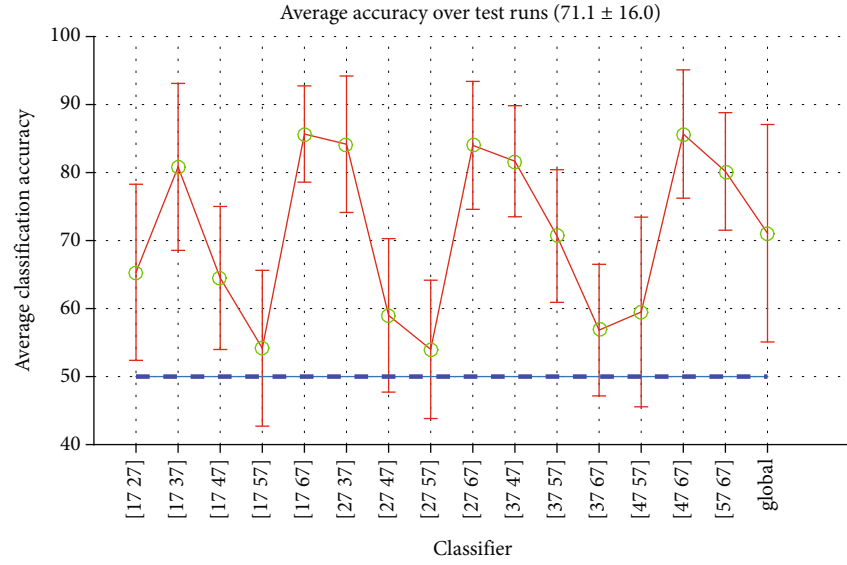
- (iv) If more than two classes had the same maximum vote counts, then output class is undefined and the input feature would be tagged to belong to special class “Unknown”

Note that if we assume that every binary classifier has a relatively good performance, then this Majority Rule is also *the Rule of Specialists*. For example, in the special classification case of two words and the silence (W1, W2 and S), if the input sample is from the S class, then the two machines SVM (W1, S) and SVM (W2, S) will realize it and the sample will be correctly assigned to S, regardless of the vote of SVM (W1, W2) machine. On the other hand, if the input belongs to one of the words’ classes, then the former two machines reject the S and vote for W1 and W2, respectively. Now the third machine, the words specialist SVM (W1, W2), would determine the correct class.

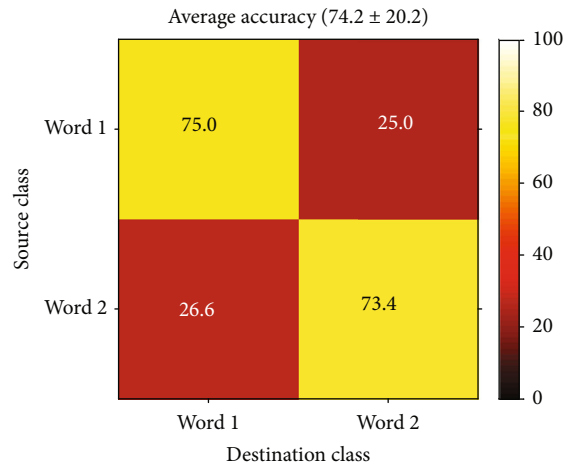
2.5. Evaluation Procedure

2.5.1. Performance Measurement. To evaluate our classifier, Monte-Carlo cross-validation was used. EEG data of proper classes were fetched from “Pre-processed EEG Database” and their features were extracted.

In each round (r) of the Monte-Carlo simulation, instant feature sets were randomly divided into two partitions: the train set, with approximately 70% of samples and the test set, with the rest 30%. As usual, SVMs were trained with train set data and used to classify features in the test set. A confusion matrix (C_r) was formed with known (true) classes as rows and resulted classes from the classifier as columns, so that the number in (i, j)-th element of the matrix is the number of features from class i , classified as class j . As mentioned above, in the case of multiclass classification, an extra column, “Unknown”,



(a)



(b)

FIGURE 11: Short-time Word-Word classification: (a) accuracies estimated for record no. 7 and (b) confusion matrix (average).

should be added to C_r . In every iteration, the accuracy of classification was computed:

$$A_r = \frac{\text{Number of correct classifications}}{\text{Total number of classifications}} = \frac{\text{trace}(C_r)}{\text{sum}(C_r)}, \quad (1)$$

where $\text{trace}(C_r)$ and $\text{sum}(C_r)$ are sums of main diagonal and all elements of C_r , respectively. When all the Monte-Carlo iterations were done, the final confusion matrix can be computed by the sum of all iteration matrices:

$$C = \sum_{r=1}^{N_r} C_r. \quad (2)$$

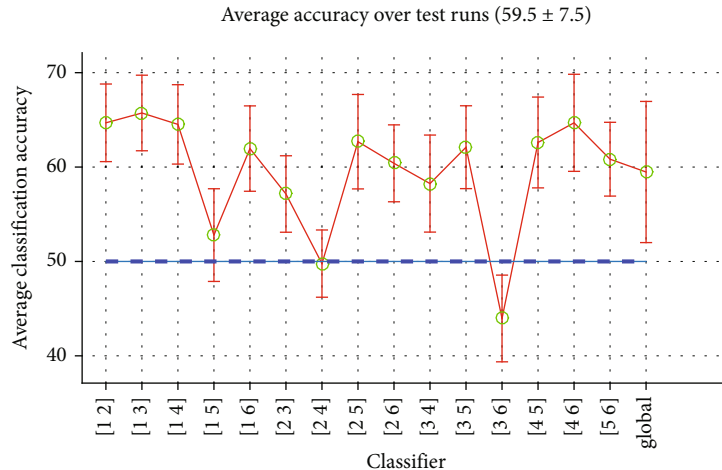
In (2), N_r is number of Monte-Carlo iterations. The final classification accuracy and its standard deviation were estimated by the following:

$$A = \frac{1}{N_r} \sum_{r=1}^{N_r} A_r \text{ and } \sigma = \sqrt{\frac{1}{N_r} \sum_{r=1}^{N_r} (A - A_r)^2} \quad (3)$$

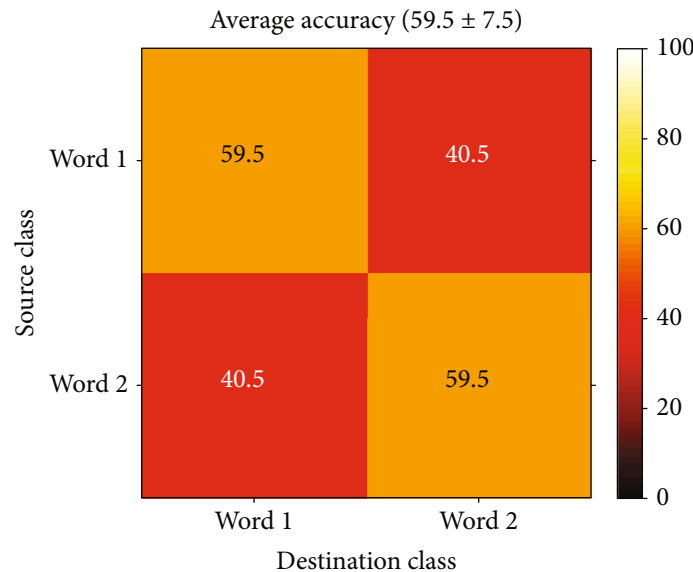
Alternatively, A can be calculated by eq. (1), with C_r replaced by C . Values of A_r were kept in a column vector called Class Measures (M_c) for further evaluation extensions.

2.5.2. Data Representation. For simplification, all names (the words or class names, EEG channel names, records, etc.) were coded with integer numbers. These codes have been used in all figures and tables of this report.

Recording parts were numbered in chronological order so that the first part of first session (the oldest part) is called



(a)



(b)

FIGURE 12: Long-time Word-Word classification: (a) accuracies of all classifiers and (b) confusion matrix (all classifiers average).

part-0 and the second part of the last session (the newest) is part-9. Similarly, every word was presented in 10 records, one record in every part, except the silence, which was present in all records. So, we may refer to part numbers of a word as “record number.” We also considered all instances of silence in a part as a single record.

Table 1 shows the class names and their code numbers. As seen in this table, each word has up to 11 class numbers. If all instances of a word in the whole database is considered, the code will be a single digit number (from 1 to 7 for six words and the silence). When only samples from a specific record have to be referenced, the code should be followed by the part number. For example, instances of word “kheir “(class 4) in the second record (part no 1) have class no 41.

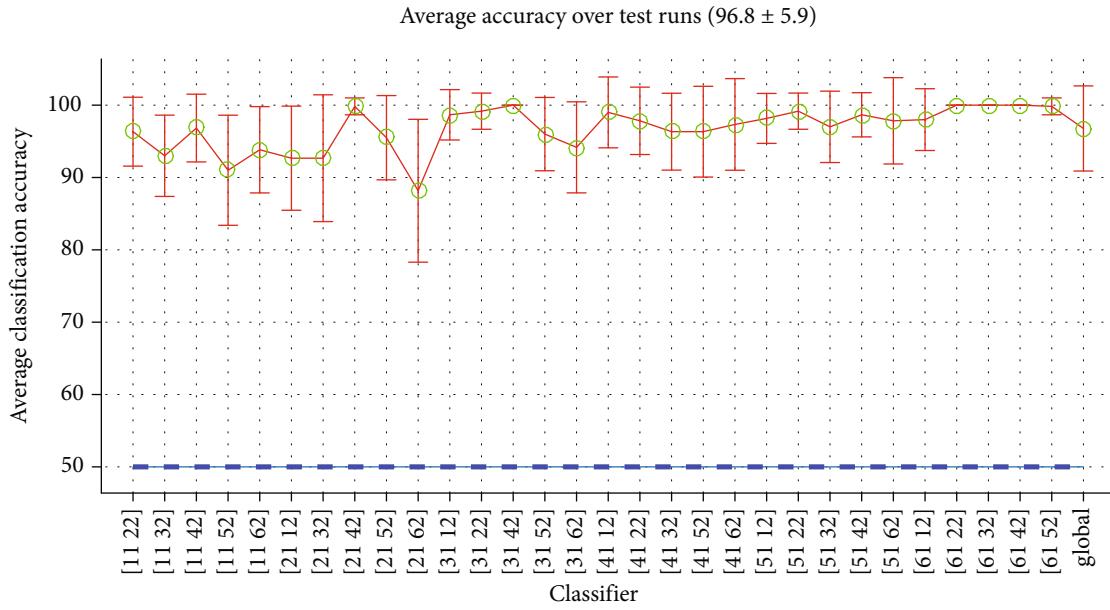
2.6. *Classification Modes.* EEG signals are not stationary, therefore the time differences between the recording times of EEG signals can directly affect the performance of classi-

fication. Based on recording time difference of inter- and intraclass samples, three *modes* of classification were distinguished in this research:

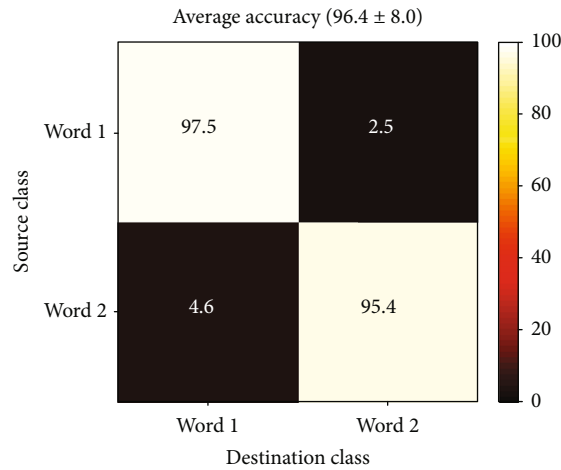
- (i) *Long-time classification* where all instances of a class in whole database are involved
- (ii) *Short-time classification* where only instances of the same record (or part) are used
- (iii) *Mixed-time classification* where samples of every class are taken from a different record

In the first mode, the difference in the recording time of samples of the same class spans from a few seconds to several weeks. This is the time during which the EEG signals of a subject were recorded. The same is true for interclass time differences.

In the second mode, intraclass time difference is at most as short as a test time (4 minutes). The interclass difference



(a)



(b)

FIGURE 13: Mixed-time Word-Word classification: (a) accuracies estimated for record no. 1 vs record no. 2 and (b) average confusion matrix (all classifiers).

would be at most equal to duration of a recording part which was kept in about half an hour.

In the third mode, the intraclass times is the same as the second mode, while the interclass time is typically many days (about an hour at least and some months at most).

3. Results Evaluation

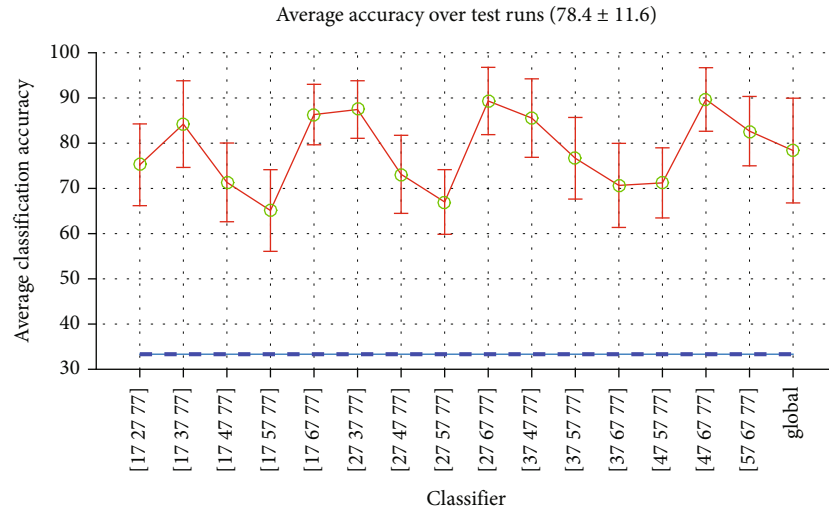
We tried several classification cases with our captured data sets, selected feature extraction methods, and designed classifiers. Some interesting results are presented below. First, the results of the simplest case of a 2-class classification are reported, then 3-class case, and, finally, some examples of the results in multiclass classifications are mentioned. In all cases, the above three classification modes are considered (Figure 7).

TABLE 4: Results of classification in Word-Word case.

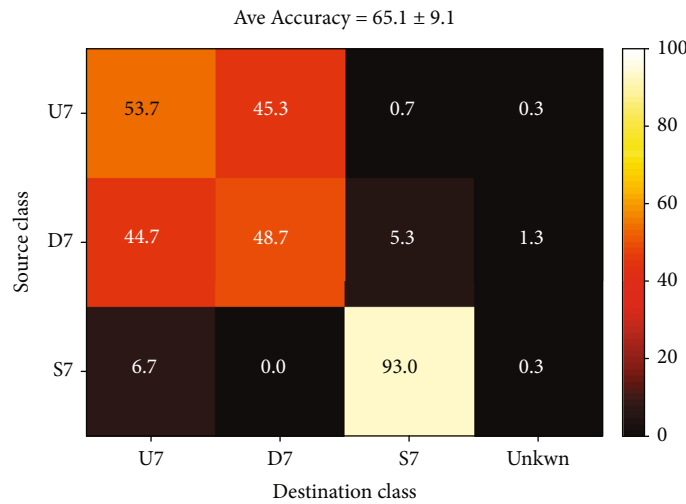
Classification mode	Classification accuracy (mean ± standard deviation)
Short-time	74.5 ± 20.4
Long-time	59.5 ± 7.5
Mixed-time	96.4 ± 8.0

TABLE 5: Results of classification in Word-Word-Silence case.

Classification mode	Classification accuracy (mean ± standard deviation)
Short-time	78.9 ± 13.9
Long-time	39.5 ± 5.4
Mixed-time	92.0 ± 11.1



(a)



(b)

FIGURE 14: Short-time Word-Word-Silence classification, (a) accuracies estimated for record no. 7, (b) Confusion matrix of the worst case in record 7 (17, 57, 77).

To avoid making the manuscript lengthy, feature extraction parameters are kept constant during this report. The selected set of parameters are shown in Table 2. It is almost optimal [29], but there may be other combinations, which produce comparable results. Furthermore, otherwise specified, the results are average of all subjects.

3.1. Two-Class Type. First, we examined the simple case of a two-word classification. In this case, there was only one SVM which was trained with features extracted from two specific classes. Different combinations of classes and/or feature parameters were examined.

3.1.1. Word-Silence. This type of classification is important for discriminating talking/not-talking states, especially in real-time silent-talk applications. Table 3 summarizes some results of classifying EEG signals corresponding to silent repeating of a single word, chosen from the above six-word set, and the “Silence” case.

(1) Short-Time Classification. To explain the accuracies in Table 3, we start with classifying between two classes 32 and 72 (instances of the word “Left” and the Silence in record no. 2, or L2 and S2, see Table 1). Here, we only work with EEG data of subject no. 1. When validating this SVM classifier by the Monte-Carlo cross-validation method, confusion matrix (C_r) is calculated in each iteration, and accuracy (A_r) is estimated by eq. (1). At last, the final confusion matrix (C) is obtained by eq. (2) and all (A_r) are concatenated to form a column vector named M_c (see Section 2.5.1).

Figure 8(a) shows a typical confusion matrix (C) for this classifier. The values in this matrix were calculated from the results of 30 Monte-Carlo iterations. Element colors are graphical illustrations of their values. As mentioned, the value in row i and column j is the percent of samples from class i , classified as class j . For example, 94.3% of total test samples of the word “Left” were correctly identified, but 5.7% of them were falsely classified as “Silence.” From eq.

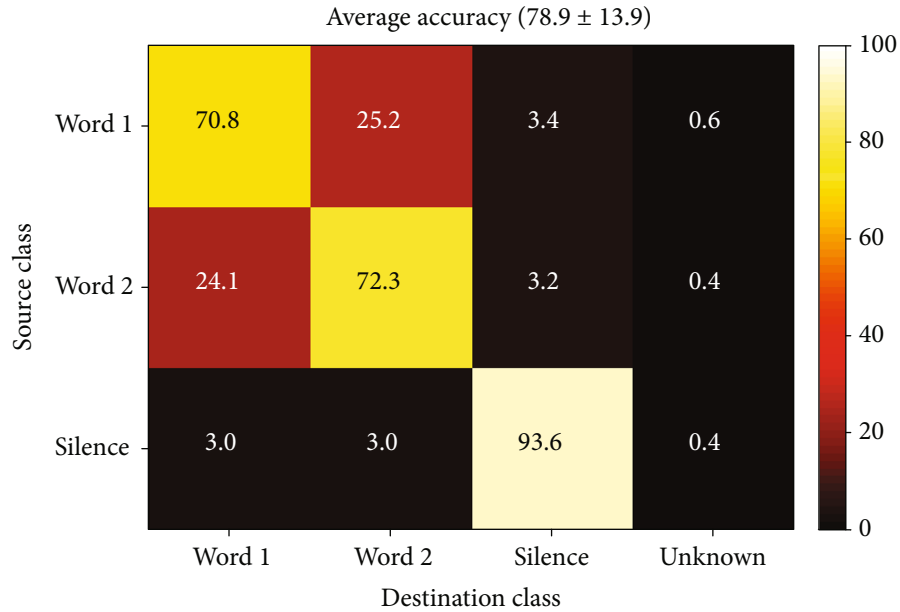


FIGURE 15: Short-time Word-Word-Silence classification confusion matrix (overall average).

(3) and data in M_c , the average accuracy for this classifier was $A = 96.2\%$ with a standard deviation $\sigma = 5.6\%$.

We repeat the same procedure for all the six words in record no. 2 (classes 12, 22, 32, 42, 52, and 62 against class 72) and concatenate all (M_c) into a matrix M . The average accuracy for each word classifier and its standard deviation can be estimated from columns of M . We can also calculate the global mean and standard deviation for all data in M as a measure of performance for the task of discriminating a word from silence in this record. These values are shown in error-bar of Figure 8(b). In this figure, green circles show classifier's average accuracy and red bars indicate the standard deviation around the mean. The last bar is the global mean and standard deviation. The horizontal dashed blue line shows the chance level accuracy, which is 50% for balanced binary classification. Regarding the definition of classification modes (Section 2.6), this is a short-time classification because only instances of words and silence in the same record are used.

A better estimation for the accuracy can be obtained by averaging the performances over all records and all subjects. This is the value recorded as the accuracy of short-time Word-Silence classification, in the first row of Table 3 ($95.7 \pm 8.3\%$).

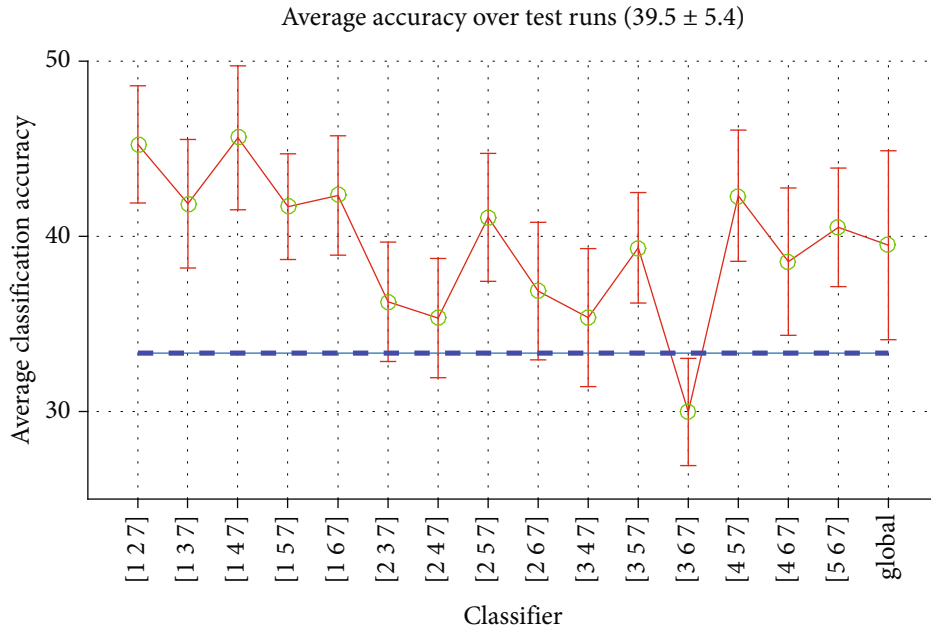
(2) *Long-Time Classification*. In the long-time mode, all instances of a word and silent from each particular subject are used to train and test the classifier. Figure 9 shows the performances of classifiers in this case. As seen, the classifiers have generally a poor performance and in some cases they even did not better than chance. The global accuracy is weakly above 50%, thanks to the relatively good performance of (1, 7, 5, and 7) classifiers.

(3) *Mixed-Time Classification*. In mixed-time mode, every class data should be picked from a separate record of the

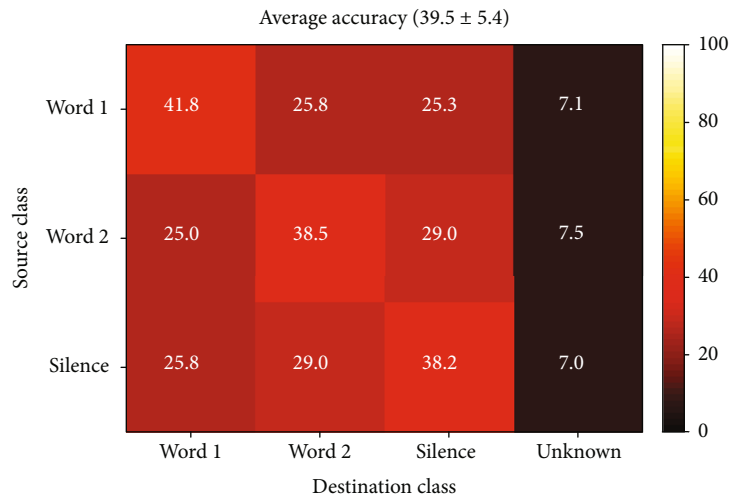
same subject, e.g., the word "Up" of the third record (class no. 12) and "Silence" from the second record (class no. 71). Hundreds of such classifiers can be defined for every subject. Figure 10 shows performances for some of these Word-Silence classifiers in mixed-time mode using EEG data of subject no. 1. As may be expected, classification performance in this mode is generally better than both previous modes. In some cases, accuracy reaches to 100%, and, in a worst case, it remains above 80%. The global average accuracy is also slightly better than the short-time mode.

3.1.2. *Word-Word*. For this case, there was also one SVM machine, trained with feature vectors of instances of a selected pair of words. The procedure and results in this case were the same as Word-Silence, with the silence replaced by a word. Three classification modes were considered, but since there were several words which can be picked up for the second class, the total number of possible combinations was much more than the previous case.

(1) *Short-Time Classification*. In this case, words instances of a recording part were compared together. In each part, there were 15 ($6 \times 5/2$) possible combinations. Figure 11(a) shows the accuracies estimated for record no. 7 of subject no. 1. There were seven classifiers with relatively good performance (80-90%), three were moderate (60-80%), and the other five had poor accuracies (50-60%). Three of the weakest results belonged to classifiers (17, 57), (27, 47) and (37, 67), specifically classifiers with words (Up, Down), (Yes, No) and (Left, Right). Note that these are the words which were paired into three designed tests T1, T2, and T3 (sec. 2.1.3) and therefore had the smallest interclass time difference, nearly the same as their intraclass distance (a few seconds up to less than 4 minutes).



(a)



(b)

FIGURE 16: Long-time estimated Word-Word-Silence classification, (a) accuracies, (b) confusion matrix.

The confusion matrix for this type of classification is shown in Figure 11(b). Global average accuracy was about 74.2 percent with a relatively large standard deviation ($\sigma = 20.1$), which means a possible but not so reliable classification.

(2) *Long-Time Classification.* Using the long-time data reduces the total iterations needed to estimate Word-Word classification performance for each subject to 15. The results are illustrated in Figure 12(a). Although the accuracy levels have decreased compared to Figure 11(a), the graph shapes are roughly the same: three classifiers with chance level accuracies, a few in middle, and a majority in higher levels of accuracy. The smallest accuracies belonged to the same classifiers working on words which were paired in tests.

Figure 12(b) is the average confusion matrix for all classifiers. The average accuracy is about 60% with a standard

deviation equal to 7.5 which again has a breaking higher performance than random classification.

(3) *Mixed-Time Classification.* In mixed time, each word in a record are compared with other words in another record. Classification results of words in record no. 1 against record no. 2 of subject no. 1 have been shown in Figure 13(a). Nearly all classifiers have excellent performance compared with other classification modes (90-100%). The overall accuracy, calculated by averaging over thousands word/record combination was 96.4% with $\sigma = 8.0$. Average confusion matrix is shown in Figure 13(b).

Table 4 summarizes overall accuracies for Word-Word classification in all the three modes. Compared with the corresponding values for Word-Silence classification, accuracies were only slightly reduced in long- and mixed-time modes.

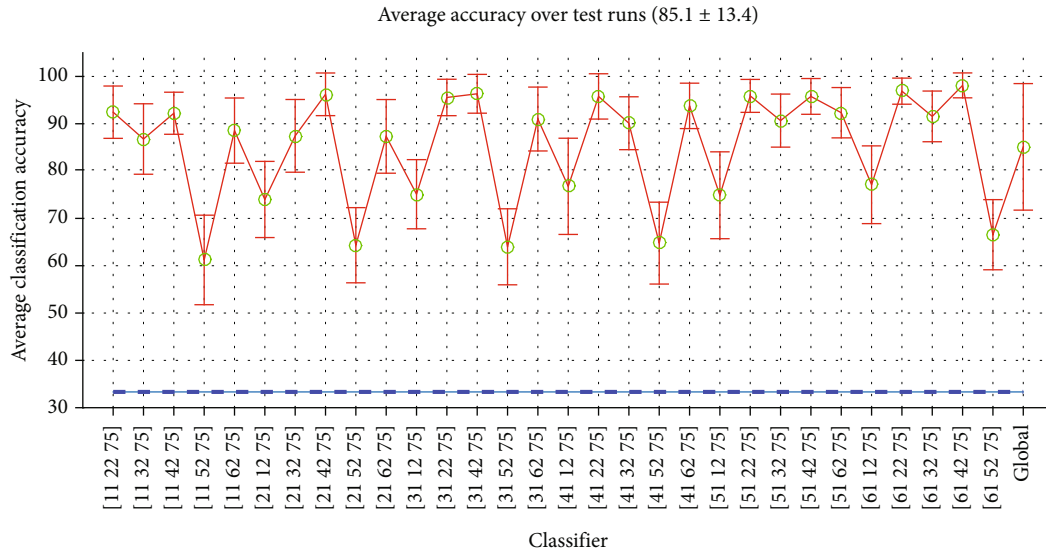


FIGURE 17: Mixed-time Word-Word-Silence classification accuracies estimated with words from records no. 1 and 2 and silence form record no. 5.

In the short-time mode, performance reduced noticeably due to low accuracies in discriminating words which were paired in tests.

3.2. Three Classes. In the three-class classification, there are three possible classes ($C1$, $C2$, and $C3$) and the classifier consists of three binary SVMs, ($C1$, $C2$), ($C1$, $C3$), and ($C2$, $C3$). If a sample with unknown class C is presented to the three SVMs, then with the majority rule which we have selected for the classifier (see Section 2.4), there would be only two possible states:

- (i) If two classifiers agree on their common class (Cx) then the unknown sample will be assigned to this class ($C = Cx$). Of course the third machine cannot agree with this choice
- (ii) If every SVM has its own choice, not common to any other machine, then the sample C will remain unknown and will be assigned to the class of "Unknowns" ($C0$)

An important case of this classifier type is the two words and silence classifier (WWS). Table 5 shows the summarized results for three-class WWS classification in all three modes.

3.2.1. Short-Time Classification. Figure 14(a) shows the classification results of WWS classification on EEG data from record no. 7 of subject no. 1. The Word-Word classification results of this record were shown in Figure 11(a). The shape of the two curves are very close together, but the accuracies in WWS case are slightly greater than the two-class case with the same words. Furthermore, the chance level in three class case is about 33% compared with 50% in binary classifiers. These made the resulting performance to be much better than Word-Word classifier.

Besides the overall good quality of classification, the worst cases were again the classifiers which worked on the

words with the minimum interclass time difference. Figure 14(b) shows confusion matrix of the worst classifier in record no. 7, the (17, 57, 77) or more clearly the (U7, D7, S7) classifier. It is clear that the classifier almost does not distinguish between the two words but has a good performance for the silence.

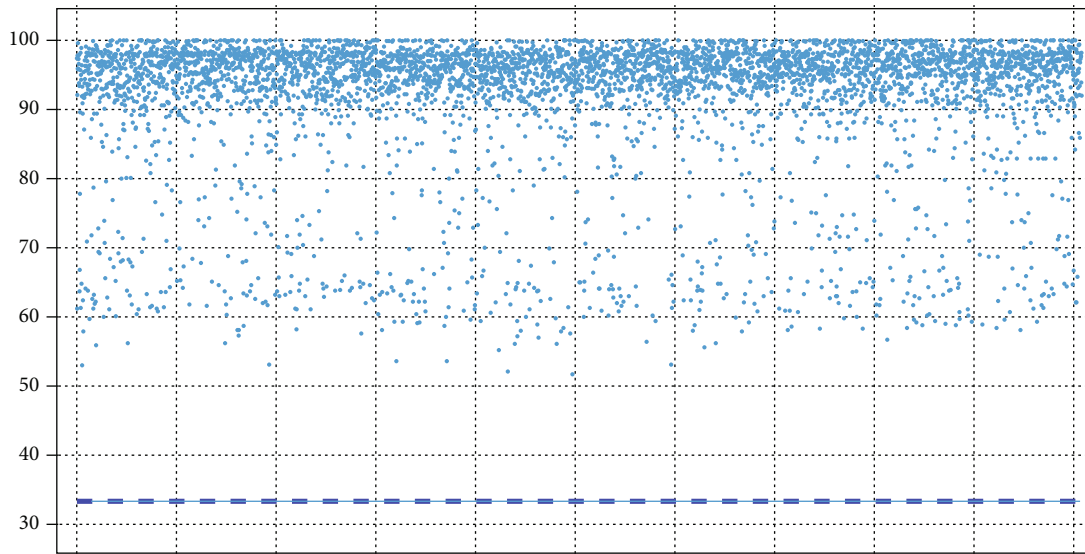
The global average accuracy was calculated for all short-time mode WWS classifiers of all records whose results indicated a rate of $78.9 \pm 13.9\%$ which is shown with its confusion matrix in Figure 15.

3.2.2. Long-Time Classification. Here the classifiers are defined over three different words in the whole recordings. Figure 16(a) shows the accuracies of all 15 classifiers in this category. The accuracies were generally low (all below 50%) and in some cases were close to random decision. The overall accuracy was above the chance by a very narrow margin (39.5% with $\sigma = 5.4$). Figure 16(b) is the average confusion matrix of all classifiers.

3.2.3. Mixed-Time Classification. Figure 17 shows some results for mixed mode three class WWS classification. The words combinations was the same as Figure 13(a) where the first word was picked from record no. 1 and the second word comes from record no. 2 of subject no. 1. Silence samples were from record no. 5.

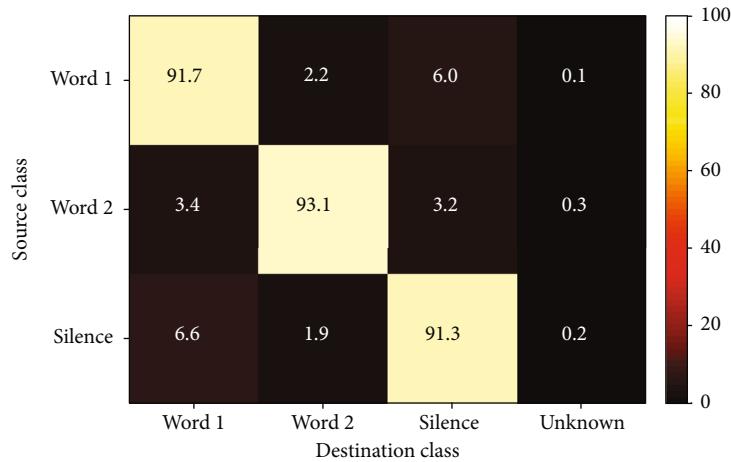
As seen in most cases, the classification accuracy was well above 85%. When second word was taken from class no. 12 the accuracy reduced to about 75% and if the word was from class no. 52 the accuracy again decreased further to 60-70%. This shows a low performance for (12, 75) and (52, 75) SVM machines. The average accuracy in about 85% which is much lower than 98% of Word-Word case.

Figure 18(a) shows a scatter of accuracies of several thousands of WWS classifiers with a unique combination of classes. Figure 18(b) is the overall confusion matrix of classifiers. The global average accuracy of all classifiers was



(a)

Average accuracy (92.0 ± 11.1)



(b)

FIGURE 18: Mixed-time Word-Word-Silence classification: (a) accuracies of thousands of WWS classifiers with unique combination of classes and (b) average confusion matrix for all classifiers.

about 92% which is not as good as Word-Silence and Word-Word classifiers but is still a good result. Note that in about 0.6 percent of cases a feature set could not be classified.

3.3. Multiclass Classification

3.3.1. Short-Time Classification. As the first example for multiclass classification, here we take a quick glance at the classification of all the seven classes in a single record. Obviously, this is a short-time classification mode with seven classes. The classifier has 21 SVM machines which will select the output class by the Majority Rule. Figure 19(a) presents the confusion matrix of such classifier for all words in record no 8 of subject no. 1. The average accuracy was about 55%. The silence had the maximum single class accuracy (about 95%), but no other single word accuracy exceeded 50%. The largest error rates were between the words paired into a test, as seen, for example, in discrimination of L8 and R8.

Note that the random selection accuracy in this classifier was $1/7$ or about 14 percent. Figure 19(b) is the average confusion matrix for all records. The average accuracy was again about 55% and ranged from 47.5% to 65.1% for single records.

3.3.2. Long-Time Classification. This case is similar to the short-time case except that similar words from all records are tagged as a single class. The accuracy of classification in this case was about 32%. Although this accuracy is about 2 times the chance level, it cannot be considered a good result. Once again, the highest error rate belonged to “Left” and “Right” classes. Figure 20 shows the confusion matrix of this classifier.

3.3.3. Mixed-Time Classification. For the classification of all the seven classes in mixed-time mode, we should choose every class sample from a different record. Since there are

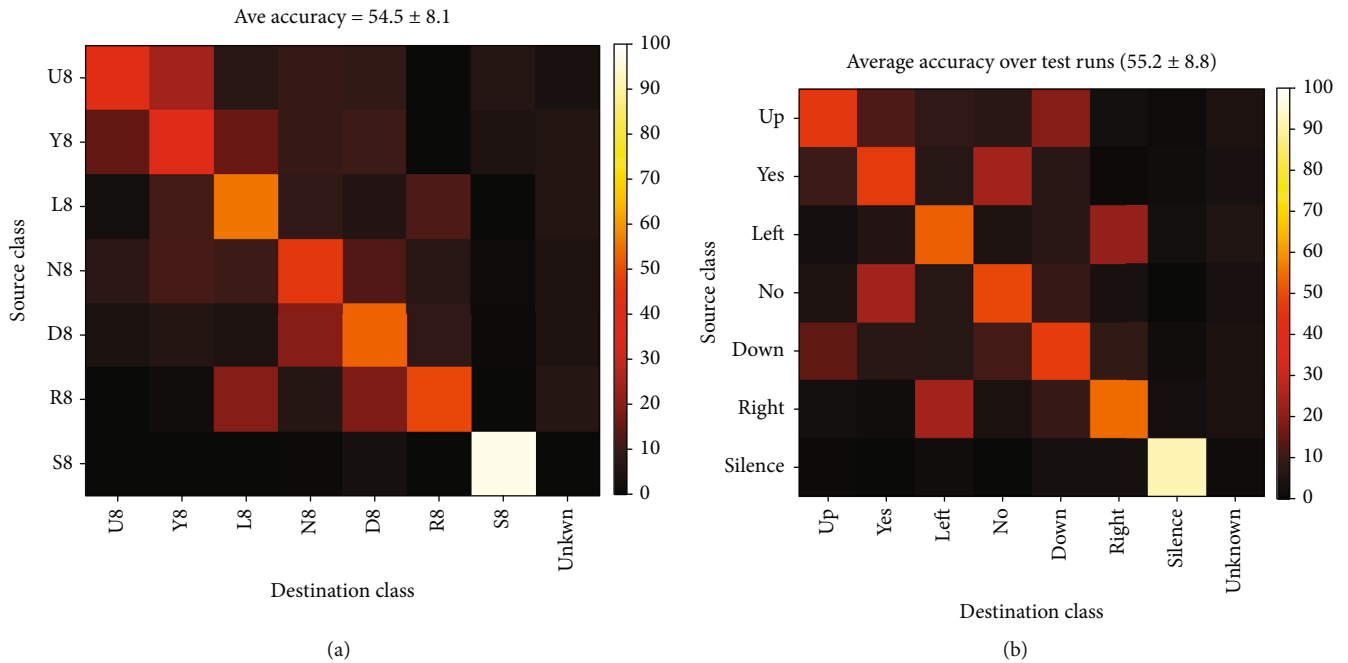


FIGURE 19: Confusion matrix of multiclass classifier of all words in a record: (a) record no. 8 and (b) average of all records.

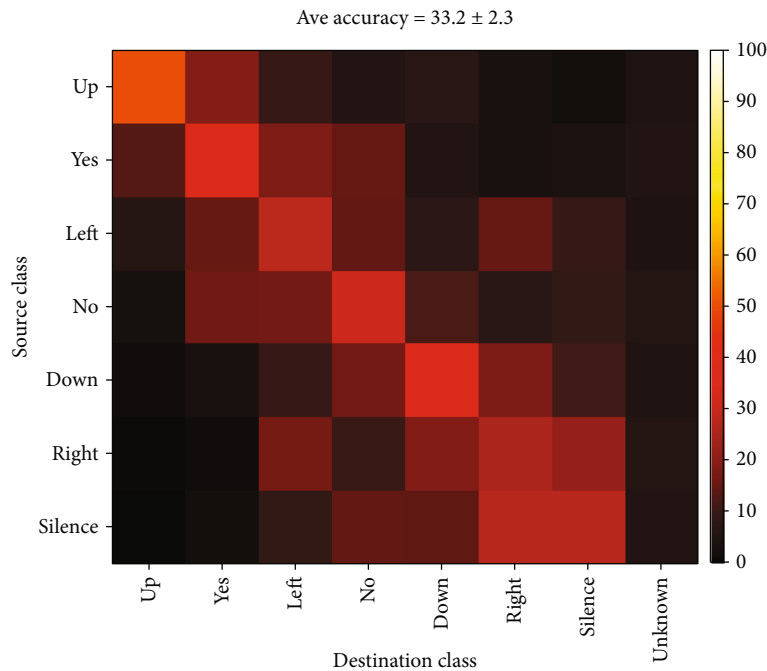


FIGURE 20: Confusion matrix for long-time classification of all 7 words.

hundreds of thousands of ways ($10!/3! = 604800$ per subject) to do this, the full estimation of classification accuracy by averaging all possible cases is impractical. Therefore, to calculate the confusion matrix shown in Figure 21, we used the average of 100 randomly selected permutations for each subject.

The overall accuracy was more than 88 percent, which is much better than both short-time and long-time modes.

Another interesting case of mixed-time multiclass classification was classifying the instances of a word in a record against the same word in other records. Figure 22(a) shows the confusion matrix of a classifier for the word “Up” in records no. 1 to 9 of subject no. 1. Note that how good the words in different records are distinguished. Figure 22(b) is a 14-class classifier of all 7 words in two different records. We can see that the two records have almost completely

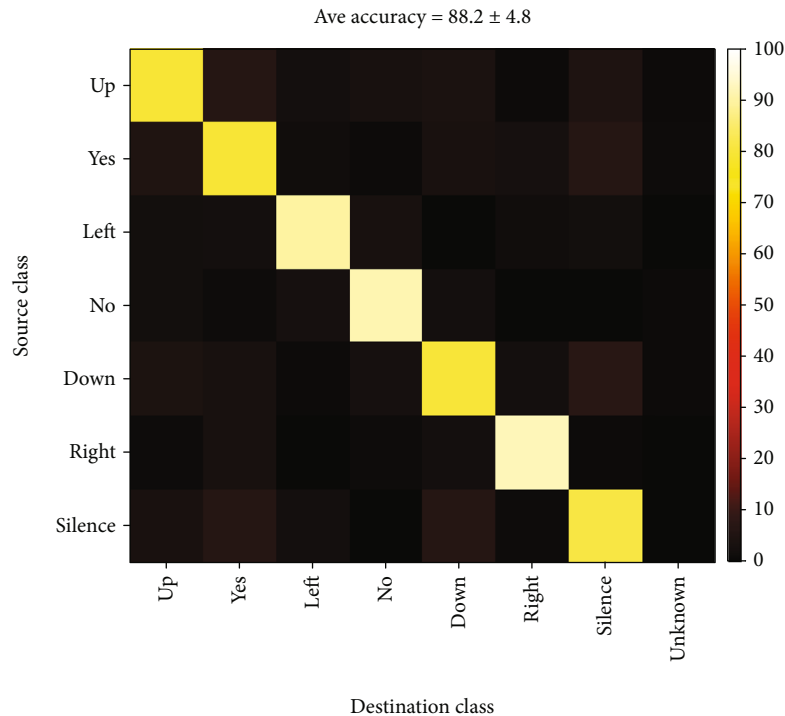


FIGURE 21: Confusion matrix for mixed-time classification of all 7 classes.

separated so that second and third quarters of the graph is almost empty. Consequently, in each part, words of a record are classified with a confusion pattern very similar to classifying the same record alone (compare patterns in Figure 19(a) with lower right section of Figure 22(b)).

4. Discussion

At first, it is worth noting that all presented algorithms and parameters, are the results selected after the evaluation of many (hundreds in some cases) possible substitutions. For example, many combinations of EEG-Channel collection and montages, tens of feature extraction methods and parameters, and a few classifier types and parameters are evaluated. Some had poor absolute results like Artificial Neural Networks (ANN) as classifier or raw EEG signal amplitudes as features. Some had weaker results than those reported here, e.g., Principle Component Analysis (PCA) features with a Minimum-Distance classifier [30], or using Banana montage for EEG signals [31]. Some did not fulfil our requirements, for example, Random Forest Classifier (RFC) needs to be separately configured and trained for each class combination. Also some combinations or parameters did not notably affect the performance and were left in their default values or set to an arbitrary value, like SVM machines basis function which was “Linear” by default (in MATLAB® version we used) or ordering of channels in EEG signal.

Table 6 shows a summary of average accuracies in all classification types and modes described in this report. As seen in each column of this table, the minimum accuracy for each type of classification accrues in long-time mode,

where the data of a relatively long-time interval were merged together. This is mainly due to the fact that EEG signals are not stationary and their statistics (or properties) change along the time. In long-time mode, samples from different times and with different properties are got together in a single class. The complexity and diversity of such large classes, causes the SVM machines to not precisely distinguish between them. This is evident in increased learning time of the machine as well as its learning accuracy. Hence, it can be concluded that we should not use (too) old EEG data in a silent-talk BCI application.

In the 2-class classification, for the short-time mode, relatively better accuracies were obtained, especially in the Word-Silent classification. This means that the pattern of brain waves, when it is busy with a word imagination (or perhaps any other), the task is much different from its background (idle) processes. So, if the subject cannot concentrate on his/her task, then the system performance will degrade in this (and generally all) type of classification. In the Word-Word classification, the large variance (20%) is due to the high correlation of EEG signals in too close time intervals, at least with some different but near tasks. This is a major shortcoming of the most imagined speech BCIs.

Finally, in the mixed-time mode, the best accuracies were obtained because in this mode we did not use both too far and too near signals.

Since the chance level was 50% in the 2-class, 33.3% in 3-class, and 14.3% in 7-class (balanced) classification, all accuracies were meaningfully above random classification (p value < 0.05), even in the worst case of the long-time mode.

In other modes, the accuracies, if not better, were comparable with the best reported values. For example, ALSaleh

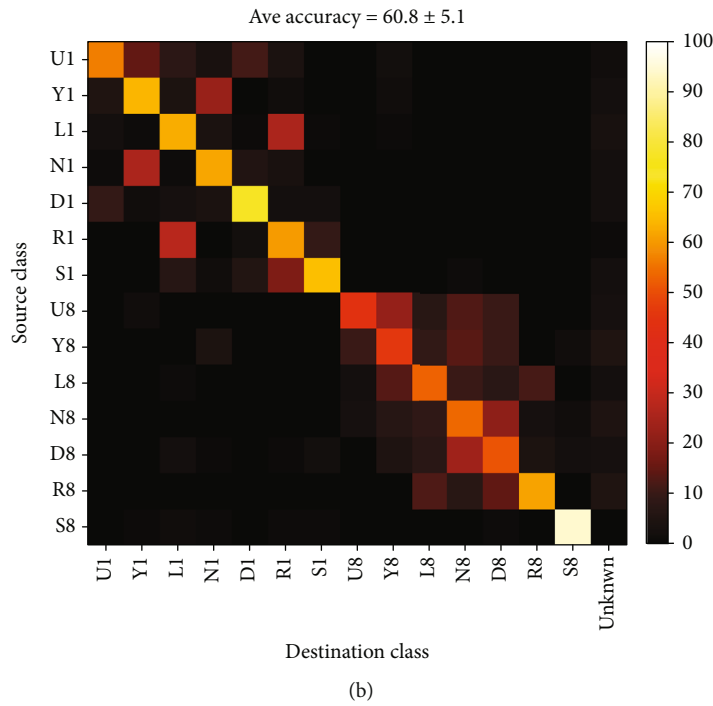
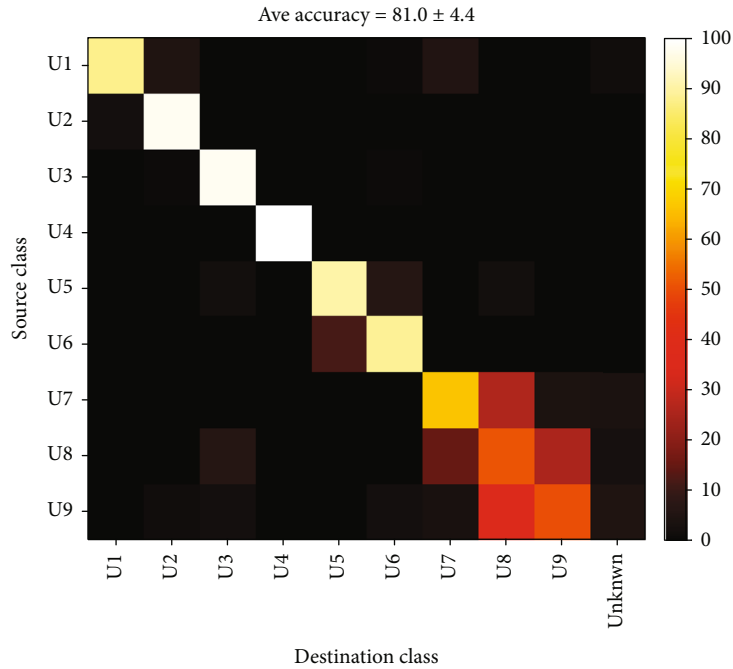


FIGURE 22: Confusion matrix of mixed-time multiclass classifications: (a) a single word in different records and (b) all words in 2 records.

TABLE 6: Summary of average accuracies in all classification types and modes.

Classification mode	Classification accuracy (mean \pm standard deviation)			
	Word-Silence	Word-Word	Word-Word-Silence	6 Words+Silence
Short-time	95.7 ± 8.3	74.5 ± 20.4	78.9 ± 13.9	55.2 ± 8.8
Long-time	57.5 ± 7.2	59.5 ± 7.5	39.5 ± 5.4	32.0 ± 2.1
Mixed-time	97.1 ± 5.1	96.4 ± 8.0	92.0 ± 11.1	88.2 ± 4.8

et al. reported an average maximum accuracy equal to 87.4% for 5 words classification over 9 subjects [32]. According to the reported data accusation schema, their classification method belonged to Mixed-time mode defined in this paper. Therefore, this value should be compared with 92 or 88 percent accuracies of 3 and 7-class types in Table 6.

DaSalla et al. had maximum accuracy about 78% for two class (vowel-silence) with three subjects [11]. In a similar manner, Brigham and Kumar achieved 68.8% for two syllables and seven subjects [16], compared to 74.5% for two words classification in our work.

Cooney et al. have published the results of their work on binary classification of word pairs [33]. Two different convolutional neural networks (CNN) and a baseline linear discriminant analysis (LDA) classifier were examined. 15 word pairs were constructed from six Spanish words of the EEG data set provided by Coretto et al. [34]. Despite of much smaller acquisition time, the data set can be considered as our long-time mode. Accuracies were 62.37% and 60.88% for deep and shallow CNNs and 57.80% for LDA classifier compared with 59.5% in our work. Also D-Y. Lee et al. reported 45% accuracy for the classification of the six words in Coretto's data set [22].

In another recent work using deep learning approaches, Panachakel et al. reported an average classification accuracy of 57.15% on 11 prompts of the KaraOne data set [35]. The prompts include seven phonemic/syllabic prompts (iy, uw, piy, tiy, diy, m, n) and four words (i.e., pat, pot, knew, and gnaw). Each prompt was presented 12 times for a total of 132 trials [36]. The recording time span was 30 to 40 minutes for each subject and therefore the classification is more similar to our short-time multiclass mode with 55.2% accuracy.

5. Conclusion

In this paper, we reported the results of a simple and extensible multi-class classifier working on self-recorded EEG signals during silent speech. Six complete and useful Persian words were the main classes, with the rest or silence for detecting talking/not-talking states. Although it seems that the language by itself may not change the way of human brain processing during speech, exploring Persian words in silent talk is new to literature in this domain of research. Also, the use of frequency spectrum as a feature and SVMs as the classifier have been reported in some papers, but the combination of this feature set (normalized to a constant length regardless of signal durations and sampling rates) and the purposed classifier has not been seen in previous studies.

Introducing classification modes, according to the time difference of EEG signals, is a new idea for discussing different accuracy values obtained in this and other related scholarly explorations. We made an attempt to expand this concept drawing on many other papers, which worked on imagined speech classification.

Data Availability

EEG data can be accessed by direct request from the corresponding author via mohammadkhani@irost.ir.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Vahid Reza Nafisi worked on the investigation. Fatemeh Darakeh wrote the methodology. Gholam Reza Mohammad Khani was the project administrator. Mohammad Reza Asghari Bejestani took care of the software used in the study. Gholam Reza Mohammad Khani supervised the work. Vahid Reza Nafisi validated the research. Mohammad Reza Asghari Bejestani was responsible for the writing of the original draft. Gholam Reza Mohammad Khani, Vahid Reza Nafisi, and Fatemeh Darakeh contributed to the writing, review, and editing of the manuscript.











References

- [1] J. J. Daly and J. E. Huggins, "Brain-computer interface: current and emerging rehabilitation applications," *Archives of Physical Medicine and Rehabilitation*, vol. 96, no. 3, pp. S1-S7, 2015.
- [2] C. DaSalla, H. Kambara, Y. Koike, and M. Sato, "Spatial filtering and single-trial classification of EEG during vowel speech imagery," in *Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology*, Singapore, Singapore, 2009.
- [3] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. Abd El-Samie, "EEG seizure detection and prediction algorithms: a survey," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, Article ID 183, 2014.
- [4] A. M. Abdulghani, A. J. Casson, and E. Rodriguez-Villegas, "Compressive sensing scalp EEG signals: implementations and practical performance," *Medical & Biological Engineering & Computing*, vol. 50, no. 11, pp. 1137-1145, 2011.
- [5] P. Fergus, D. Hignett, A. Hussain, D. Al-Jumeily, and K. Abdel-Aziz, "Automatic epileptic seizure detection using scalp EEG and advanced artificial intelligence techniques," *BioMed Research International*, vol. 2015, Article ID 986736, 17 pages, 2015.
- [6] L. Mondada, M. Karim, and F. Mondada, "Electroencephalography as implicit communication channel for proximal interaction between humans and robot swarms," *Swarm Intelligence*, vol. 10, no. 4, pp. 247-265, 2016.
- [7] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123-1134, 2001.
- [8] B. Rebsamen, E. Burdet, C. Guan et al., "Controlling a wheelchair indoors using thought," *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 18-24, 2007.
- [9] K. Tanaka, K. Matsunaga, and H. Wang, "Electroencephalogram based control of an electric wheelchair," *IEEE Transactions on Robotics*, vol. 21, no. 4, pp. 762-766, 2005.
- [10] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 4, pp. 317-326, 2008.
- [11] C. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial

- patterns,” *Neural Networks*, vol. 22, no. 9, pp. 1334–1339, 2009.
- [12] B. Min, J. Kim, H. Park, and B. Lee, “Vowel imagery decoding toward silent speech BCI using extreme learning machine with electroencephalogram,” *BioMed Research International*, vol. 2016, no. 1, Article ID 2618265, pp. 1–11, 2016.
- [13] B. Idrees and O. Farooq, “Vowel classification using wavelet decomposition during speech imagery,” in *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India, 2016.
- [14] S. Deng, R. Srinivasan, T. Lappas, and M. D’Zmura, “EEG classification of imagined syllable rhythm using Hilbert spectrum methods,” *Journal of Neural Engineering*, vol. 7, no. 4, article 046006, 2010.
- [15] J. Kim, S. Lee, and B. Lee, “EEG classification in a single-trial basis for vowel speech perception using multivariate empirical mode decomposition,” *Journal of Neural Engineering*, vol. 11, no. 3, article 036010, 2014.
- [16] K. Brigham and B. Kumar, “Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy,” in *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, Chengdu, China, 2010.
- [17] M. D’Zmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan, “Toward EEG Sensing of Imagined Speech,” in *Human-Computer Interaction. New Trends. HCI 2009*, J. A. Jacko, Ed., vol. 5610 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2009.
- [18] L. Wang, X. Zhang, X. Zhong, and Y. Zhang, “Analysis and classification of speech imagery eeg for bci,” *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 901–908, 2013.
- [19] E. González-Castañeda, A. Torres-García, C. Reyes-García, and L. Villaseñor-Pineda, “Sonification and textification: proposing methods for classifying unspoken words from EEG signals,” *Biomedical Signal Processing and Control*, vol. 37, pp. 82–91, 2017.
- [20] K. Mohanchandra and S. Saha, “A communication paradigm using subvocalized speech: translating brain signals into speech,” *Augmented Human Research*, vol. 1, no. 1, p. 3, 2016.
- [21] P. Saha, M. Abdul-Mageed, and S. Fels, “SPEAK YOUR MIND! Towards Imagined Speech Recognition With Hierarchical Deep Learning,” in *Interspeech 2019, 20th Annual Conference of the International Speech Communication Association*, Graz, Austria, 2019.
- [22] D.-Y. Lee, M. Lee, and S.-W. Lee, “Decoding imagined speech based on deep metric learning for intuitive BCI communication,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1363–1374, 2021.
- [23] P. Suppes, Z.-L. Lu, and B. Han, “Brain wave recognition of words,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 26, pp. 14965–14969, 1997.
- [24] A. Porbadnigk, M. Wester, J.-P. Calliess, and T. Schultz, “EEG-based Speech Recognition - Impact of Temporal Effects,” in *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, Porto, Portugal, 2009.
- [25] N. Yoshimura, A. Satsuma, C. S. DaSalla, T. Hanakawa, M.-A. Sato, and Y. Koike, “Usability of EEG cortical currents in classification of vowel speech imagery,” in *2011 International Conference on Virtual Rehabilitation*, Zurich, Switzerland, 2011.
- [26] F. Nijboer, E. Sellers, J. Mellinger et al., “A P300-based brain-computer interface for people with amyotrophic lateral sclerosis,” *Clinical Neurophysiology*, vol. 119, no. 8, pp. 1909–1916, 2008.
- [27] E. Donchin and Y. Arbel, “P300 based brain computer interfaces: a progress report,” in *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience. FAC 2009*, D. D. Schmorow, I. V. Estabrooke, and M. Grootjen, Eds., vol. 5638 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2009.
- [28] S. R. Sinha, “EEG instrumentation,” 2021, <https://neupsykey.com/eeg-instrumentation/>.
- [29] M. R. Asghari Bejestani, G. R. Mohammadkhani, and V. Nafisi, “Multi-class classification of Persian imagines words,” *The CSI Journal on Computing Science and Information Technology*, vol. 18, no. 2, pp. 94–104, 2021.
- [30] M. R. Asghari Bejestani, G. R. Mohammadkhani, S. Gorgin, V. R. Nafisi, and G. R. Farahani, “Classification of EEG signals for discrimination of two imagined words,” *Journal of Signal and Data Processing*, vol. 17, no. 2, p. 120, 2020.
- [31] T. Yamada and E. Meng, *Practical Guide for Clinical Neurophysiologic Testing: EEG*, Lippincott Williams & Wilkins (LWW), Philadelphia, PA, 2nd edition, 2017.
- [32] M. AlSaleh, R. Moore, H. Christensen, and M. Arvaneh, “Examining Temporal Variations in Recognizing Unspoken Words using EEG Signals,” in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Miyazaki, Japan, 2019.
- [33] C. Cooney, F. Raffaella, and D. Coyle, “Classification of Imagined Spoken Word-Pairs Using Convolutional Neural Networks,” in *Proceedings of the 8th Graz Brain Computer Interface Conference 2019: Bridging Science and Application*, Graz, 2019.
- [34] G. A. P. Coretto, I. Gareis, and H. L. Rufiner, “Open access database of EEG signals recorded during imagined speech,” in *12th International Symposium on Medical Information Processing and Analysis*, Tandil, Argentina, 2017.
- [35] J. T. Panachakel, A. Ramakrishnan, and T. Ananthapadmanabha, “Decoding Imagined Speech using Wavelet Features and Deep Neural Networks,” in *2019 IEEE 16th India Council International Conference (INDICON)*, Rajkot, India, 2019.
- [36] S. Zhao and F. Rudzicz, “Classifying phonological categories in imagined and articulated speech,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Australia, 2015.

Research Article

A U-Net Approach to Apical Lesion Segmentation on Panoramic Radiographs

Ibrahim S. Bayrakdar ¹, **Kaan Orhan** ^{2,3}, **Özer Çelik** ⁴, **Elif Bilgir** ¹, **Hande Sağlam** ¹,
Fatma Akkoca Kaplan ¹, **Sinem Atay Görür** ¹, **Alper Odabaş** ⁴, **Ahmet Faruk Aslan** ⁴,
and Ingrid Różyło-Kalinowska ⁵

¹Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Eskisehir Osmangazi University, Eskisehir 26040, Turkey

²Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Ankara University, Ankara 06560, Turkey

³Ankara University Medical Design Application and Research Center (MEDITAM), Ankara 06560, Turkey

⁴Department of Mathematics and Computer Science, Faculty of Science, Eskisehir Osmangazi University, Eskisehir 26040, Turkey

⁵Department of Dental and Maxillofacial Radiodiagnosics, Medical University of Lublin, Lublin 20-093, Poland

Correspondence should be addressed to Ingrid Różyło-Kalinowska; rozylo.kalinowska@umlub.pl

Received 1 July 2021; Revised 13 December 2021; Accepted 16 December 2021; Published 15 January 2022

Academic Editor: Lei Zhang

Copyright © 2022 Ibrahim S. Bayrakdar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of the paper was the assessment of the success of an artificial intelligence (AI) algorithm formed on a deep-convolutional neural network (D-CNN) model for the segmentation of apical lesions on dental panoramic radiographs. A total of 470 anonymized panoramic radiographs were used to progress the D-CNN AI model based on the U-Net algorithm (CranoCatch, Eskisehir, Turkey) for the segmentation of apical lesions. The radiographs were obtained from the Radiology Archive of the Department of Oral and Maxillofacial Radiology of the Faculty of Dentistry of Eskisehir Osmangazi University. A U-Net implemented with PyTorch model (version 1.4.0) was used for the segmentation of apical lesions. In the test data set, the AI model segmented 63 periapical lesions on 47 panoramic radiographs. The sensitivity, precision, and F1-score for segmentation of periapical lesions at 70% IoU values were 0.92, 0.84, and 0.88, respectively. AI systems have the potential to overcome clinical problems. AI may facilitate the assessment of periapical pathology based on panoramic radiographs.

1. Introduction

Chronic apical periodontitis is an infection of tissues surrounding the dental apex induced by pulpal disease, mostly because of bacterial disease in the root canal complex developing during untreated or incorrectly treated dental caries [1–3]. Apical periodontitis is common, and its prevalence increases with age. Epidemiological studies have reported that apical periodontitis is present in 7% of teeth and 70% of the general population. The diagnosis of acute apical periodontitis is made clinically, but the detection of chronic apical periodontitis is done by radiography [4]. In general, following root canal treatment, complete healing of periapical lesions is expected or at least improvement in the form of

a decrease of the size of periapical lesion [1, 5]. Radiographically, apical periodontitis manifests as a widened periodontal ligament space or visible lesions. Such radiolucencies, also called apical lesions, tend to be detected incidentally or by radiographic follow-up of endodontically treated teeth [6, 7]. Radiolucency in radiographs is an important feature of apical periodontitis [2]. Apical periodontitis can be detected on periapical and panoramic radiographs and by cone-beam computed tomography (CBCT). CBCT has superior discriminatory power but is costly and exposes the patient to radiation burden [6, 8]. Periapical and panoramic radiographs are the most frequently used techniques in the diagnosis and treatment of apical lesions [2]. Panoramic radiography generates two-dimensional (2D) tomographic

images of the entire maxillomandibular area [9], enabling the evaluation of all teeth simultaneously. Also, panoramic radiography requires a far lower dose of radiation than CBCT imaging [6, 10]. Besides, panoramic radiography is painless, unlike intraoral radiographs, thus well tolerated by patients [9, 11]. One of the many recent technological advances in artificial intelligence (AI) and its applications are expanding rapidly, also in the area of medical management and medical imaging [12]. AI uses computational networks (neural networks (NNs)) that mimic biological nervous systems [13]. NNs were developed as one of the first types of AI algorithms. The computing power of NNs varies depending on the character and amount of training data. Networks using many large layers are termed deep learning NNs [14]. A deep convolutional neural network (D-CNN) was used to process large and complex images [15]. Deep learning networks, including CNNs, have displayed superior achievement in terms of object, face, and activity recognition [16]. Medical organ and lesion segmentation are an important application of imaging modalities [17, 18]. The detection and classification performance of deep learning-based CNNs concerning retinopathy caused by diabetes, skin cancer, and tuberculosis is very high [19, 20]. CNNs have also been applied in dentistry for tooth detection and numbering, as well as an assessment of periodontal bone loss and periapical pathology [21–25]. U-Net and pixel-based image segmentation, which is a different architecture created from CNN layers, are more successful than classical models even if there are few training images. The presentation of this architecture has been realized with biomedical images. The traditional U-Net architecture, extended to handle volumetric input, has two phases: the coder portion of the network where it learns representational features at unlikely scale and gather-dependent information, and the decoder portion where the network extracts knowledge from the noticed situation and formerly learned features. The jump links used between the corresponding encoder and decoder layers allow deep parts of the network to be trained efficiently and compare the same receiver characteristics with different receiver areas [26].

The study is aimed at assessing the diagnostic success of U-Net approach for the segmentation of apical lesions in panoramic images.

2. Material and Methods

2.1. Radiographic Data Preparation. The panoramic radiographs used in the study were derived from the archives of the Faculty of Dentistry of Eskisehir Osmangazi University; 470 anonymized panoramic radiographs were applied. The radiographs were obtained from January 2018 to January 2019 for a variety of reasons. Images with artifacts of any type were excluded. The study design was authorized by the Non-Interventional Clinical Research Ethics Committee of Eskisehir Osmangazi University (decision date and number: 06.08.2019/14). The study was conducted following the regulations of the Declaration of Helsinki. The Planmeca Promax 2D (Planmeca, Helsinki, Finland) panoramic imag-

ing system was used to obtain panoramic radiographs with the following parameters: 68 kVp, 16 mA, and 13 s.

2.2. Image Annotation. Three dental radiologists (I.S.B. and E. B. with 10 years of experience and F.A.K. with 3 years of experience) annotated ground truth images with the common decision on all images using CranioCatch Annotation software (Eskisehir, Turkey). The polygonal boxes were used to determine the locations of the apical lesions.

2.3. Deep CNN Architecture. The deep learning was performed using a U-Net implemented with the PyTorch model (version 1.4.0). The U-Net architecture is used for semantic segmentation assignments (Figure 1).

The U-Net architecture consists of four block levels, including two convolutional layers with batch normalization and a rectified linear unit activation function (ReLU). There is a maximum pool layer in the encoding section and upconvolution layers in the decoding section. Each block has 32, 64, 128, or 256 convolutional filters. Besides the bottleneck, the layer comprises 512 convolutional filters. Skip connections to the corresponding layers from the encoding layers are present in the decoding part [26]. The Adam Optimizer was used to train the U-Net.

2.4. Model Pipeline. PyTorch library was used for model development on the Python open-source programming language (v. 3.6.1; Python Software Foundation, Wilmington, DE, USA; retrieved on August 1, 2019, from <https://www.python.org/>). An AI model (CranioCatch, Eskisehir-Turkey) was developed to automatically segment apical lesions on panoramic radiographs. The training process was performed using an individual computer implemented with 16 GB RAM and an NVIDIA GeForce GTX 1060Ti graphic card.

- (i) Split: 470 panoramic radiographs were divided into train, validation, and test group
 - (a) Training group: 380
 - (b) Validation group: 43
 - (c) Test group: 47
- (ii) Augmentation: 1140 images from the 380 original training group images were derived using data augmentation. Augmentation was applied on the training data set, and augmentations were horizontal flip and vertical flip (total images: 1140 (=380 × 3)) (size: 2943 × 1435)
- (iii) Cropping (preprocessing step): then, all images of the train were divided into 4 parts as upper right, upper left, lower right, and lower left (size: 1000 × 530)
 - (a) Training group: 1140 × 4 = 4560

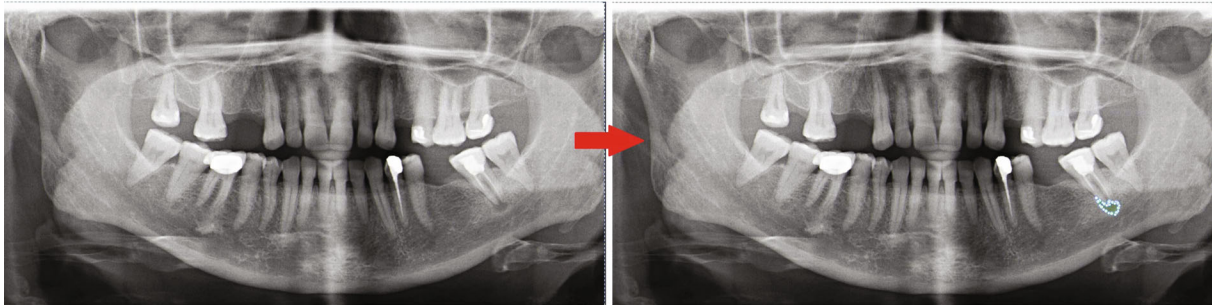


FIGURE 1: Annotation of the apical lesion using polygonal box method.

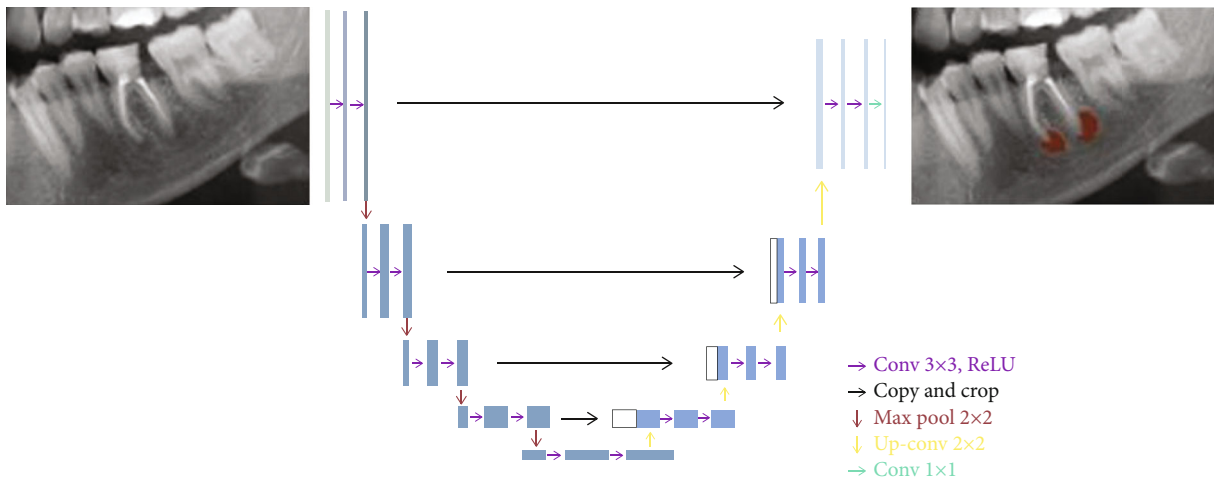


FIGURE 2: The U-Net architecture for the semantic segmentation task.

(b) Validation group: $43 \times 4 = 172$

(c) Test group: 47

(c) Test group: 47

(iv) Remove full black masks (preprocessing step): the regions without lesions of all data set were deleted

(vi) Resize (preprocessing step): the resolution of each piece divided into 4 (1000×530) was resized to 512×256

(a) Training group: 1629

(a) Training group: 1629

(b) Validation group: 59

(b) Validation group: 59

(c) Test group: 47

(c) Test group: 47

(v) Contrast Limited Adaptive Histogram Equalization (CLAHE) (preprocessing step): CLAHE has applied all images to improve image contrast and enable the identification of apical lesions

The segmentation model with PyTorch U-Net was trained with 95 epochs; the model based on 43 epochs showed the best performance and was thus used in the experiment. The model pipeline is summarized in Figure 2.

(a) Training group: 1629

2.5. *Statistical Analysis.* The confusion matrix was used to assess the achievement of the model. This matrix is a meaningful table that summarizes the predicted and actual situations. The performance of model is frequently assessed using the data in the confusion matrix [27]. The metrics used to evaluate the success of the model were as follows:

(b) Validation group: 59

- (1) True Positive (TP): apical lesion was segmented, correctly
- (2) False Positive (FP): apical lesions were not detected
- (3) False Negative (FN): without apical lesions, lesions were nevertheless segmented
- (4) TP, FP, and FN were determined; then, the following metrics were computed:

- (i) Sensitivity (recall): $TP/(TP + FN)$
- (ii) Precision: $TP/(TP + FP)$
- (iii) F1 score: $2TP/(2TP + FP + FN)$

3. Results and Discussion

3.1. Results. The AI model segmented 63 apical lesions on 47 radiographs in the test data set (True Positives) (Figures 3–5).

Twelve apical lesions were not detected (False Negatives). In 5 cases without apical lesions, lesions were nevertheless segmented by the AI model (False Positives) (Table 1).

The sensitivity, precision, and F1-score values at 70% IoU value were 0.92, 0.84, and 0.88, respectively (Table 2).

3.2. Discussion. AI has rapidly improved the interpretation of medical and dental images, including via the application of deep learning models and CNNs [28, 29]. Deep learning has been developing rapidly thus recently attracting considerable attention [28–34]. The deep CNN architecture appears to be the most used deep learning approach. This is most likely due to its effective self-learning models and high computing capacity, which provide superior classification, detection, and quantitative performance based on imaging data [28–35]. CNNs have been used in dentistry for cephalometric landmark detection, dental structure segmentation, tooth classification, and apical lesion detection [36–39].

Tuzoff et al. presented a novel CNN algorithm for automatic tooth detection and numbering on panoramic radiographs. They found the sensitivity and specificity value of tooth numbering as 0.9893 and 0.9997, respectively. The findings showed the ability of current CNN architectures for automatic dental radiographic interpretation and diagnosis on panoramic radiographs [25]. Chen et al. detected and numbered teeth in dental periapical films using faster region proposal CNN networks (faster R-CNN). Faster R-CNN performed unusually well for tooth detection and localization, showing good precision and recall and overall performance like that of a younger dentist [24]. Miki et al. assessed the utility of deep CNN for classifying teeth based on dental CBCT images; the accuracy was 91.0%. The system rapidly and automatically produces diagrams for forensic recognition [38]. Two previous studies investigated the utility of AI systems for detecting periapical lesions. Ekert et al. investigated the capability of deep CNN algorithm to detect apical lesions on dental panoramic radiographs. CNNs

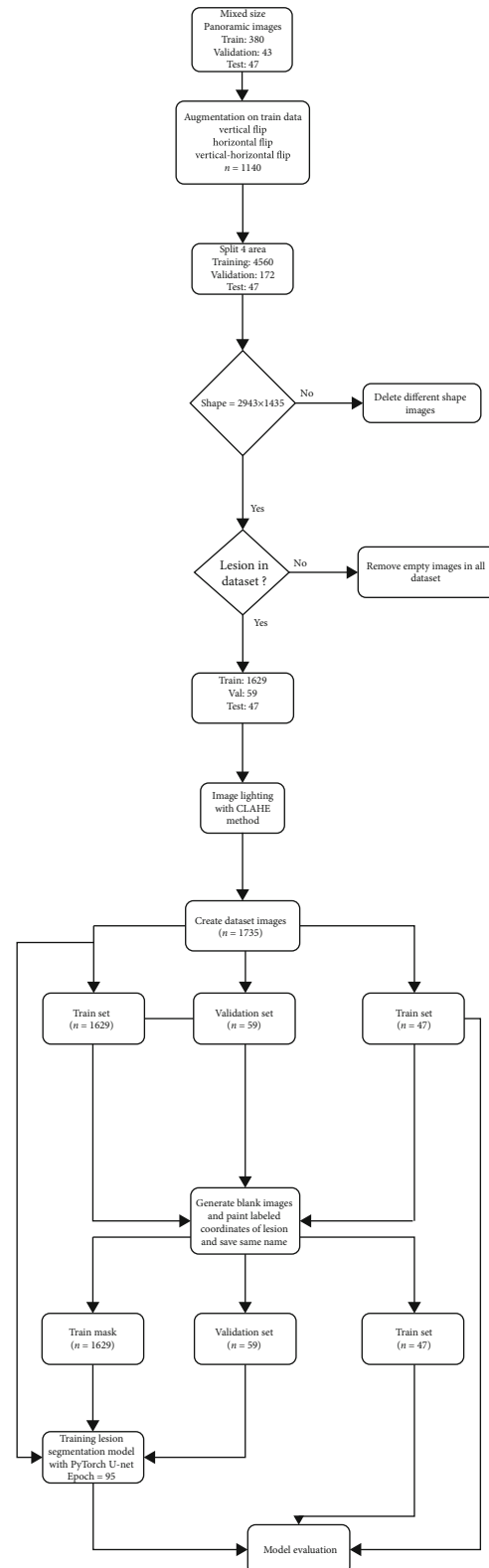


FIGURE 3: Model pipeline for apical lesion segmentation (CraneoCatch, Eskisehir, Turkey).

detected the lesions despite the small number of data sets [6]. Orhan et al. [39] compared the diagnostic ability of a deep CNN algorithm to that of volume measurements based

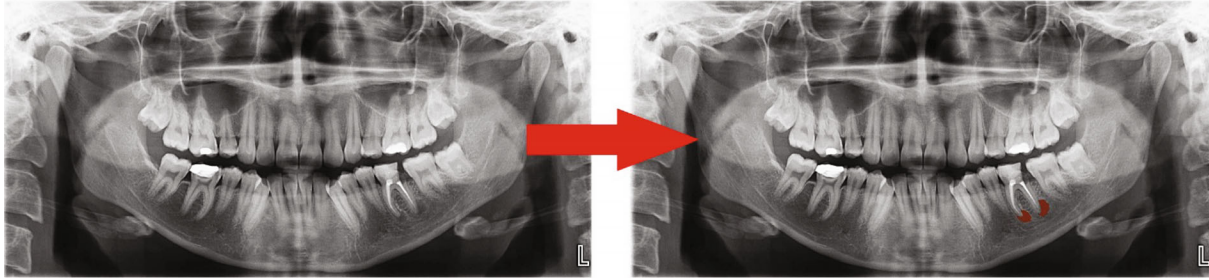


FIGURE 4: Automatically apical lesion segmentation using AI model (CranioCatch, Eskisehir, Turkey).

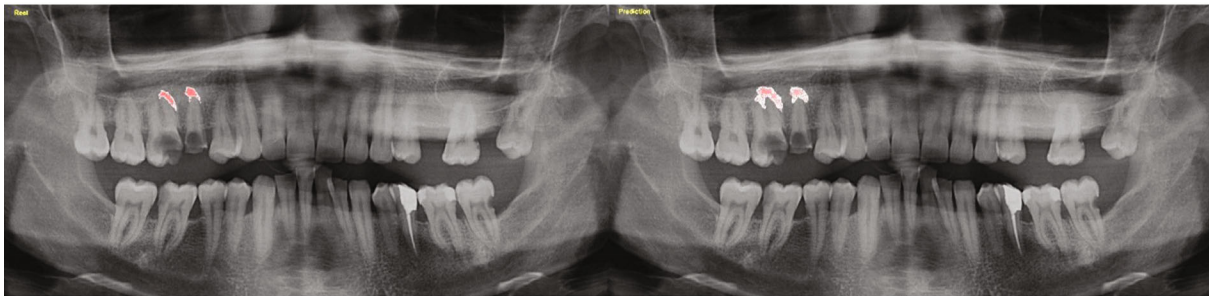


FIGURE 5: An example real-prediction image comparison.

TABLE 1: The number of segmented apical lesions with AI model (CranioCatch, Eskisehir, Turkey).

Metrics	Number
True Positives (TP)	63
False Negatives (FN)	12
False Positives (FP)	5

TABLE 2: The prediction performance measurement of the AI model (CranioCatch, Eskisehir, Turkey).

Measure	Value	Derivations
Sensitivity (recall)	0.92	$TP/(TP + FN)$
Precision	0.84	$TP/(TP + FP)$
F1 score	0.88	$2TP/(2TP + FP + FN)$
IoU value	0.79	$TP/(TP + FP + FN)$
Dice coefficient	0.88	$2TP/(2TP + FP + FN)$

on CBCT images in the context of periapical pathology. The rate of detection of periapical lesions of the CNN model was 92.8%, and the volumetric and manual segmentation measurements were similar [39]. Endres et al. [40] created a model using 2902 deidentified panoramic radiographs. The presence of periapical radiolucencies on panoramic radiographs was evaluated by 24 oral and maxillofacial surgeons. They show that the deep learning algorithm has better success than 14 of 24 oral and maxillofacial surgeons. The success metrics for this model were as follows: the precision of 0.60 and an F1 score of 0.58 corresponding to a positive predictive value of 0.67 and True Positive rate of 0.51. Setzer

et al. performed a study to use a deep learning proposal using U-Net architecture for the automatic segmentation of periapical lesions on CBCT images [41]. Segmentation of lesion accuracy was found as 0.93 with a specificity of 0.88, a positive predictive value of 0.87, and a negative predictive value of 0.93. They concluded that the DL algorithm trained in a limited CBCT images presented wonderful results in lesion detection accuracy. In the presented study, we created a segmentation model with PyTorch U-Net AI architecture on panoramic radiograph. It segmented 63 apical lesions on 47 radiographs in the test data set. Twelve apical lesions were not detected. In 5 cases without apical lesions, lesions were nevertheless segmented by the AI model. The sensitivity, precision, and F1-score values at 70% IoU value were 0.92, 0.84, and 0.88, respectively. Our results showed that AI deep learning algorithms can have service ability in the clinical dental setting. However, the present study had some limitations. Only one radiography machine and standard parameters were used to image acquisitions. Besides, study groups included all size of periapical images. The external test group was not used to assess the model's success. We used the U-Net algorithm to model development, only. Future studies should be used using larger study samples and images taken from different radiography equipment. Comparative experiments should be planned to use different CNN algorithms, and AI model performance should be compared to different human observers which have different level professional experiences.

4. Conclusions

Deep learning AI models enable the evaluation of periapical pathology based on panoramic radiographs. The application

of AI for apical lesion detection and segmentation can reduce the burden on clinicians.

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.

References

- [1] J. Segura-Egea, J. Martín-González, and L. Castellanos-Cosano, "Endodontic medicine: connections between apical periodontitis and systemic diseases," *International Endodontic Journal*, vol. 48, no. 10, pp. 933–951, 2015.
- [2] P. Velvart, H. Hecker, and G. Tillinger, "Detection of the apical lesion and the mandibular canal in conventional radiography and computed tomography," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, vol. 92, no. 6, pp. 682–688, 2001.
- [3] C. Ridaio-Sacie, J. Segura-Egea, A. Fernández-Palacín, P. Bullón-Fernández, and J. Ríos-Santos, "Radiological assessment of periapical status using the periapical index: comparison of periapical radiography and digital panoramic radiography," *International Endodontic Journal*, vol. 40, no. 6, pp. 433–440, 2007.
- [4] S. Patel and C. Durack, "Radiology of apical periodontitis. Essential endodontology: prevention and treatment of apical periodontitis," in *Essential Endodontology: Prevention and Treatment of Apical Periodontitis*, pp. 179–210, Blackwell, 2019.
- [5] T. Connert, M. Truckenmüller, A. ElAyouti et al., "Changes in periapical status, quality of root fillings and estimated endodontic treatment need in a similar urban German population 20 years later," *Clinical Oral Investigations*, vol. 23, no. 3, pp. 1373–1382, 2019.
- [6] T. Ekert, J. Krois, L. Meinhold et al., "Deep learning for the radiographic detection of apical lesions," *Journal of Endodontics*, vol. 45, no. 7, pp. 917–922.e5, 2019.
- [7] S. Kanagasigam, H. Hussaini, I. Soo, S. Baharin, A. Ashar, and S. Patel, "Accuracy of single and parallax film and digital periapical radiographs in diagnosing apical periodontitis—a cadaver study," *International Endodontic Journal*, vol. 50, no. 5, pp. 427–436, 2017.
- [8] S. Patel, J. Brown, M. Semper, F. Abella, and F. Mannocci, "European Society of Endodontology position statement: use of cone beam computed tomography in Endodontics," *International Endodontic Journal*, vol. 52, no. 12, pp. 1675–1678, 2019.
- [9] X. Du, Y. Chen, J. Zhao, and Y. Xi, "A convolutional neural network based auto-positioning method for dental arch in rotational panoramic radiography," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2615–2618, Honolulu, HI, USA, 2018.
- [10] K. Leonardi Dutra, L. Haas, A. L. Porporatti et al., "Diagnostic accuracy of cone-beam computed tomography and conventional radiography on apical periodontitis: a systematic review and meta-analysis," *Journal of Endodontics*, vol. 42, no. 3, pp. 356–364, 2016.
- [11] S. C. White and M. J. Pharoah, *Oral radiology-E-Book: Principles and interpretation*, Elsevier Health Sciences, St. Louis, 2014.
- [12] B. Allen, S. E. Seltzer, C. P. Langlotz et al., "A road map for translational research on artificial intelligence in medical imaging: from the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop," *Journal of the American College of Radiology*, vol. 16, no. 9, pp. 1179–1189, 2019.
- [13] S. Wong, H. al-Hasani, Z. Alam, and A. Alam, "Artificial intelligence in radiology: how will we be affected?," *European Radiology*, vol. 29, no. 1, pp. 141–143, 2019.
- [14] J. R. Burt, N. Torosdagli, N. Khosravan et al., "Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks," *The British Journal of Radiology*, vol. 91, no. 1089, 2018.
- [15] J. J. Hwang, Y. H. Jung, B. H. Cho, and M. S. Heo, "An overview of deep learning in the field of dentistry," *Imaging Science in dentistry*, vol. 49, no. 1, pp. 1–7, 2019.
- [16] J. E. Sklan, A. J. Plassard, D. Fabbri, and B. A. Landman, "Toward content-based image retrieval with deep convolutional neural networks," *Proceedings of SPIE—the International Society for Optical Engineering*, vol. 9417, p. 94172C, 2015.
- [17] E. Neri, N. de Souza, A. Brady et al., "What the radiologist should know about artificial intelligence—an ESR white paper," *Insights into imaging*, vol. 10, no. 1, 2019.
- [18] J. H. Lee, D. H. Kim, S. N. Jeong, and S. H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *Journal of Dentistry*, vol. 77, pp. 106–111, 2018.
- [19] V. Gulshan, L. Peng, M. Coram et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [20] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [21] J. Krois, T. Ekert, L. Meinhold et al., "Deep learning for the radiographic detection of periodontal bone loss," *Scientific Reports*, vol. 9, no. 1, pp. 1–6, 2019.
- [22] A. Davies, F. Mannocci, P. Mitchell, M. Andiappan, and S. Patel, "The detection of periapical pathoses in root filled teeth using single and parallax periapical radiographs versus cone beam computed tomography - a clinical study," *International Endodontic Journal*, vol. 48, no. 6, pp. 582–592, 2015.
- [23] A. Zakirov, M. Ezhov, M. Gusarev, V. Alexandrovsky, and E. Shumilov, "End-to-end dental pathology detection in 3D cone-beam computed tomography images," 2018, <http://arxiv.org/abs/1810.10309>.
- [24] H. Chen, K. Zhang, P. Lyu et al., "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Scientific Reports*, vol. 9, no. 1, pp. 1–11, 2019.
- [25] D. V. Tuzoff, L. N. Tuzova, M. M. Bornstein et al., "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," *Dentomaxillofacial Radiology*, vol. 48, no. 4, 2019.
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

- [27] U. Ö. Osmanoğlu, O. N. Atak, K. Çağlar, H. Kayhan, and T. C. Can, "Sentiment analysis for distance education course materials: a machine learning approach," *Journal of Educational Technology and Online Learning*, vol. 3, no. 1, pp. 31–48, 2020.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, Santiago, Chile, 2015.
- [29] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.
- [30] L. Zhang, D. Arefan, Y. Guo, and S. Wu, "Fully automated tumor localization and segmentation in breast DCEMRI using deep learning and kinetic prior," in *SPIE Medical Imaging*, p. 113180Z, Houston, Texas, USA, 2020.
- [31] L. Zhang, A. A. Mohamed, R. Chai, Y. Guo, B. Zheng, and S. Wu, "Automated deep learning method for whole-breast segmentation in diffusion-weighted breast MRI," *Journal of Magnetic Resonance Imaging*, vol. 51, no. 2, pp. 635–643, 2020.
- [32] L. Zhang, Z. Luo, R. Chai, D. Arefan, J. Sumkin, and S. Wu, "Deep-learning method for tumor segmentation in breast DCE-MRI," in *SPIE Medical Imaging*, p. 6, San Diego, California, USA, 2019.
- [33] O. Russakovsky, J. Deng, H. Su et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [34] H. Hricak, "2016 New Horizons Lecture: Beyond Imaging—Radiology of Tomorrow," *Radiology*, vol. 286, no. 3, pp. 764–775, 2018.
- [35] S. Soffer, A. Ben-Cohen, O. Shimon, M. M. Amitai, H. Greenspan, and E. Klang, "Convolutional neural networks for radiologic images: a radiologist's guide," *Radiology*, vol. 290, no. 3, pp. 590–606, 2019.
- [36] S. Ö. Arik, B. Ibragimov, and L. Xing, "Fully automated quantitative cephalometry using convolutional neural networks," *Journal of Medical Imaging*, vol. 4, no. 1, 2017.
- [37] C. W. Wang, C. T. Huang, J. H. Lee et al., "A benchmark for comparison of dental radiography analysis algorithms," *Medical Image Analysis*, vol. 31, pp. 63–76, 2016.
- [38] Y. Miki, C. Muramatsu, T. Hayashi et al., "Classification of teeth in cone-beam CT using deep convolutional neural network," *Computers in Biology and Medicine*, vol. 80, pp. 24–29, 2017.
- [39] K. Orhan, I. S. Bayrakdar, M. Ezhov, A. Kravtsov, and T. Ozyurek, "Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans," *International Endodontic Journal*, vol. 53, no. 5, pp. 680–689, 2020.
- [40] M. G. Endres, F. Hillen, M. Salloumis et al., "Development of a deep learning algorithm for periapical disease detection in dental radiographs," *Diagnostics*, vol. 10, no. 6, p. 430, 2020.
- [41] F. C. Setzer, K. J. Shi, Z. Zhang et al., "Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images," *Journal of Endodontia*, vol. 46, no. 7, pp. 987–993, 2020.

Research Article

An Expert System for COVID-19 Infection Tracking in Lungs Using Image Processing and Deep Learning Techniques

Umashankar Subramaniam ¹, M. Monica Subashini,² Dhafer Almakhlles,¹
Alagar Karthick ³, and S. Manoharan ⁴

¹Department of Communications and Networks, Prince Sultan University, 11586 Riyadh, Saudi Arabia

²Department of Control and Automation, School of Electrical Engineering, Vellore Institute of Technology, 632014 Vellore, India

³Renewable Energy Lab, Department of Electrical and Electronics Engineering, KPR Institute of Engineering and Technology, Arasur, Coimbatore 641407, Tamilnadu, India

⁴Department of Computer Science, School of Informatics and Electrical Engineering, Institute of Technology, Ambo University, Ambo Post Box No.: 19, Ethiopia

Correspondence should be addressed to Umashankar Subramaniam; usubramaniam@psu.edu.sa and S. Manoharan; manoharan.subramanian@ambou.edu.et

Received 5 May 2021; Revised 29 September 2021; Accepted 16 October 2021; Published 13 November 2021

Academic Editor: Lei Zhang

Copyright © 2021 Umashankar Subramaniam et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The proposed method introduces algorithms for the preprocessing of normal, COVID-19, and pneumonia X-ray lung images which promote the accuracy of classification when compared with raw (unprocessed) X-ray lung images. Preprocessing of an image improves the quality of an image increasing the intersection over union scores in segmentation of lungs from the X-ray images. The authors have implemented an efficient preprocessing and classification technique for respiratory disease detection. In this proposed method, the histogram of oriented gradients (HOG) algorithm, Haar transform (Haar), and local binary pattern (LBP) algorithm were applied on lung X-ray images to extract the best features and segment the left lung and right lung. The segmentation of lungs from the X-ray can improve the accuracy of results in COVID-19 detection algorithms or any machine/deep learning techniques. The segmented lungs are validated over intersection over union scores to compare the algorithms. The preprocessed X-ray image results in better accuracy in classification for all three classes (normal/COVID-19/pneumonia) than unprocessed raw images. VGGNet, AlexNet, Resnet, and the proposed deep neural network were implemented for the classification of respiratory diseases. Among these architectures, the proposed deep neural network outperformed the other models with better classification accuracy.

1. Introduction

Emerging pathogens are a big concern for global public health, and technology may help classify potential cases more rapidly in order to bring in timely treatments [1, 2]. The Envision 2030 agenda of the United Nations have included 17 sustainable development goals towards a promising future for persons with disabilities aligned with Saudi Vision 2030. As per the SD goals set and implemented by the United Nations, the proposed work targets promoting the transformation of disabilities. This work promises good health and well-being (SDG 3) by diagnosing respiratory

diseases at the earlier stages based on chest X-ray images. To achieve this, the authors have utilized innovation techniques, infrastructure (SDG 9), and international partnerships (SDG 17) to transform the world into a better place to live.

The novel coronavirus identified in December resulted in significant quarantines worldwide, including large cities, villages, and public areas [3]. The huge impact of COVID-19 is due to a lack of testing and medical errors. COVID-19 is a disease that mainly affects the lungs [4] apart from pneumonia. This can exhibit various patterns or pathological symptoms in the lungs based on different causes, and no particular symptom can indicate its severe impact on the lungs. Hence, the

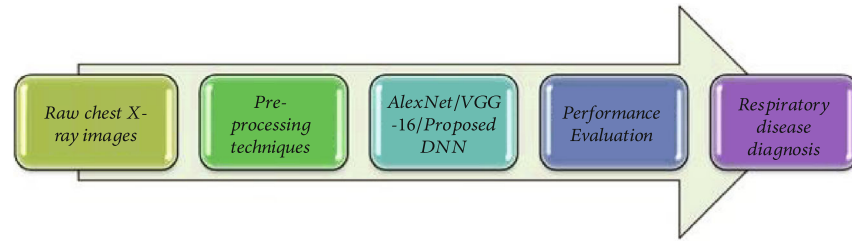


FIGURE 1: Process flow.

diagnosis should be initiated at an early stage. Based on the symptoms, when a person approaches a clinic, few tests are done for confirmation of the disease along with X-ray images. X-ray images help in analyzing the condition of the lung and the stage of the disease. Preprocessing of these images is absolutely necessary for accurate diagnosis.

2. Contributions and Limitations

This paper focuses on image preprocessing techniques applied on pneumonia, COVID-19 dataset (X-ray images), and diagnosis of diseases from the processed X-ray images

To diagnose the disease using X-ray images, different CNN architectures were implemented and compared to find out the best architecture

In this study, deep neural network architecture is proposed to diagnose normal/pneumonia/COVID-19 or bacterial infected disease

This model is trained on COVID-19 X-ray images, pneumonia X-ray images, and normal lung X-ray images

The failure to detect the early stages of COVID-19 is one of the most significant limitations of chest radiography research

But well-trained deep learning models can concentrate on points that the human eye cannot notice, and this method was successful in achieving the same

The only limitation to this method is the availability of datasets (training/testing) on normal/COVID-19/pneumonia X-ray images. The literature survey enabled the use of resources in fetching the dataset from various medical repositories

3. Related Work

COVID-19 has to be detected properly without any negligence else can lead to a severe impact on the country's economy and country's citizen health [5]. The person who is suspicious of COVID-19 is suggested to undergo a chest X-ray. Analysis of X-rays by humans can lead to various human errors, which can lead to a huge impact on patients and society. So, a computer-aided system can help the doctors for proper analysis of lungs of the COVID-19 affected human. Throughout underdeveloped and developing nations, where the number of patients is high and medical care cannot be adequately delivered, these programs may be a tremendous benefit [6, 7]. The authors have worked on X-ray imaging techniques for the detection of bone fractures. They have applied edge detection and segmentation

TABLE 1: Dataset for building a model.

Chest X-ray image	No. of images used for training the model	No. of images used for testing the model
Normal lung	110	30
COVID-19	110	30
Pneumonia	110	30

techniques to ease the process of the diagnosis system. These methods will reduce the processing time and other physical evaluation procedures [8]. So, while working with X-ray images, we need to consider the noises which have to be reduced. The random noises occurring during the process of image acquisition degrades the image quality leading to an incorrect diagnosis. Researchers suggest the application of the temporal recursive filter. Also, they propose an improved self-adaptive filter. This was a combination of FPGA with image processing techniques [9]. The authors have recommended region localization which offers a close level of precision. Few other image preprocessing techniques are adaptive histogram-based equalization, adaptive contrast enhancement, and histogram equalization. There is the presence of multiple noises during capturing the images because of device mobility and motion artefact [10]. But in X-ray images mostly Gaussian, salt and pepper noises are present. To reduce the noise, a digital median filtering technique is used as per the researches.

Chest X-rays aid in the diagnosis of pneumonia. Researchers seek the help of CNN in classifying normal and abnormal X-rays [11]. The feature extracted from the lung X-ray improves the functionality of the classifier. This method is useful where a large dataset is received. In another similar work, deep learning techniques are applied for the analysis of chest X-rays. Pulmonary infections are easily identified using these radiography images. This is extended in the detection of coronavirus disease [12]. The authors have brought hope in applying artificial intelligence in the early detection of the disease. Supervised learning techniques have been applied in the classification of a normal/abnormal pneumonia dataset. A labelled dataset aids this process in reducing the error [13]. CNN has been trained with nonaugmented data. The researchers have suggested a novel deep learning architecture for pneumonia detection [14]. They have applied transfer learning for efficient classification. The features were extracted and pretrained on ImageNet.

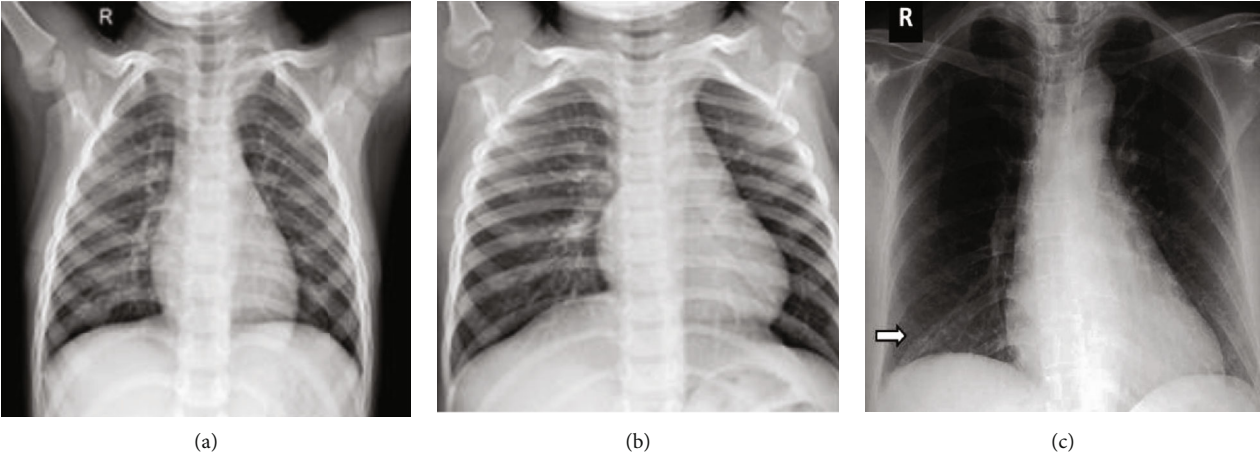


FIGURE 2: Sample image from training dataset: (a) normal; (b) pneumonia; (c) COVID-19.

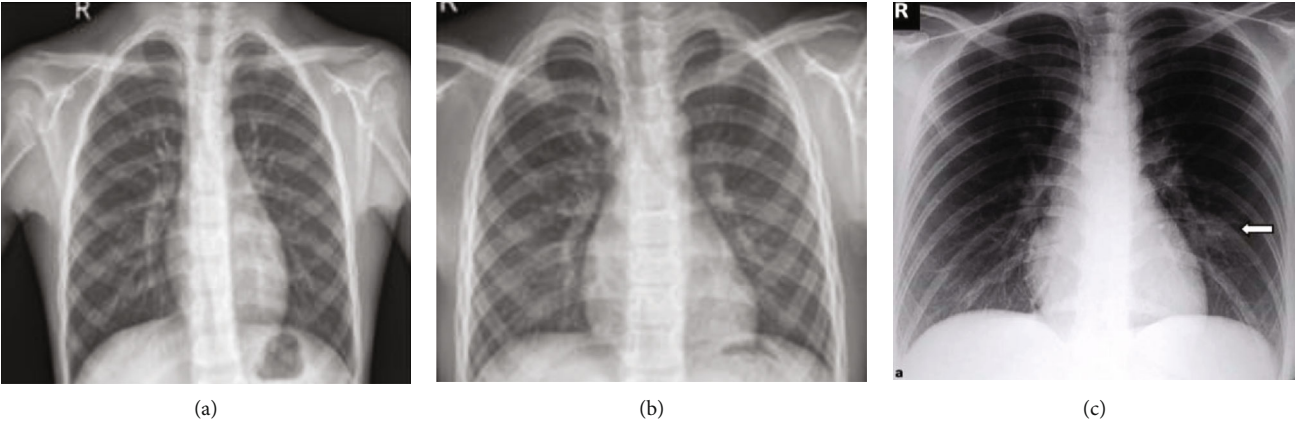


FIGURE 3: Sample image from testing dataset: (a) normal; (b) pneumonia; (c) COVID-19.

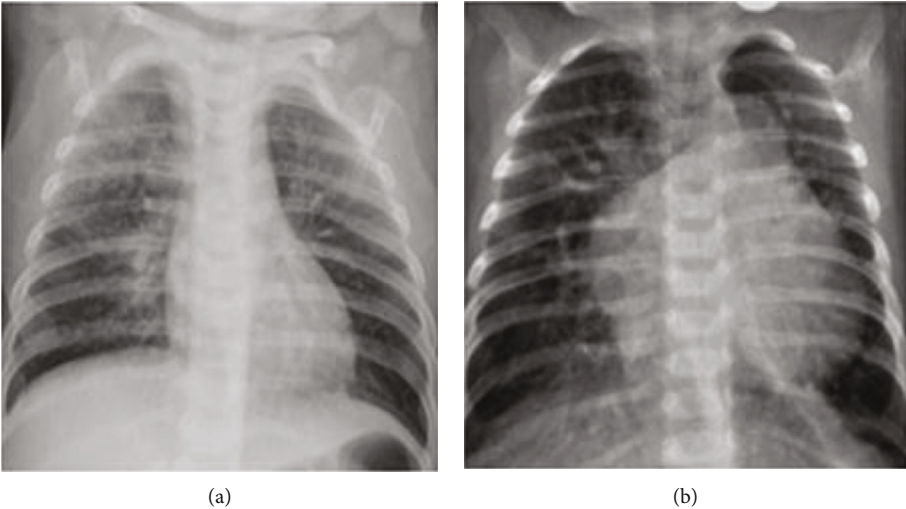


FIGURE 4: Sample image from validated dataset: (a) normal; (b) pneumonia.

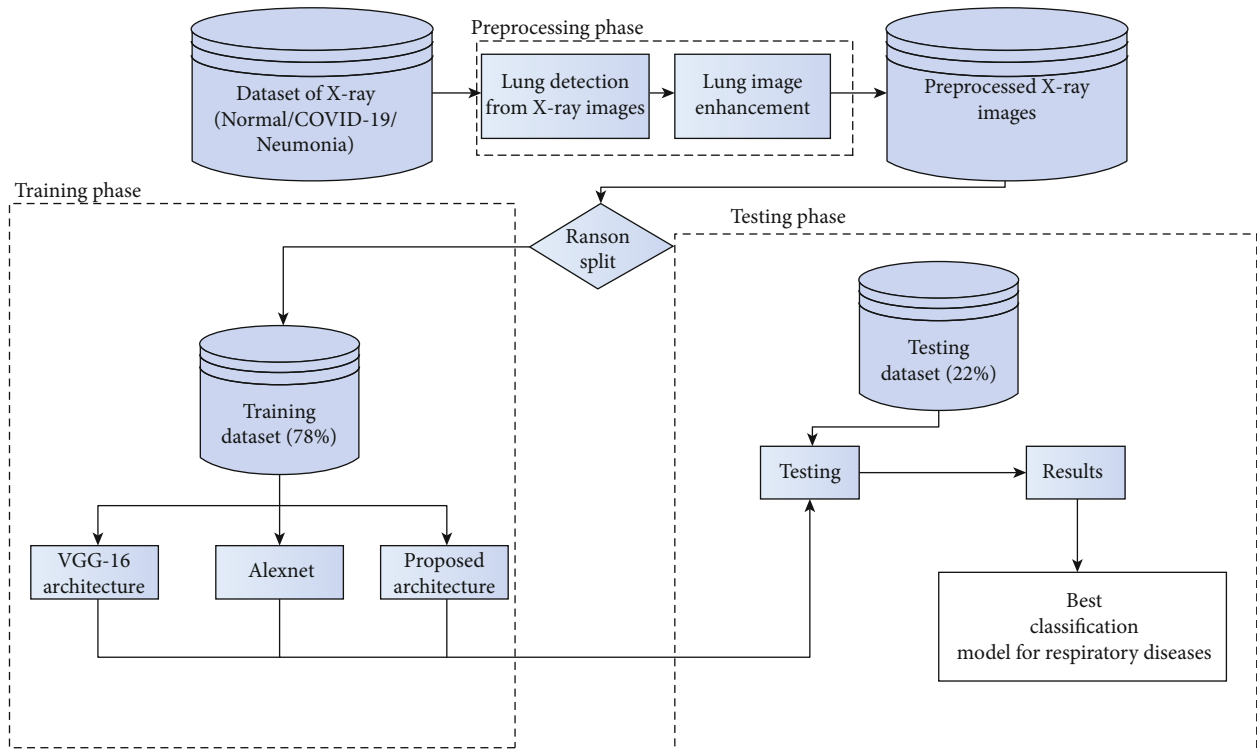


FIGURE 5: Model development.

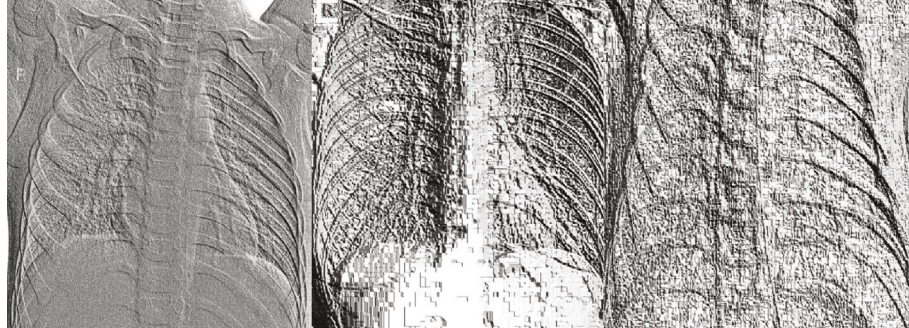


FIGURE 6: Feature described images: (a) normal, (b) pneumonia, and (c) COVID-19.

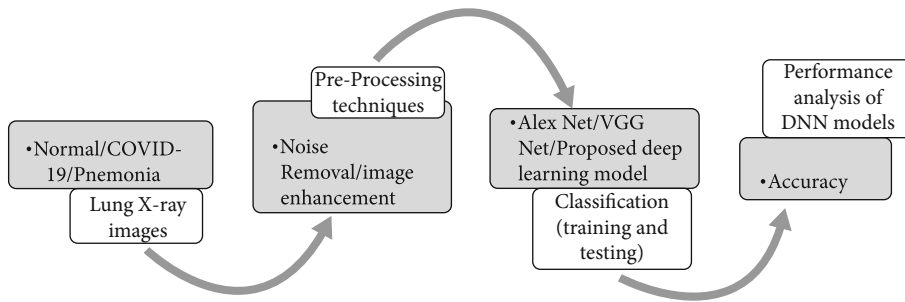


FIGURE 7: Block diagram on diagnosing the disease.

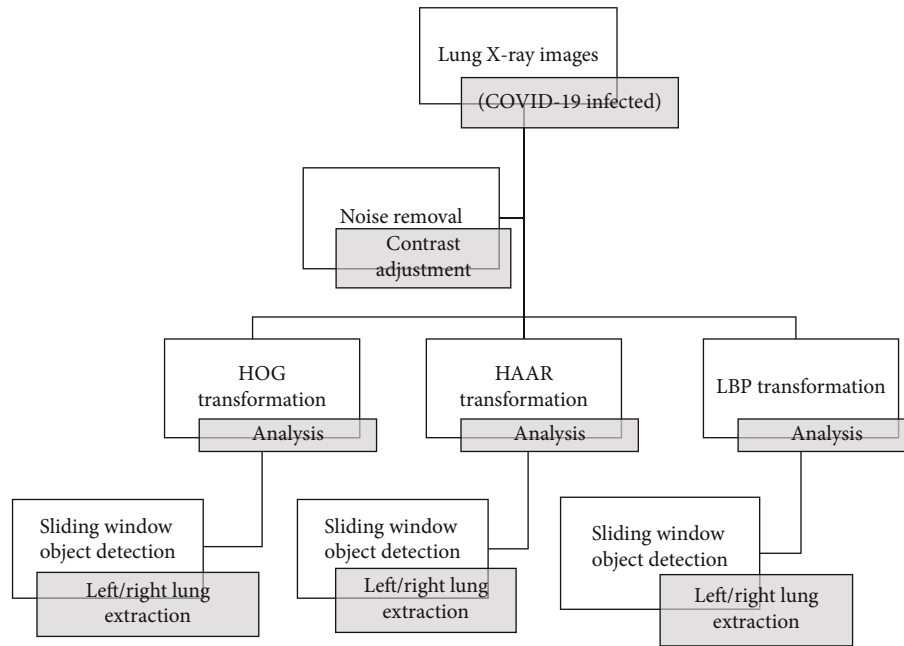


FIGURE 8: Preprocessing of the X-ray images.

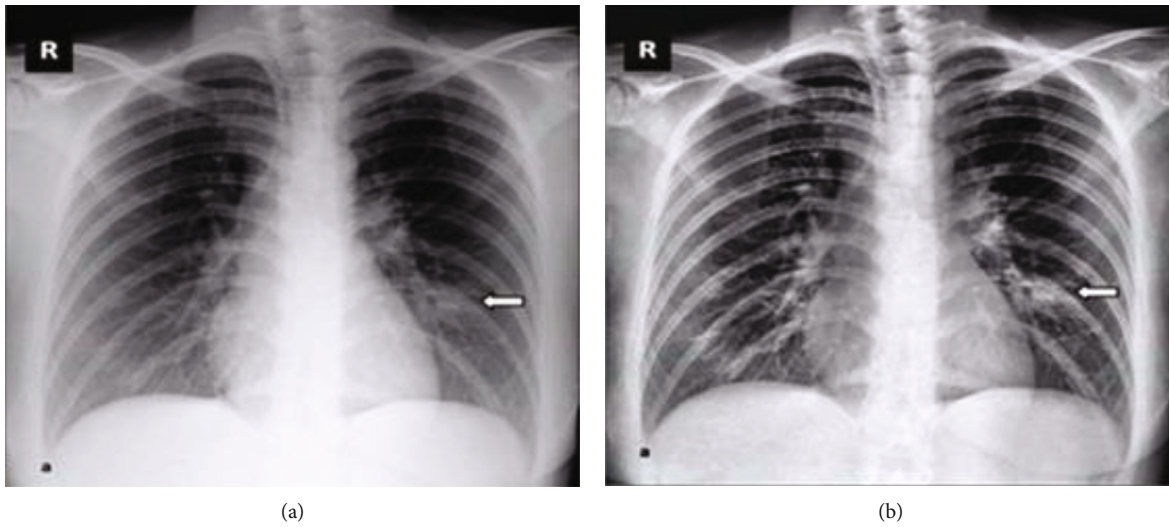


FIGURE 9: (a) Left image shows normal lung image; (b) right image shows processed image using contrast adjustment.

A classifier predicted the possibility of pneumonia. The developed ensemble model provided an accuracy of 96.4%. They have utilized the Guangzhou Women and Children’s Medical Center dataset. Ozturk et al. [15] proposed a model to provide diagnostics on normal, COVID-19, and pneumonia with binary and multiclass classification. Their model provided classification accuracy of 98.08% for binary classes and 87.02% for multiclass cases. In a recent research, the authors have implemented a deep neural network to differentiate COVID-19-induced pneumonia and other virus-induced pneumonia from chest X-ray images [16]. Researches have analyzed the effectiveness of the deep learning model VGG16 for the detection of COVID-19 and pneumonia [17]. The authors have claimed the work as a screening test based on the sensitivity and degree of specific-

ity. In a similar study [18], the authors have presented an automatic COVID-19 screening system which used radiomic texture descriptors to differentiate the CXR images to identify the normal, COVID-19, and suspected infected patients. Researchers have predicted the severity of COVID-19 from chest X-ray images [19]. Their method can measure lung infections as well as monitor treatment in ICU. Sharma et al. [20] have applied transfer learning for the classification of respiratory diseases including tuberculosis, pneumonia, COVID-19, and normal lung. Hence, based on the recent literature [21–29], the proposed method has implemented few efficient image preprocessing algorithms and deep learning models to diagnose respiratory diseases. The deep learning models developed is efficient when the input images are pre-processed with contrast enhancement, segmentation, and

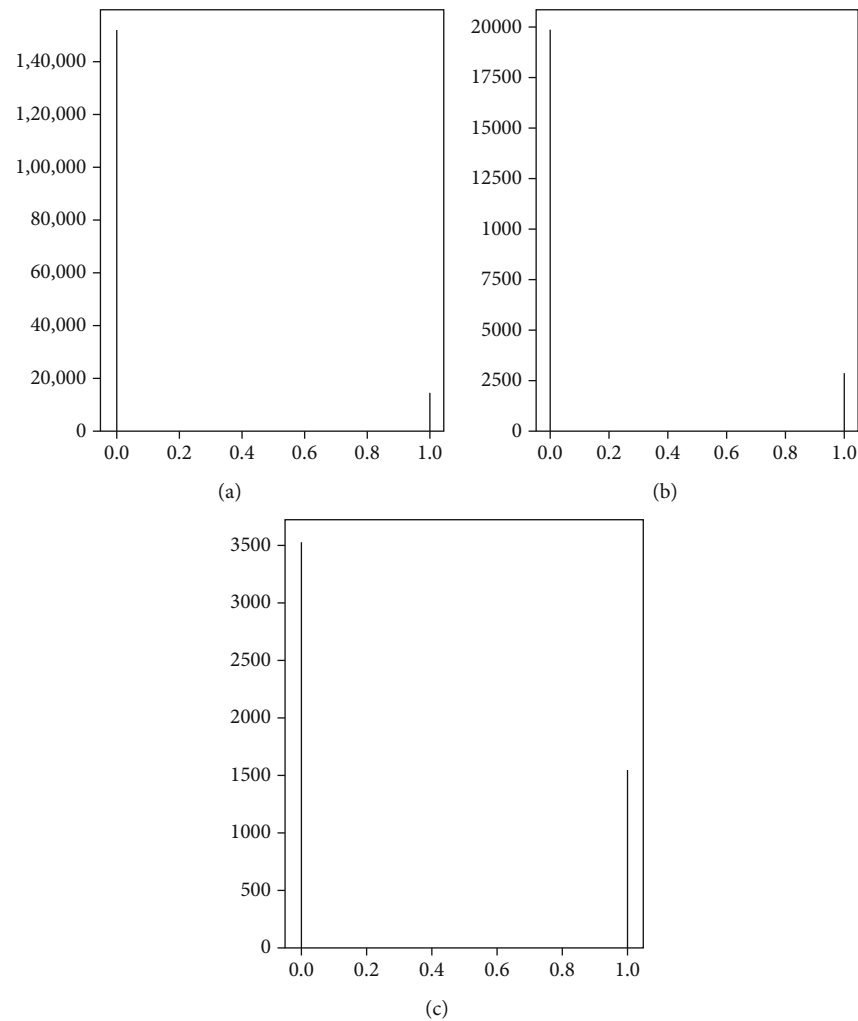


FIGURE 10: (a) Normal; (b) pneumonia; (c) COVID-19.

feature extraction. The accuracy improved in the classification of respiratory diseases.

4. Methodology

The proposed method involves preprocessing and classification of lung X-ray images using image processing algorithms and deep learning techniques as shown in Figure 1.

The X-ray images are preprocessed to locate the lungs. As per the literature, the stages of COVID-19 and pneumonia can be detected from the lung. Hence, the left lung and right lung have to be thoroughly analyzed to differentiate normal/abnormal images. Various algorithms on preprocessing have been applied, and the performance is evaluated. The processed data would be used as labelled data for the convolutional neural network model for the classification of respiratory diseases.

5. Dataset

Data is the first step in establishing any method or therapy for the diagnosis. The X-ray images of a COVID-19/pneumonia patient's lung are used for preprocessing. The data

is collected from a public database consisting of X-ray and CT scan images for various diseases, a project approved by the University of Montreal's Ethics Committee. The images of COVID-19 are drawn out from a public database and used for further processing. 78% of COVID-19 X-ray images are randomly split into the training dataset for training the classifier, and 22% of COVID-19 X-ray images are used for validation. The lung part in images of the training dataset is segmented physically to train the machine learning model to segment the lungs from the COVID-19 lung X-rays. Similarly, a total of 1200 images including normal and pneumonia (training and testing) were utilized for the proposed method and shown in Table 1. Also, the sample dataset is shown in Figures 2-4.

6. The Proposed Model Development for Respiratory Disease Detection

The following steps have been followed in the model development and depicted in Figure 5:

- (i) Apply deep learning edge detection model on COVID-19 images

- (ii) Apply deep learning edge detection model on pneumonia images
- (iii) Apply deep learning edge detection model on normal images
- (iv) Make COVID-19 directory which contains training (110 images) and testing (30 images) directories
- (v) Make pneumonia directory which contains training (110 images) and testing (30 images) directories
- (vi) Make a directory which contains normal training (110 images) and testing (30 images) directories
- (vii) Develop a CNN model using pretrained model
- (viii) Develop architecture for CNN model
- (ix) Develop a hybrid model with machine learning and deep learning

7. Preprocessing Techniques

7.1. Haar Cascade Classifier. Lung detection using Haar feature-based cascade classifiers is an effective location technique proposed by Paul Viola and Michael Jones in 2001. It is an AI-based methodology where a cascade work is prepared from a great deal of positive and negative pictures. It is then used to distinguish questions in different pictures.

In this proposed method, the discovery of lungs in X-ray images is defined. At first, the algorithm needs positive (pictures of appearances) and negative (pictures without lungs) to prepare the classifier followed by feature extraction from the images to train the model. The obtained Haar features are utilized. The features are mostly the same as the convolutional kernel. Each element is a solitary worth acquired by deducting the aggregate of pixels under a white rectangle shape from the entirety of pixels under the dark rectangle shape.

The cascading of the classifiers helps to evaluate even the subimages with the greatest likelihood for all Haar transform that differentiates an entity. This also helps in adjusting a classifier's accuracy. Viola and Jones were able to detect a human face with an accuracy rate of 95 per cent using only 200 simple features. Hence, the Haar cascade classifier is trained to classify lungs using the extracted features. This gentle AdaBoost algorithm and Haar feature algorithms have been implemented to train these classifiers, and in this process, more than 6000 features were included to classify with greater accuracy.

7.2. HOG for Lung Segmentation. HOG, or histogram of oriented gradients, is an element descriptor that is frequently used to extricate features from picture information. It is broadly utilized in PC vision errands for object location. We should take a gander at some significant parts of HOG that make it unique in relation to other component descriptors.

The HOG descriptor centers on the structure or the state of an item. Hoard can give the edge course too. This is finished by extracting the slope and direction of the edges. Furthermore, these directions are determined in "restricted" parcels. This implies that the total picture is separated into

TABLE 2: Proposed deep neural network layers.

Layer (type)	Output shape	Param #
conv2d (Conv2D)	(None, 224, 224, 64)	1792
conv2d_1 (Conv2D)	(None, 224, 224, 64)	36928
max_pooling2d (MaxPooling2D)	(None, 112, 112, 64)	0
conv2d_2 (Conv2D)	(None, 112, 112, 128)	73856
conv2d_3 (Conv2D)	(None, 112, 112, 128)	147584
max_pooling2d_1 (MaxPooling2)	(None, 56, 56, 128)	0
conv2d_4 (Conv2D)	(None, 56, 56, 256)	295168
conv2d_5 (Conv2D)	(None, 56, 56, 256)	590080
conv2d_6 (Conv2D)	(None, 56, 56, 256)	590080
max_pooling2d_2 (MaxPooling2)	(None, 28, 28, 256)	0
conv2d_7 (Conv2D)	(None, 28, 28, 512)	1180160
conv2d_8 (Conv2D)	(None, 28, 28, 512)	2359808
conv2d_9 (Conv2D)	(None, 28, 28, 512)	2359808
max_pooling2d_3 (MaxPooling2)	(None, 14, 14, 512)	0
conv2d_10 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_11 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_12 (Conv2D)	(None, 14, 14, 512)	2359808
max_pooling2d_4 (MaxPooling2)	(None, 7, 7, 512)	0
flatten (flatten)	(None, 25088)	0
dense (dense)	(None, 4096)	102764544
dense_1 (dense)	(None, 4096)	16781312
dense_2 (dense)	(None, 3)	12291
Total params: 134272835;		
trainable params: 134272835;		
nontrainable params: 0		

littler locales, and for every area, the inclinations and direction are determined. The HOG would produce a histogram for every one of these locales independently. The histograms are made utilizing the slopes and directions of the pixel esteems, thus the name "histogram of oriented gradients"

- (i) Preprocess the data (64×128)
- (ii) Calculating gradients (directions x and y)
- (iii) Calculate the magnitude and orientation

Using each pixel's magnitude and orientation in the X-ray, it is graded into a lung image or not. Support vector machine (SVM) helps in classifying an image and works more efficiently with HOG transform. The SVM model is essentially a reflection in multidimensional space of various classes in a hyperplane. The hyperplane is created by SVM iteratively to minimize the cost function. SVM is aimed at segmenting the datasets into classes in order to find a maximal absolute hyperplane.

7.3. LBP Transform for Lung Segmentation. The substance of an individual passes on huge information of data about the

TABLE 3: AlexNet: deep neural network layers.

Layer (type)	Output shape	Parameter #
conv2d_13 (Conv2D)	(None, 54, 54, 96)	34944
activation (activation)	(None, 54, 54, 96)	0
max_pooling2d_5 (MaxPooling2)	(None, 27, 27, 96)	0
conv2d_14 (Conv2D)	(None, 17, 17, 256)	2973952
activation_1 (activation)	(None, 17, 17, 256)	0
max_pooling2d_6 (MaxPooling2)	(None, 8, 8, 256)	0
conv2d_15 (Conv2D)	(None, 6, 6, 384)	885120
activation_2 (activation)	(None, 6, 6, 384)	0
conv2d_16 (Conv2D)	(None, 4, 4, 384)	1327488
activation_3 (activation)	(None, 4, 4, 384)	0
conv2d_17 (Conv2D)	(None, 2, 2, 256)	884992
activation_4 (activation)	(None, 2, 2, 256)	0
max_pooling2d_7 (MaxPooling2)	(None, 1, 1, 256)	0
flatten_1 (flatten)	(None, 256)	0
dense_3 (dense)	(None, 4096)	1052672
activation_5 (activation)	(None, 4096)	0
dropout (dropout)	(None, 4096)	0
dense_4 (dense)	(None, 4096)	16781312
activation_6 (activation)	(None, 4096)	0
dropout_1 (dropout)	(None, 4096)	0
dense_5 (dense)	(None, 1000)	4097000
activation_7 (activation)	(None, 1000)	0
dropout_2 (dropout)	(None, 1000)	0
dense_6 (dense)	(None, 3)	3003
Total params: 28040483; trainable parameters: 28040483; nontrainable parameters: 0		

lung and enthusiastic condition of the individual. The simplest operator is to set the middle pixel “Pm” value as a threshold and associate it with the pixels in the neighbourhood, Pn. The value of Pm is calculated. All threshold values of neighbourhood points are multiplied by the corresponding proportion and added. The LBP code for k -bit can be calculated as

$$\text{LBP} = \sum_{i=0}^{i=k-1} F(\text{Pn} - \text{Pm})2^i. \quad (1)$$

Pn represents the n^{th} neighbouring pixel, and “m” represents the center pixel. The histogram features of size 2^i are extracted from the obtained local binary pattern code.

The lung region is first partitioned into little areas from which local binary patterns (LBP) and histograms are extricated and linked into a solitary component vector. This element vector shapes a proficient portrayal of the lung and is utilized to quantify likenesses between pictures.

TABLE 4: VGGNet: deep neural network layers.

Layer (type)	Output shape	Parameter #
conv2d (Conv2D)	(None, 444, 444, 32)	2432
max_pooling2d (MaxPooling2D)	(None, 222, 222, 32)	0
conv2d_1 (Conv2D)	(None, 220, 220, 64)	18496
max_pooling2d_1 (MaxPooling2)	(None, 110, 110, 64)	0
dropout (dropout)	(None, 110, 110, 64)	0
conv2d_2 (Conv2D)	(None, 108, 108, 128)	73856
max_pooling2d_2 (MaxPooling2)	(None, 54, 54, 128)	0
dropout_1 (Dropout)	(None, 54, 54, 128)	0
conv2d_3 (Conv2D)	(None, 52, 52, 512)	590336
max_pooling2d_3 (MaxPooling2)	(None, 26, 26, 512)	0
dropout_2 (Dropout)	(None, 26, 26, 512)	0
conv2d_4 (Conv2D)	(None, 24, 24, 512)	2359808
conv2d_5 (Conv2D)	(None, 22, 22, 128)	589952
conv2d_6 (Conv2D)	(None, 20, 20, 64)	73792
max_pooling2d_4 (MaxPooling2)	(None, 10, 10, 64)	0
dropout_3 (dropout)	(None, 10, 10, 64)	0
flatten (flatten)	(None, 6400)	0
dense (dense)	(None, 4096)	26218496
dense_1 (dense)	(None, 1024)	4195328
dropout_4 (dropout)	(None, 1024)	0
dense_2 (dense)	(None, 3)	3075
Total parameters: 34125571; trainable parameters: 34125571; nontrainable parameters: 0		

The LBP transformed image has features that can be used for classification, and sample images are shown in Figures 6(a)–6(c). Feature extraction helps the classification model in classifying the image with better accuracy.

7.4. Diagnosis of Disease. The dataset consists of normal lung images, bacterial pneumonia affected lung images, and COVID-19 affected lung images. The images are preprocessed to remove noise and to enhance the image quality. The images are scaled and cropped for increasing the performance of the proposed deep learning model. After preprocessing, AlexNet, VGGNet, and the proposed neural network are trained on normal/abnormal lung images. After training, the models are tested with data and the algorithm with the best accuracy is selected for further diagnosing the process. The process is represented in the block diagram in Figure 7.

7.5. Implementation

7.5.1. Preprocessing Algorithms. COVID-19 lung images are used to train the classifier in the detection of lungs from

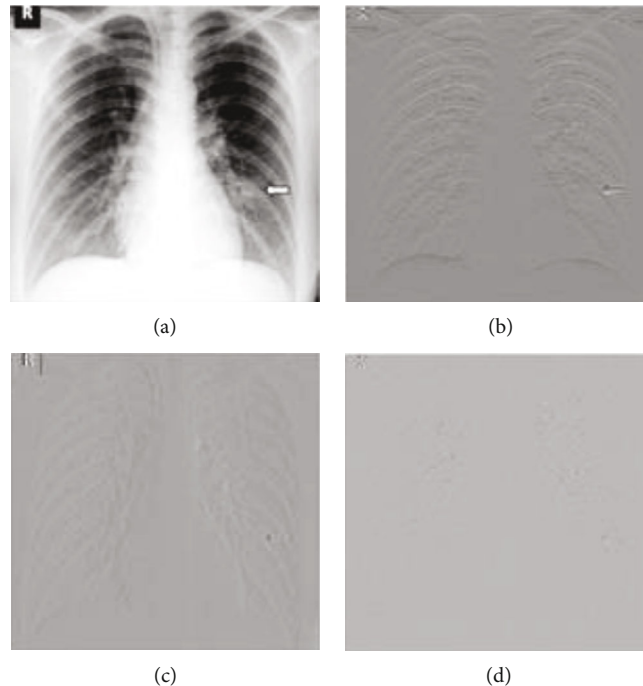


FIGURE 11: (a) Original image; (b) horizontal detail; (c) vertical detail; (d) diagonal detail.

the X-ray images. Various transformations like HOG, Haar, and LBP were applied on images, and depending on the accurate segmentation of lungs, the best accurate algorithm for the detection has been chosen. In Figure 8, the process is depicted on preprocessing of COVID-19 images, and extraction of the lungs using the sliding window technique is highlighted. The preprocessed images (contrast enhancement) are shown in Figure 9.

7.5.2. Noise Removal. The noises from the dataset are removed to get the best features from the COVID-19/pneumonia lung images. Some of the preprocessing techniques that can be applied to images are contrast adjustment, intensity adjustment, histogram equalization, binarization, morphology operation, etc. For lung X-ray images, contrast adjustment and histogram equalization perform better when compared with other preprocessing techniques. Contrast adjustment of a COVID-19 lung image has produced a better image to extract the features from the lungs. These features can further be used for the detection of lungs from X-ray images using various lung detection algorithms. As shown in Figures 10(a)–10(c), histogram equalized images represent normal, pneumonia, and COVID-19 with gray level and frequency count. The histogram equalization on an image produces images with better quality without loss of information. This contrast adjustment using histogram equalization helps in extracting better features in lung detection algorithms.

7.5.3. Detection of Lungs Using HOG Feature Vector, Haar Transform, and LBP Transform. The algorithm used for lung detection is the sliding window technique where a window of size starting from $2 \times 2 \times 3$ to $N \times N \times 3$ is used to detect

the left and right lungs of an image. The window of constant size is allowed to slide over the entire image, and the window will be utilized with transform techniques to locate lungs. The algorithm shown below can be used in various applications to get better accuracy.

Step 1. Contrast adjustment of X-ray image.

Step 2. Sliding window with window size from 2 to size of the image (N).

Step 3. Classification of image in window into left the lung using HOG feature descriptor, Haar transform, and LBP transform.

Step 4. Classification of image in window into right lung using HOG feature descriptor, Haar transform, and LBP transform.

Step 5. Repeat steps from Steps 2 to 4 until the left lung and right lungs are extracted properly with every algorithm.

Step 6. Comparing the extracted images of lungs to predict the best algorithm for feature description.

8. Proposed Deep Neural Network

The proposed DNN model is a powerful network which is capable of getting higher accuracy on classification of respiratory diseases. This architecture is used to classify X-ray images of lungs into COVID-19, pneumonia caused by bacteria, and a normal lung image. This architecture can also be extended to various object classification models.

In medical image analysis, accuracy is the major factor for architecture to be implemented in any system. This model consists of 7 sequentially connected convolutional layers with different size filters. These convolutional layers are sequentially connected to three fully connected layers. The major characteristics of this architecture are features

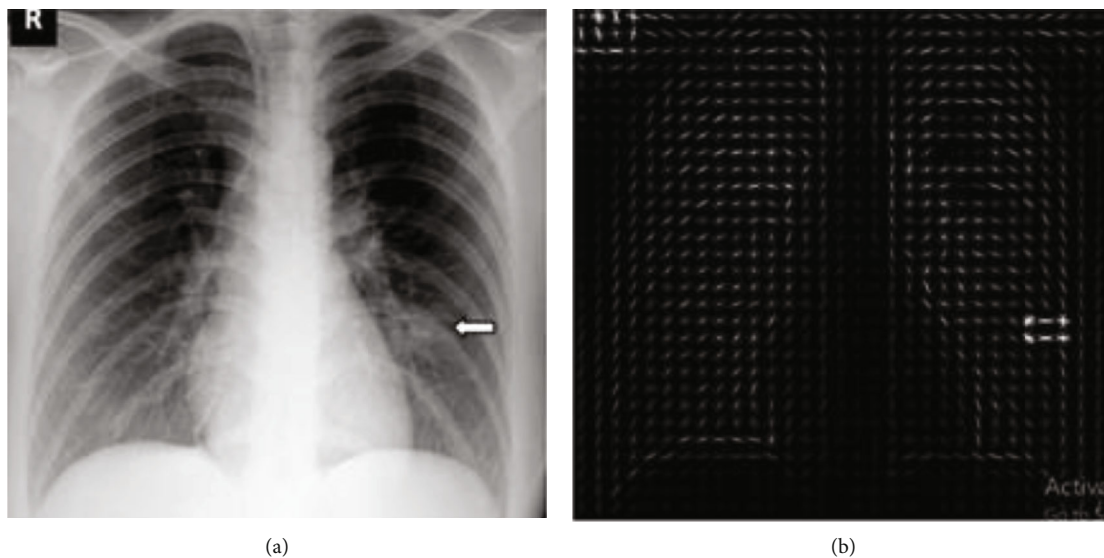


FIGURE 12: (a) Original image; (b) histogram of oriented gradients.

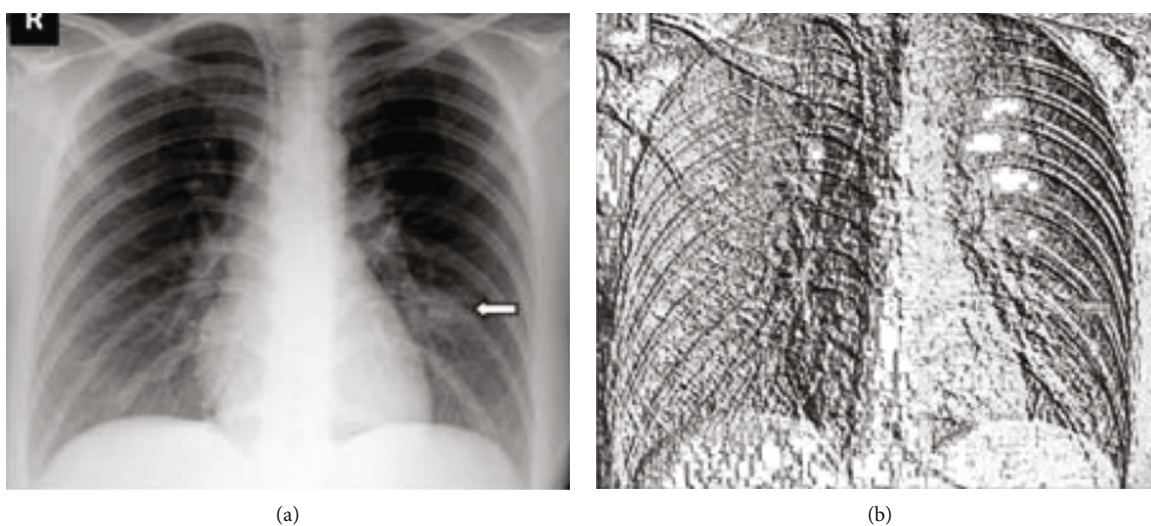


FIGURE 13: (a) Original image; (b) local binary pattern image.

of the image passed through strided convolutions and max pooling layers. The architecture has used 5×5 and 3×3 size filters. The architecture has convolutions, max pooling, dropout, and ReLU activation layers.

8.1. Architecture. The proposed architecture has 10 layers where 7 layers are convolutional layers and the remaining 3 are fully connected layers and shown in Table 2. A fixed size of $448 \times 448 \times 3$ RGB lung image of COVID-19 should be given as input to the network. After preprocessing of the image, the image has been subjected to the network for classification. The first layer in architecture uses " 5×5 " size filter with stride 1 pixel. All other layers in architecture use " 3×3 " size filter with a stride of 1 pixel. These filters can help to get all features of an image. Max pooling is performed on different layers over 2×2 size filter with stride 1. Each convolutional

layer is followed by the Rectified Linear Unit (ReLU) to add nonlinearity to the model. This nonlinearity can classify the images with high accuracy. The architecture contains 3 fully connected dense layers where the first layer is of size "4096" and the second layer is of size "1024." The final layer which classifies the image is of size "3". Softmax function is applied to the output layer to classify the image.

9. AlexNet Deep Neural Network

AlexNet is an incredibly adaptable model which, on particularly problematic datasets, can achieve high accuracy. Regardless, dispensing with the whole of the traditional layers will profoundly degenerate AlexNet's introduction. AlexNet is a fundamental plan for any movement and can have enormous consequences for issues with automated

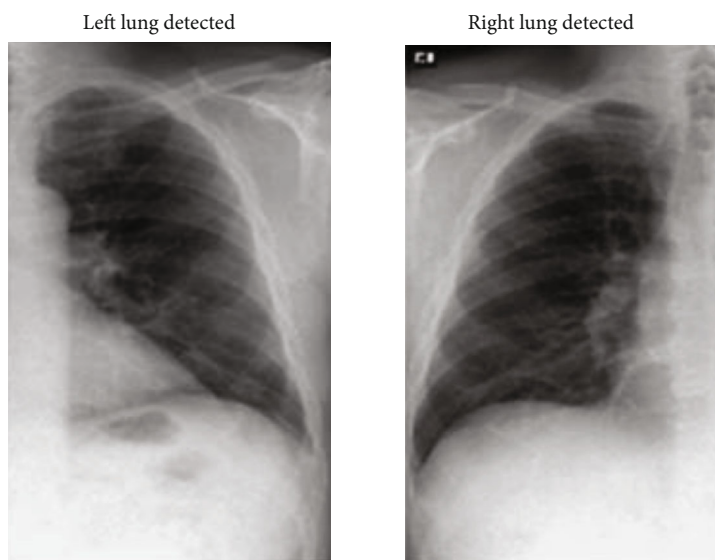


FIGURE 14: Left and right lung detection using Haar transform.

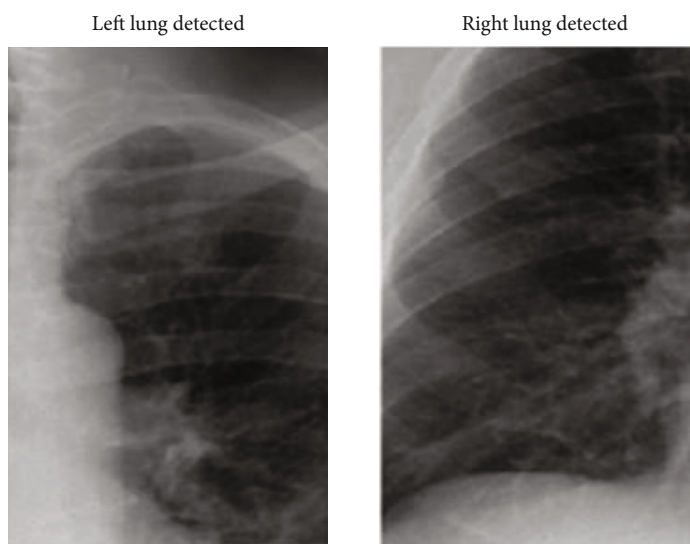


FIGURE 15: Left and right lung detection using local binary pattern algorithm.

TABLE 5: Existing methods on classification models.

Author	Methodology	Year of publication	Accuracy on multiclass (%)
Ozturk et al.	Darknet model	2020	87.02%
Civit-Masot et al.	VGG-16	2020	90%
Cohen et al.	Regression model	2020	1.14 mean absolute error

thinking, keeping watch for PC vision. The model includes five progressively associated convolutional layers of lessening channel size, followed by three totally associated layers. One of AlexNet’s fundamental features is the incredibly speedy down testing of the widely appealing depictions through strided convolutions and max pooling layers. The delayed consequences of AlexNet (Table 3)

have demonstrated that on an uncommonly testing dataset, an immense significant convolutionary neural association is good for achieving record-breaking results using coordinated learning. AlexNet is made out of eight layers. The underlying 5 are convolutionary, and the last 3 are completely related layers. We also have couple “layers,” called pooling and order.

TABLE 6: Intersection over union scores for detected left and right lungs.

Algorithm	Right lung (IOU)	Left lung (IOU)
Hog feature descriptor	0.59	0.65
Haar transform	0.86	0.81
Local binary pattern	0.36	0.54

10. VGGNet

VGG is a contraction for the Oxford University Visual Geometric Group, and VGG-16 is a 16-layer network proposed by the Visual Geometric Group. These 16 layers contain likely limits and various layers; for instance, the max pool layer is available (Table 4), yet no workable limits are joined.

10.1. Architecture. The plan is essential, followed by a most extreme pooling layer of two flanking squares of 2 convolution layers; by then, it has three adjoining squares of 3 convolution layers, followed by max pooling; finally, we have three thick layers. The last 3 convolution layers have obvious profundities in different models. The fundamental examination is that after each most extreme pooling measure, the size is reduced considerably. A fixed size (224 * 224) RGB picture is given as a commitment to this association which infers that the structure has been made (224, 224, 3). The primary preplanning is that they deducted the mean RGB regard from each pixel, assessed over the whole getting ready grouping. Pieces with a size of 3 * 3 and a phase size of 1 pixel were used to cover the entire idea of the image. With stage 2, over 2 * 2 pixel windows, max pooling is accomplished. Three totally related layers were progressively connected to these convolutionary layers, the underlying two of which were 4096 in size, and afterward, the last layer was a layer with 1000 channels for 1000-way ILSVRC gathering and a softmax feature.

11. Results and Discussion

11.1. Extracting Features from Haar Transform. The original image in Figure 11(a) is taken, and various Haar transform filters are applied to obtain different image characteristics. As shown in Figures 11(a)–11(d), few features are extracted from the original X-ray image in horizontal, vertical, and diagonal, and the respective feature image is created. These characteristics can be extracted using various kernels. This process with several thousand features extracted from images provides better accuracy.

11.2. HOG Feature Descriptor. For each pixel, the gradient direction and magnitudes are determined, and feature vectors are extracted using histograms. The vector of the extracted feature from an image helps extract the features from the image. In classifying a picture into a lung image or a nonlung image, these features are helpful. The illustration in Figure 12(a) is a pictorial representation of the magnitudes and directions of the gradient. In Figure 12(b), whenever there is no shift in contrast, hog gradients are almost zero.

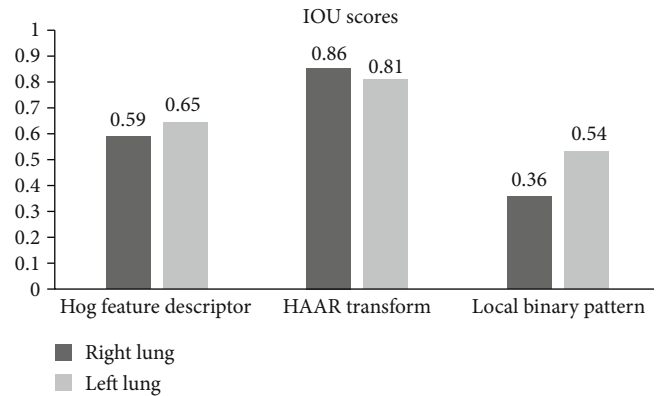


FIGURE 16: IOU scores of proposed image processing algorithms.

TABLE 7: Performance of deep learning models.

Deep learning models	Accuracy on X-ray images on classification of diseases
Proposed deep neural network	91.40%
AlexNet	75.00%
VGGNet	87.5%

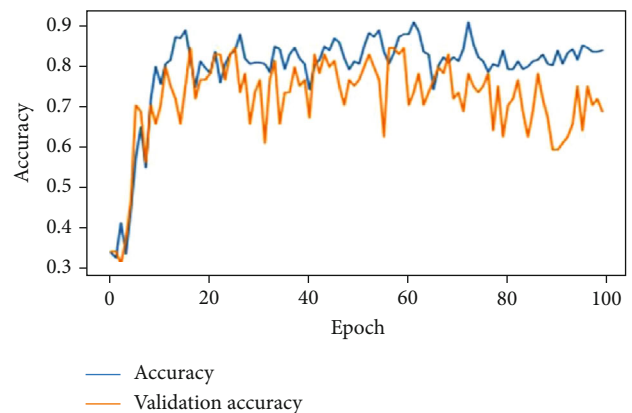


FIGURE 17: VGGNet after 100 epochs.

11.3. LBP Feature Descriptor. The local binary pattern is a plain and gray-scale invariant feature descriptor which is highly used for extracting features for an image in a classification model. In LBP, by thresholding its neighbourhood pixels to either 0 or 1 depending on the center pixel value, a binary code is created at each pixel. Figure 13(b) is an image which contains all the features described from the original COVID-19 lung image in Figure 13(a).

In Figure 14, the left and right lungs are detected after extracting thousands of features from the image in which the lungs have to be detected. Cascade classification is used to classify the image into lung or not. The cascade classifier uses the features extracted using Haar transform to get the best accuracy in the classification process. In Figure 15, the lungs are detected using the feature description vector of every pixel in the original COVID-19 image. Through the support vector machine algorithm, the image would be

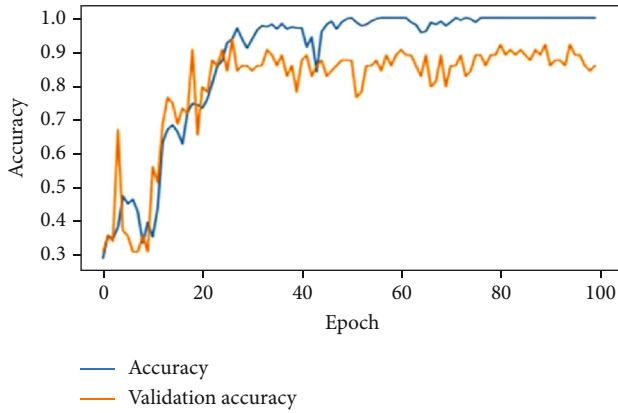


FIGURE 18: AlexNet after 100 epochs.

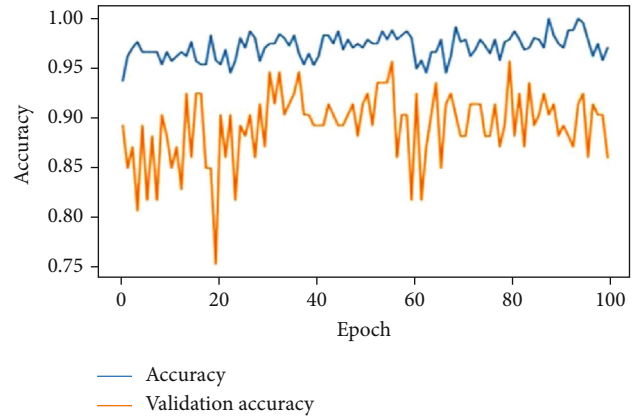


FIGURE 19: Proposed DNN after 100 epochs.

subjected to identification of lungs. The support vector machine helps in classifying an image with better accuracy. The LBP transform helps in extracting features at every pixel with the help of neighbouring pixels. Various Machine Learning Algorithms can be applied to LBP transformed images to classify a COVID-19 image. The results in Figure 14 are classified using the support vector machine and provided better accuracy.

11.4. Result Analysis and Performance Evaluation. The objective of this proposed method is to classify lung X-ray images into normal (without any viral/bacterial infections), COVID-19 disease, and pneumonia disease using deep learning techniques. The study revealed that the preprocessing of data improved the performance of the developed model in diagnosing respiratory diseases. This study also implemented three deep learning models for the classification. AlexNet and VGG-16 were the pretrained models, and the proposed deep neural networks were the developed model. A comparative analysis of these three models revealed that the developed model outperformed the pretrained model in terms of IoU scores.

The results are promising (91%) in multiclass classification when compared to the results of existing works in respiratory system disease diagnosis and tabulated in Table 5.

To measure the best algorithm out of the above three algorithms, intersection over union scores are calculated. Intersection over union can be calculated by

$$IOU = \frac{\text{Area of overlap}}{\text{Area of union}} \tag{2}$$

In equation (2), the area of overlap is the intersection of the area of ground truth image with the area of the predicted image, and the area of union is the union of the area of ground truth image with the area of the predicted image. The algorithm with IOU scores above 0.5 can be used as the detector algorithm. As per Table 6, detection of lungs using Haar transform was performing better when compared with the other two algorithms

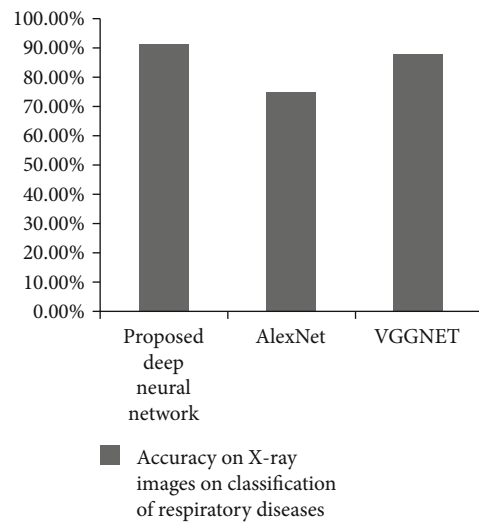


FIGURE 20: Analysis of classification accuracy.

In Figure 16, the applied algorithms for feature extraction are plotted with their IOU scores for the left and right lungs.

The results clearly depict that Haar transform outperformed the Hog feature descriptor and local binary pattern methods.

11.5. Evaluation of Classification Models. The feature extracted X-ray lung images which are subjected to the proposed deep neural network, AlexNet and VGGNet. The developed architectures classify the images into three classes of respiratory diseases as normal, COVID-19, and pneumonia. The accuracy of the lung X-ray images on test data is given in Table 7. As per the results, the proposed deep neural network was performing better when compared with the other 2 predefined networks. Figures 17–19 show the accuracy and validation accuracy per epoch which can help to determine the best model. The validation accuracy is almost on par with the model accuracy which proves that the trained model is efficient in the classification of respiratory diseases.

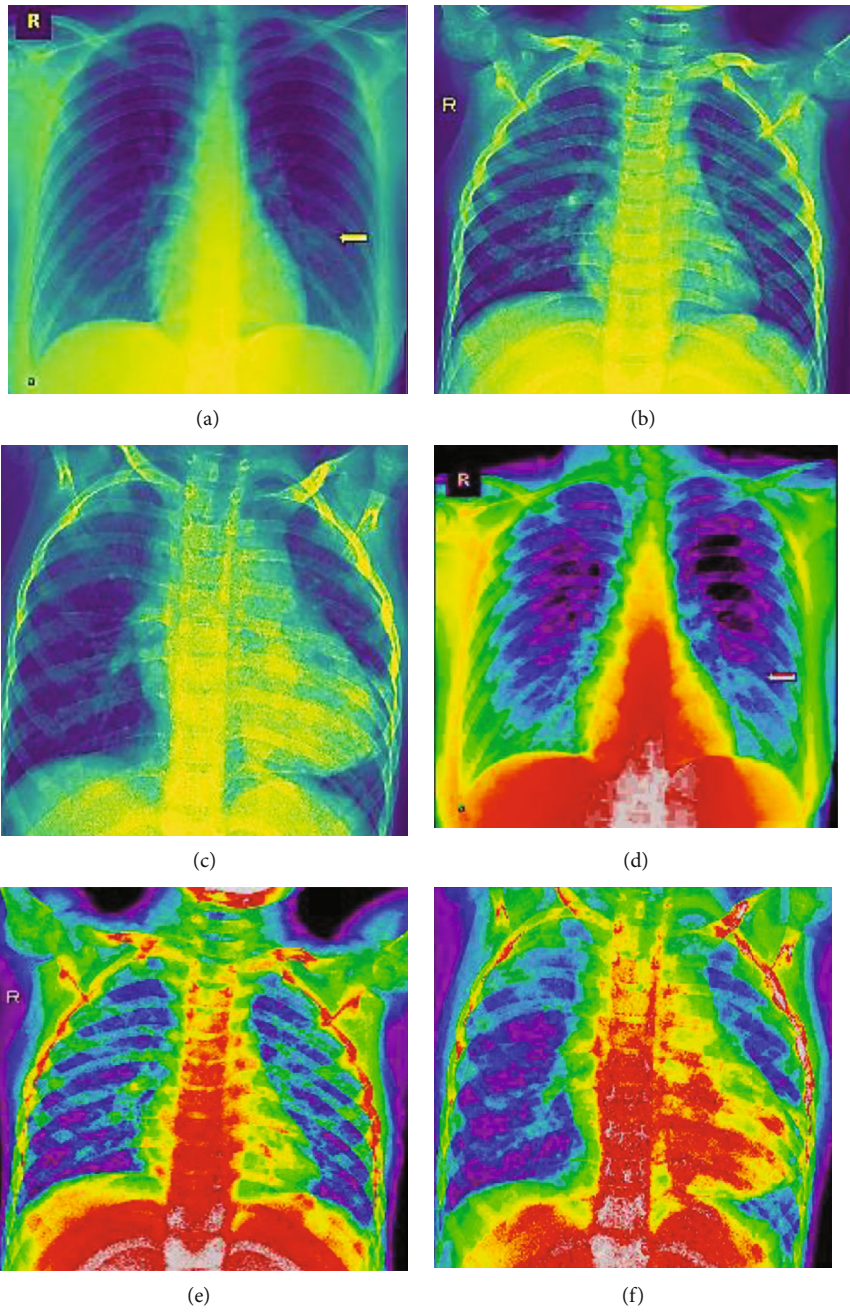


FIGURE 21: Continued.

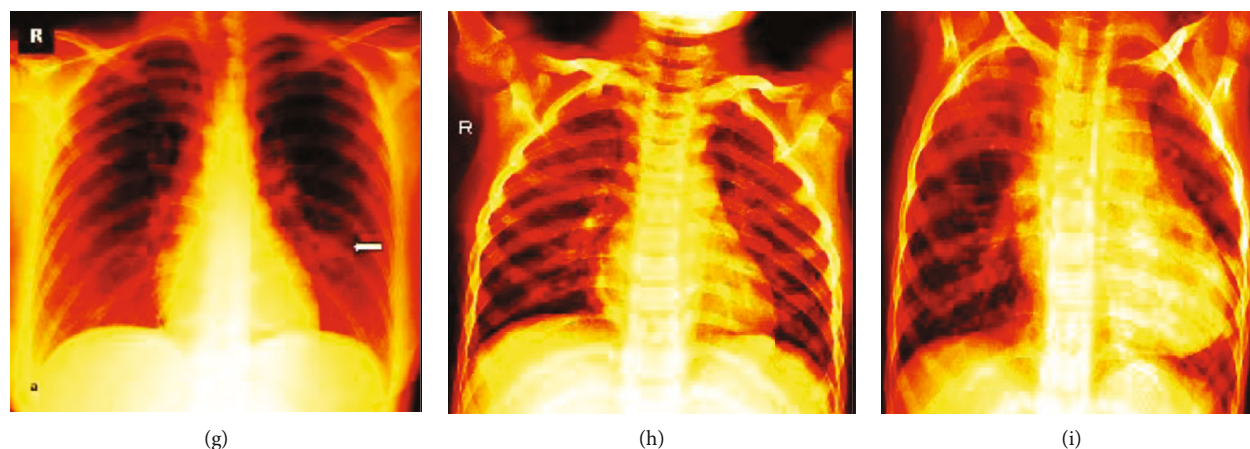


FIGURE 21: (a) Normal-colorbars; (b) COVID-19-colorbars; (c) pneumonia-colorbars. (d) Normal nipy_spectral; (e) COVID-19 nipy_spectral; (f) pneumonia nipy_spectral. (g) Normal-viridis. (h) COVID-19-viridis. (i) Pneumonia-viridis (other image preprocessing results).

The image preprocessing techniques shown in Figure 20 along with feature descriptors (Figure 8) applied on the dataset also aid in the classification accuracy of the proposed deep neural network model. Pseudocolor is a useful tool for contrast enhancement and visualization of data. The contrast of the X-ray images was poor, and hence to improve, the proposed method included addition of data visualization.

The grayscale images are processed using Matplotlib's pyplot. This is a powerful method suitable for large datasets. Figures 21(a)–21(c) display colorbars and indicate the data values as a colormap. The mapping highlights the most dominant features in identifying the disease. A sequential color-mapping technique, Viridis (Figures 21(d)–21(f)), highlights the infected portions. So, when a user refers to the developed model, he/she can be visually supported with the classification. The nipy_spectral method (Figures 21(g)–21(i)) further discriminates the anatomy of the lung with respect to normal/COVID-19 and pneumonia affected regions. The dataset is huge, and when we subject these three image processing techniques along with segmentation, feature descriptors, and finally with classification, the results are obvious and easy to predict the disease without the support of human assistance.

12. Discussion

The proposed method incorporated preprocessing techniques comprising noise removal and image enhancements. Segmentation and filtering were applied to identify left/right lungs. Then, the preprocessed images were subjected to classifiers for the classification of respiratory diseases. The results were compared to identify the most accurate preprocessing technique and classification model. Haar transform techniques outperformed other techniques based on the IOU scores. The proposed DNN accurately classified the diseases with 91.4% with runtime of the program to be 1944 seconds (32 minutes). The processor information is Intel® core™, 16GB RAM, 64-bit operating system. The problem statement requires a large dataset for training the deep learning model. This is time-consuming when trained using a 32/64-bit processor. This might further reduce when simulated in GPU. Hence, based on the results, we have improved

the accuracy of the classifiers by applying efficient feature extraction techniques and image preprocessing algorithms. The existing techniques on the classification of respiratory diseases are 90% accurate, and our proposed method provides an in-depth analysis of preprocessing techniques to improve the DNN model performance.

13. Conclusion

A novel method has been proposed for the efficient classification of respiratory diseases from chest X-ray lung images. The existing technologies focus on COVID-19 diagnosis whereas this proposed method focuses on all bacterial and viral infection lung diseases. In this pandemic situation, it is necessary to differentiate COVID-19 from pneumonia. The proposed method targets preprocessing and feature descriptors to efficiently classify life-threatening lung diseases. The chest X-ray images are preprocessed by applying various image processing algorithms. Then, the preprocessed images are subjected to Haar transform filters with various kernels to extract predominant features of the X-ray image. Gradient direction and magnitudes are calculated for every pixel, and feature vectors are also extracted using histograms. Similarly, a local binary pattern is also applied for extracting features for an image in the classification model.

Based on the features extracted by three different techniques and using an appropriate classification model, the left and right lungs are segmented.

- (i) The features extracted through Haar transform provide the best accuracy in segmentation of left and right lungs
- (ii) Cascade classification and support vector machine aid in identifying the presence of lungs
- (iii) Segmentation of lungs using Haar cascade classifier outperforms with an average IoU score of 81% when compared with the other two algorithms
- (iv) The feature described X-ray images are then subjected to the proposed deep neural network,

pretrained AlexNet and VGGNet for classification of lung diseases (normal/COVID-19/pneumonia)

The proposed deep neural network performed well with 91% of accuracy in classifying images into pneumonia bacteria, COVID-19, and normal. This model can help doctors to diagnose, study the diseases, and provide appropriate treatment. This architecture can be also used for diagnosing various other life-threatening diseases from medical images. Thus, the developed method focused towards achieving the major goals of the United Nations in promoting good health and well-being through diagnosis of COVID-19 and other respiratory diseases.

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 article.

Acknowledgments

The authors thank Prince Sultan University, Saudi Arabia, and Vellore Institute of Technology, India, for providing the resources in successfully completing the work. This work was supported by the research grants (COVID-19 Emergency research program), Prince Sultan University, Saudi Arabia (COVID19-CE-2020-53), Umashankar Subramaniam, <https://ric.psu.edu.sa/covid19-emergency-research-program.php>.

References

- [1] J. Segars, Q. Katler, D. B. McQueen et al., "Prior and novel coronaviruses, coronavirus disease 2019 (COVID-19), and human reproduction: what is known?," *Fertility and Sterility*, vol. 113, no. 6, pp. 1140–1149, 2020.
- [2] G. Li, R. Hu, and X. Gu, "A close-up on COVID-19 and cardiovascular diseases," *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 30, no. 7, pp. 1057–1060, 2020.
- [3] W. Wei, D. Zheng, Y. Lei et al., "Radiotherapy workflow and protection procedures during the Coronavirus Disease 2019 (COVID-19) outbreak: Experience of the Hubei Cancer Hospital in Wuhan, China," *Radiotherapy and Oncology*, vol. 148, pp. 203–210, 2020.
- [4] C. Sohrabi, Z. Alsafi, N. O'Neill et al., "World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19)," *International Journal of Surgery*, vol. 76, pp. 71–76, 2020.
- [5] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial intelligence (AI) applications for COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 337–339, 2020.
- [6] R. O. Panicker, K. S. Kalmady, J. Rajan, and M. K. Sabu, "Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 3, pp. 691–699, 2018.
- [7] B. Oh and J. Lee, "A case study on scene recognition using an ensemble convolution neural network," in *2018 20th International Conference on Advanced Communication Technology (ICACT)*, Chuncheon, Korea (South), 2018.
- [8] C. A. Trimbak, "Analysis of image processing for digital X-ray," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 5, 2016.
- [9] H. Zhu, W. Sun, M. Wu, G. Guan, and Y. Guan, "Pre-processing of X-ray medical image based on improved temporal recursive self-adaptive filter," in *The 9th International Conference for Young Computer Scientists*, pp. 758–763, Hunan, China, 2008.
- [10] V. Kajla, A. Gupta, and A. Khatak, "Analysis of X-Ray Images with Image Processing Techniques: A Review," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India, 2018.
- [11] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia detection using CNN based feature extraction," in *IEEE International Conference on Electrical, Computer and Communication Technologies*, pp. 1–7, Coimbatore, India, 2019.
- [12] K. Hammoudi, H. Benhabiles, M. Melkemi et al., "Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19," 2020, <https://arxiv.org/abs/2004.03399>.
- [13] S. Rajaraman and S. Antani, "Weakly labeled data augmentation for deep learning: a study on COVID-19 detection in chest X-rays," *Diagnostics*, vol. 10, no. 6, 2020.
- [14] V. Chouhan, S. K. Singh, A. Khamparia et al., "A novel transfer learning based approach for pneumonia detection in chest X-ray images," *Applied Sciences*, vol. 10, no. 2, 2020.
- [15] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Computers in Biology and Medicine*, vol. 121, article 103792, 2020.
- [16] G. Jain, D. Mittal, D. Thakur, and M. K. Mittal, "A deep learning approach to detect Covid-19 coronavirus with X-ray images," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1391–1405, 2020.
- [17] J. Civit-Masot, F. Luna-Perejón, M. Domínguez Morales, and A. Civit, "Deep learning system for COVID-19 diagnosis aid using X-ray pulmonary images," *Applied Sciences*, vol. 10, no. 13, 2020.
- [18] T. B. Chandra, K. Verma, D. K. Singh, D. Jain, and S. S. Netam, "Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble," *Expert Systems with Applications*, vol. 165, article 113909, 2021.
- [19] J. Cohen, L. Dao, K. Roth et al., "Predicting COVID-19 Pneumonia Severity on Chest X-ray With Deep Learning," *Cureus*, vol. 12, no. 7, 2020.
- [20] A. Sharma, S. Rani, and D. Gupta, "Artificial intelligence-based classification of chest X-ray images into COVID-19 and other infectious diseases," *International Journal of Biomedical Imaging*, vol. 2020, Article ID 8889023, 10 pages, 2020.
- [21] A. S. Al-Waisy, S. Al-Fahdawi, M. A. Mohammed et al., "COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images," *Soft Computing*, pp. 1–16, 2020.
- [22] P. M. Kumar, R. Saravanakumar, A. Karthick, and V. Mohanavel, "Artificial neural network-based output power prediction of grid-connected semitransparent photovoltaic

- system,” *Environmental Science and Pollution Research*, pp. 1–10, 2021.
- [23] M. A. Mohammed, K. H. Abdulkareem, A. S. al-Waisy et al., “Benchmarking methodology for selection of optimal COVID-19 diagnostic model based on entropy and TOPSIS methods,” *IEEE Access*, vol. 8, pp. 99115–99131, 2020.
- [24] V. Chandran, M. G. Sumithra, A. Karthick et al., “Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images,” *BioMed Research International*, vol. 2021, Article ID 5584004, 15 pages, 2021.
- [25] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa et al., “Voice pathology detection and classification using convolutional neural network model,” *Applied Sciences*, vol. 10, no. 11, 2020.
- [26] R. Kabilan, V. Chandran, J. Yogapriya et al., “Short-term power prediction of building integrated photovoltaic (BIPV) system based on machine learning algorithms,” *International Journal of Photoenergy*, vol. 2021, Article ID 5582418, 11 pages, 2021.
- [27] M. K. Abd Ghani, M. K. Mohammed, M. A. Arunkumar et al., “Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques,” *Neural Computing and Applications*, vol. 32, no. 3, pp. 625–638, 2020.
- [28] N. Y. Jayalakshmi, R. Shankar, U. Subramaniam et al., “Novel multi-time scale deep learning algorithm for solar irradiance forecasting,” *Energies*, vol. 14, no. 9, p. 2404, 2021.
- [29] O. I. Obaid, M. A. Mohammed, M. K. A. Ghani, A. Mostafa, and F. Taha, “Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer,” *International Journal of Engineering & Technology*, vol. 7, pp. 160–166, 2018.

Research Article

Combined Feedback Feedforward Control of a 3-Link Musculoskeletal System Based on the Iterative Training Method

Amin Valizadeh  and Ali Akbar Akbari 

Department of Mechanical Engineering, Ferdowsi University of Mashhad, Iran

Correspondence should be addressed to Ali Akbar Akbari; akbari@um.ac.ir

Received 9 April 2021; Revised 9 September 2021; Accepted 25 September 2021; Published 8 November 2021

Academic Editor: Qiushi Zhao

Copyright © 2021 Amin Valizadeh and Ali Akbar Akbari. 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 investigation and study of the limbs, especially the human arm, have inspired a wide range of humanoid robots, such as movement and muscle redundancy, as a human motor system. One of the main issues related to musculoskeletal systems is the joint redundancy that causes no unique answer for each angle in return for an arm's end effector's arbitrary trajectory. As a result, there are many architectures like the torques applied to the joints. In this study, an iterative learning controller was applied to control the 3-link musculoskeletal system's motion with 6 muscles. In this controller, the robot's task space was assumed as the feedforward of the controller and muscle space as the controller feedback. In both task and muscle spaces, some noises cause the system to be unstable, so a forgetting factor was used to a convergence task space output in the neighborhood of the desired trajectories. The results show that the controller performance has improved gradually by iterating the learning steps, and the error rate has decreased so that the trajectory passed by the end effector has practically matched the desired trajectory after 1000 iterations.

1. Introduction

The reaching movement is accounted for a huge part of hand movements. In all these activities, a swift and complex process occurs in the brain, and after processing, the generated control signals are transmitted to body motors, namely, muscles. This complex process in the brain comprises some levels. First, the desired trajectory is determined for reaching an object, and in the second step, the coordinates of the specified trajectory estimated by vision are converted into the body coordinates; in the last step, control commands are sent to the muscular system to go through the desired trajectory. Investigating the body's musculoskeletal system's control mechanism can lead us to develop a robust control technique that can be applied to rehabilitation robotics. The design process and application of the actuators in such robots are similar to the simulation of the human body's neural control system. Many controllers have been introduced and employed to control such systems and produce motions similar to the human movement, which are of different design methods and performances based on their

design space (robot task space, joint space, and muscle space) [1, 2]. Each of these spaces has its features and complexity, and as we move from the task space toward the muscle space, it will be difficult to design the controller because of the increasing space order. It should be noted that the controller design in the muscle space should be carried out carefully so that the forces out of the body are ignored, and the model approaches reality. The joint redundancy causes each angle to have no unique solution in return for an arbitrary trajectory of the arm's end effector [3]. Another problem caused by redundancy is the lack of a unique solution for model forces to generate unique torque [4]. Many optimization techniques have been proposed to overcome this problem in classification [5], biology and robotics [6–9].

On the other hand, in everyday life, we can easily carry out the most complex movements with the highest possible accuracy in the presence of the same redundancies with the least possible. One solution to overcome this complexity is that the central nervous system activates a small group of muscles—called muscle synergy [10]—which allows the control of body movements with less computational cost by

reducing the required independent degrees of freedom. In recent years, Suetani and Morimoto [11] presented an innovative hypothesis under the title of “Virtual Spring-Damper Hypothesis,” where there is no need to optimize the redundancy criterion to overcome the redundancy problem. However, the previous problem is solved by applying this hypothesis, but due to the redundancy of muscles, we will deal with other problems that require the application of the muscle nonlinear models. In 2013, aiming to find the synergies of reaching and balancing movements on the musculoskeletal system of the hand, Tahara et al. [12] conducted a research study to investigate muscular integrity force data and the data of body receptors (proprioception and vision). The canonical correlation analysis (CCA) method, which follows the natural behavior of the body, was used to obtain the relation between the data related to muscles and body receptors. In a real system, the time delay and noise should be considered in the body’s actuator and sensor systems. Hence, this study is aimed at examining the effects of time delay and noise on determining the synergies of the hand’s musculoskeletal system. The results of this study not only can be used to understand the biological data of the motor control system but also can be applied as an artificial controller for a high-DOF robot. In this study, the motion equations for the 3-link musculoskeletal system of the human arm and the iterative learning controller are presented in Section 2. The results obtained from the simulation of the ILC (iterative learning control) with the neuro-fuzzy controller are discussed in Section 4. In Section 6, the remarks concluded from this study are described.

2. DOF Human Musculoskeletal Arm Model

The 3-DOF human musculoskeletal arm model used in this study consisting of three solid links and six monoarticular muscles is shown in Figure 1. Since this arm moves on the horizontal plane, the effect of the gravity force can be ignored. As shown in Figure 1, this model consists of six muscles that can only apply tensile forces so that each joint moves by some of these related muscles. Muscles are assumed to be without weight and designed based on the Hill model, which are directly connected to links as

$$\begin{aligned}
 f_m &= P\bar{\alpha} - P\{A(\bar{\alpha})C + C_0\}\dot{l}, \\
 \bar{\alpha} &= (\bar{\alpha}_1, \bar{\alpha}_2, \dots, \bar{\alpha}_6) \in R^6, \\
 P &= \text{diag}(p_1, p_2, \dots, p_6) \in R^{6 \times 6}, \\
 A(\bar{\alpha}) &= \text{diag}(\bar{\alpha}_1, \bar{\alpha}_2, \dots, \bar{\alpha}_6) \in R^{6 \times 6}, \\
 C &= \text{diag}(c_1, c_2, \dots, c_6) \in R^{6 \times 6}, \\
 C_0 &= \text{diag}(c_{01}, c_{02}, \dots, c_{06}) \in R^{6 \times 6}.
 \end{aligned} \tag{1}$$

From Ref. [13], parameter f_m is the muscles’ contractile force, which is the nonlinear function of the muscle’s contractile velocity and the control input produced in the central nervous system.

Considering L_1 , L_2 , and L_3 to be the first, second, and third links, respectively, as well as their relation angle with

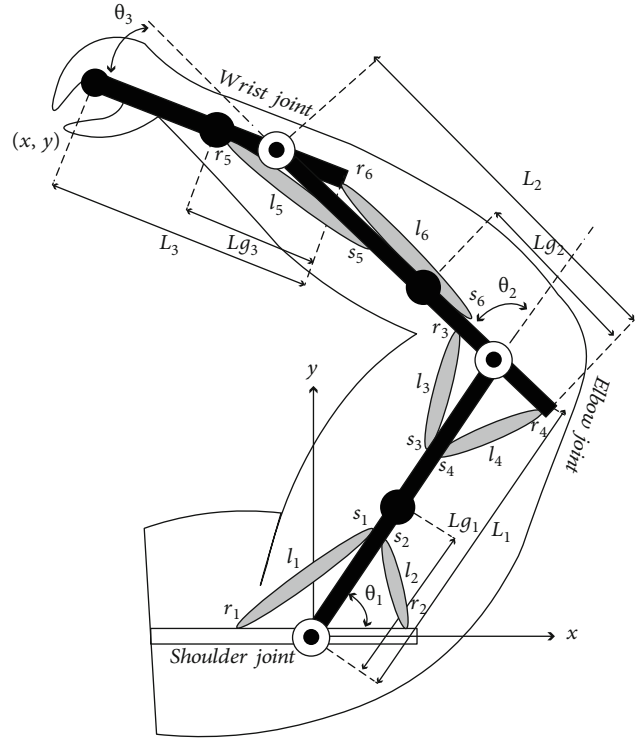


FIGURE 1: Schematic view of the 3-DOF musculoskeletal model for the hand.

respect to the x -axis, first link, and second link as θ_1 , θ_2 , and θ_3 , respectively, the arm’s end effector position to the joint angles is expressed by the following equation:

$$X = J(\theta) = \begin{bmatrix} L_1 \cos(\theta_1) + L_2 \cos(\theta_1 + \theta_2) + L_3 \cos(\theta_1 + \theta_2 + \theta_3) \\ L_1 \sin(\theta_1) + L_2 \sin(\theta_1 + \theta_2) + L_3 \sin(\theta_1 + \theta_2 + \theta_3) \end{bmatrix} \in R^2. \tag{2}$$

Defining length vectors of the muscles as $l = [l_1 \ l_2 \ l_3 \ l_4 \ l_5 \ l_6]^T$ results in

$$l = W(\theta) = \begin{bmatrix} r_1 \left(\pi - \theta_1 - a \cos\left(\frac{r_1}{s_1}\right) \right) + \sqrt{s_1^2 - r_1^2} \\ r_2 \left(\frac{\pi}{2} + \theta_1 - a \cos\left(\frac{r_2}{s_2}\right) \right) + \sqrt{s_2^2 - r_2^2} \\ r_3 \left(\pi - \theta_2 - a \cos\left(\frac{r_3}{s_3}\right) \right) + \sqrt{s_3^2 - r_3^2} \\ r_4 \left(\frac{\pi}{2} + \theta_2 - a \cos\left(\frac{r_4}{s_4}\right) \right) + \sqrt{s_4^2 - r_4^2} \\ r_5 \left(\frac{\pi}{2} + \theta_3 - a \cos\left(\frac{r_5}{s_5}\right) \right) + \sqrt{s_5^2 - r_5^2} \\ r_6 \left(\frac{\pi}{2} + \theta_3 - a \cos\left(\frac{r_6}{s_6}\right) \right) + \sqrt{s_6^2 - r_6^2} \end{bmatrix} \in R^6, \tag{3}$$

where r_{1-6} and s_{1-6} represent the torque levers, as shown in Figure 1. The following equations are obtained by taking the time derivative of equations (2) and (3) with respect to time:

$$\dot{X} = J\dot{\theta}, \quad (4)$$

$$\dot{l} = W^T\dot{\theta}. \quad (5)$$

$\dot{X} = [\dot{x} \ \dot{y}]^T$ is the end effector velocity of the arm, $\dot{\theta} = [\dot{\theta}_1 \ \dot{\theta}_2 \ \dot{\theta}_3]^T$ is the angular velocity of joints, and $\dot{l} = [\dot{l}_1 \ \dot{l}_2 \ \dot{l}_3 \ \dot{l}_4 \ \dot{l}_5 \ \dot{l}_6]^T$ represents the stretch rate of muscles. Also, $J \in R^{2 \times 3}$ is the Jacobian matrix that shows the relation between the linear velocities of the arm's end effector and angular velocities while $W^T \in R^{6 \times 3}$ is the Jacobian matrix, which relates the contractile rate of muscles to the angular velocity of the joints as

$$J = \begin{bmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \end{bmatrix},$$

$$J_{11} = -L_1 \sin(\theta_1) - L_2 \sin(\theta_1 + \theta_2) - L_3 \sin(\theta_1 + \theta_2 + \theta_3),$$

$$J_{12} = -L_2 \sin(\theta_1 + \theta_2) - L_3 \sin(\theta_1 + \theta_2 + \theta_3),$$

$$J_{13} = -L_2 \sin(\theta_1 + \theta_2) - L_3 \sin(\theta_1 + \theta_2 + \theta_3),$$

$$J_{21} = L_1 \cos(\theta_1) + L_2 \cos(\theta_1 + \theta_2) + L_3 \cos(\theta_1 + \theta_2 + \theta_3),$$

$$J_{22} = L_2 \cos(\theta_1 + \theta_2) + L_3 \cos(\theta_1 + \theta_2 + \theta_3),$$

$$J_{23} = L_3 \cos(\theta_1 + \theta_2 + \theta_3),$$

$$W = \begin{bmatrix} -r_1 & r_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -r_3 & r_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & -r_5 & r_6 \end{bmatrix}. \quad (6)$$

By assuming J as a full-rank matrix, the inverse of equations (2) and (4) is obtained as follows:

$$\theta = G_x^{-1}(x) \in R^2, \quad (7)$$

$$\dot{\theta} = J^{-1}\dot{x} \in R^2. \quad (8)$$

$G_x^{-1}(x)$ represents a vector with nonlinear functions, which shows the inverse kinematics from the task space to joint space. Also, J^{-1} shows the inverse kinematics from the task space velocity to the joint's angular velocity. By substituting equation (8) into equations (3) and (5), we can state that

$$l = G_i(G_x^{-1}(x)) \in R^6, \quad (9)$$

$$\dot{l} = W^T J^{-1}\dot{x} \in R^6. \quad (10)$$

Equation (9) demonstrates the inverse kinematics from the task space to the muscle space, which is applied to the controller's feedforward behavior.

By applying the principle of virtual work, the work done by muscle torque is defined as follows:

$$T = W f_m \in R^2, \quad (11)$$

where $f_m = [f_1 \ f_2 \ f_3 \ f_4 \ f_5 \ f_6]^T$ is the vector representing the tensile forces of muscles and $T = [T_1 \ T_2 \ T_3]^T$ is the joint torque vector.

By assuming that $W \in R^{3 \times 6}$ is a row full-rank matrix during movement, the inverse of equation (11) is expressed as follows:

$$f_m = W^+ T \in R^6, \quad (12)$$

$$W^+ = W^T (W W^T)^{-1} \in R^{6 \times 3}. \quad (13)$$

Besides, the static relation between T and the output vector of forces applied to the arm's endpoint in the space $F \in R^2$ is expressed as follows:

$$T = J^T F \in R^3. \quad (14)$$

By substituting equation (14) into equation (13), it is concluded that

$$f_m = W^+ J^T F \in R^6. \quad (15)$$

Equation (15) demonstrates the static inverse relation between f_m and F .

3. Iterative Learning Control

An ILC strategy of the PI type has been introduced in Reference [14] to trace an arbitrary time-dependent trajectory using the robotic arm model. The errors related to the position and velocity in a test are stored to be tuned for the next test by an input correction strategy. The data stored in the first step are multiplied by a factor and added to the input in the next test. Implementing a simple task space feedback control for a 2-DOF arm is considered by Tahara et al. to address the muscle space redundancy problem on the contractile output force [15]. They also studied multiple space variables to enhance the robustness of the 2-DOF arm exposed to sensory noises. Despite the nonlinear equations of the human arm's motion, the suggested method sufficiently improves the system's robustness regarding the traditional ILC methods [16]. Therefore, the proposed method is considered in our study. As discussed in the previous section, to compensate for the iterative learning controller's input, there are three representatives of the state space, namely, muscle space, joint space, and task space. Therefore, any space that can better compensate for the control input is of great importance in achieving the desired performance. Furthermore, it should be noted that many noises cause damage to sensory information, and its huge impact on the movement of the musculoskeletal system is inevitable. Therefore, the system's robustness to deal with the noise varies depending on the space in which the system is modeled.

A new control strategy based on iterative learning, which uses the sum of state-space variables, is employed to improve the robustness of the system against noise. In the present paper, a case study is performed by considering the task space and muscle space as the spaces for feedback and feedforward behaviors, respectively. The control input of the muscles in the i th test is defined as follows:

$$u_i = -W_i^+ J_i^T (K_p \Delta x_i - K_v \Delta \dot{x}_i) + v_i, \quad (16)$$

where index i represents the test number, $K_p = \text{diag} [k_{p1}, k_{p2}] \in R^{2 \times 2} > 0$ and $K_v = \text{diag} [k_{v1}, k_{v2}] \in R^{2 \times 2} > 0$ are the feedback coefficients of position and velocity in the task space, and v_i is the feedforward parameter obtained from the iterative learning process. The error of position and velocity is defined as $\Delta x_i = x_i - x_d \in R^2$ and $\Delta \dot{x}_i = \dot{x}_i - \dot{x}_d \in R^2$, respectively; x_d and \dot{x}_d also represent the end effector's position and velocity, respectively. The feedforward parameter, $v_i \in R^6$, is not designed in the task space similar to feedback behavior, but it is modeled in the muscle space and updated as follows:

$$v_i = \begin{cases} 0, & i = 1, \\ (1 - \beta)v_{i-1} - (\Phi \Delta l_{i-1} + \Psi \Delta i_{i-1}), & i > 1, \end{cases} \quad (17)$$

where $\Phi = \text{diag} [\phi_1, \phi_2, \dots, \phi_6] \in R^{6 \times 6} > 0$ and $\Psi = \text{diag} [\psi_1, \psi_2, \dots, \psi_6] \in R^{6 \times 6} > 0$ are the coefficient matrices of position and velocity, respectively; besides, the position error is defined as $\Delta l_i = l_i - l_d$ and the velocity errors in the muscle space are expressed as $\Delta i_i = i_i - i_d$. $l_d \in R^6$ and $i_d \in R^6$ are the length of muscles and their contraction rate relative to the position and velocity of the end effector, respectively. These parameters are obtained by calculating the inverse dynamic as

$$\begin{aligned} \Delta l_i &= G_i(G_x^{-1}(x_i)) - G_i(G_x^{-1}(x_d)), \\ \Delta i_i &= W_i^T J_i^{-1} \Delta \dot{x}_i. \end{aligned} \quad (18)$$

In this study, the Gaussian noise is used as a noise which is applied to sensory information. An error in the initial conditions of two consecutive tests and dynamic oscillations due to different types of noises causes the general system to be unstable using the iterative learning controller. Therefore, to overcome these noises, Suetani and Morimoto [11] introduced a forgetting factor to update the iterative learning controller. Using this forgetting factor ensures that the final converged trajectory after good learning is in the desired trajectory neighborhood. In equation (17), β is the forgetting factor that has to satisfy the condition of $0 < \beta < 1$. It is assumed that the muscle's length and end effector position and velocity signals include Gaussian noise individually. Due to Refs. [15, 17], the magnitude of the noise existing in the end effector's position and velocity is 4% of real data, and the magnitude of noise existing in the length of the muscle and its contraction rate is 50% of the real data. This is because the data related to the end effector's position and velocity are obtained through observation, which is relatively accurate. However, the data related to the muscle's length and contraction rate are received through

TABLE 1: Numerical parameters of the model.

	Length (m)	Mass (kg)	Inertial moment (kg.m ²)	CoM position (m)
1 st link	0.31	1.93	0.0141	0.165
2 nd link	0.27	1.32	0.0120	0.135
3 rd link	0.15	0.35	0.0010	0.075

TABLE 2: Geometric parameters of the muscles.

Muscle	Value (m)	
l_1	$r_1 = 0.055$	$s_1 = 0.080$
l_2	$r_2 = 0.055$	$s_2 = 0.080$
l_3	$r_3 = 0.030$	$s_3 = 0.120$
l_4	$r_4 = 0.030$	$s_4 = 0.120$
l_5	$r_5 = 0.035$	$s_5 = 0.220$
l_6	$r_6 = 0.040$	$s_6 = 0.250$

TABLE 3: The parameters associated with the controller.

Parameter	Value
Feedback gain	$K_p = [8080]$
Feedback gain	$K_v = [5050]$
Learning gain	$\Phi_1 = \dots = \Phi_6 = 250$
Learning gain	$\Psi_1 = \dots = \Psi_6 = 140$
Forgetting factor	$\beta = 0.3$

the muscular bulk which has large electrical noise leading to inaccurate results [18].

4. Results

The simulation results are presented in this section. Tables 1 and 2 demonstrate the numerical values associated with the 3-link model and the values related to the muscles' physical properties, respectively. Also, the coefficients of the controller are listed in Table 3.

The controller is aimed at tracing a semicircular trajectory. Therefore, we consider the following trajectory:

$$\begin{cases} x = 0.2 + 0.1 \cos(t), \\ y = 0.55 + 0.1 \sin(t). \end{cases} \quad (19)$$

The simulation's total time is assumed to be $T = \pi s$, and the hand is initially located at point $(-0.1, 0.55)$. Therefore, during the aforementioned period, the robot is expected to cover the semi-semicircular trajectory fully. For evaluating the robustness of the presented model against uncertainties, the simulation parameters have been changed by 5%. To compare the controller's performance with similar counterparts, the model control results are compared to the neuro-fuzzy control method presented in our other paper [19]. In the cited article, the similar given trajectory was precisely

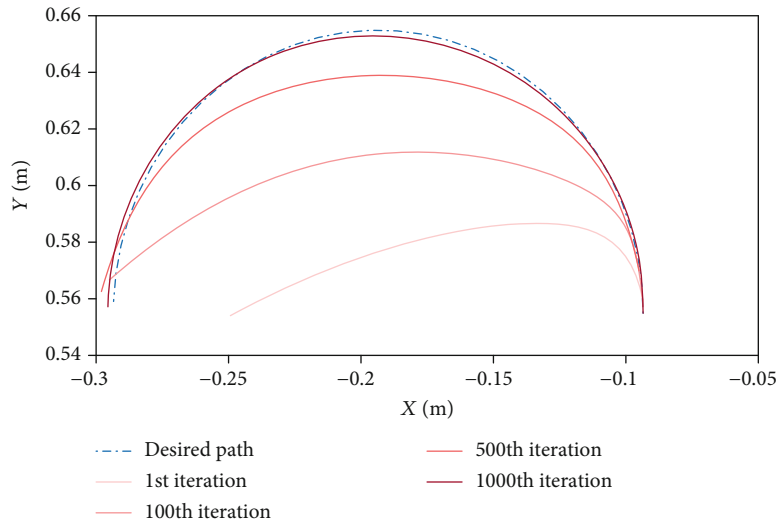


FIGURE 2: The trajectories passed by the model per 1000 iterations.

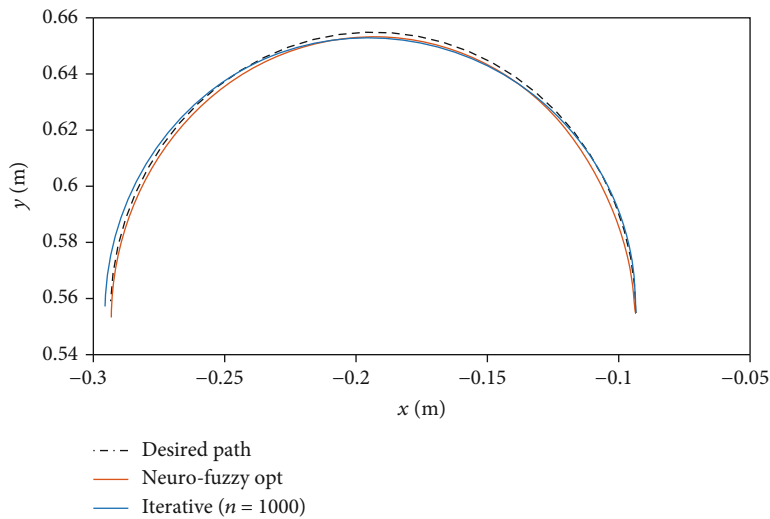


FIGURE 3: Comparison of trajectories covered by the neuro-fuzzy adaptive controller and ILC.

followed by the muscle optimization, and the results concluded appropriate compliances with the hand's natural motion. The model simulation was performed in MATLAB version 2021a running on an Intel Core i7 (2.8GHz and 16 Gb RAM). For all simulations, the variable-step MATLAB ODEs solver ode45 with relative solver tolerance 1×10^{-4} was implemented, which took 4.817 s for the ILC controller compared to 10.045 s for the neuro-fuzzy controller.

Figure 2 depicts different trajectories that the controller has taken over 1000 iterations to reach the desired trajectory. As can be seen, as the controller's performance is improved, the error is reduced gradually. Therefore, in iteration no. 1000, the trajectory is adjusted to the desired trajectory. Such a process is similar to learning and muscle memory that can perfectly go through a trajectory with practice and repetition. The trajectories that both controllers have gone through at the same time are shown in Figure 3. The results

show that the proposed controller has better performance. In other words, if we exceed 1000 iterations in training the controller, we will observe a further improvement in the controller results. However, it should be noted that the simulation time increases with increasing the number of iterations.

Figure 4 displays the displacement of different joints during the movement scenario. The displacement of joints is similar to another. The adaptive controller performance is based on the optimization of the cost function and the iterative controller performance on learning; hence, Figure 4 designates that the proposed controller performance is acceptable compared to the adaptive controller's performance using the neuro-fuzzy adaptive controller.

Finally, the magnitude of forces applied to each muscle during the desired trajectory is illustrated in Figure 5. The neuro-fuzzy controller uses muscle optimization; therefore, its force diagrams are much more ideal. On the other hand,

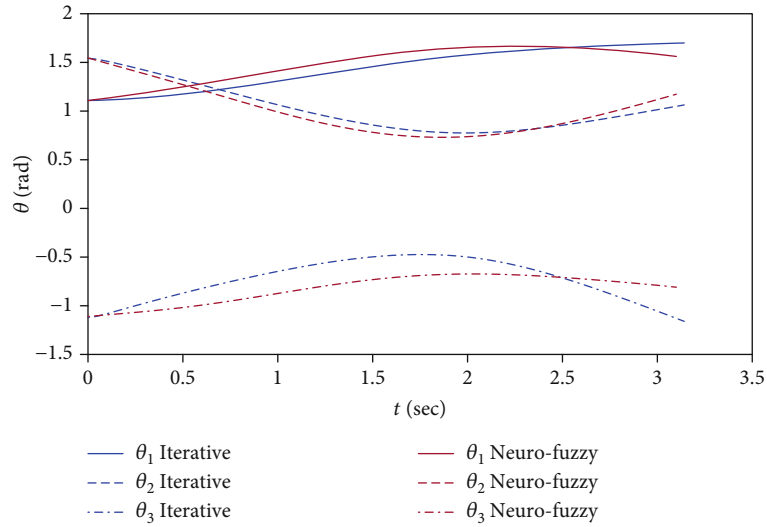


FIGURE 4: Comparison of joint displacement between the neuro-fuzzy adaptive controller and ILC.

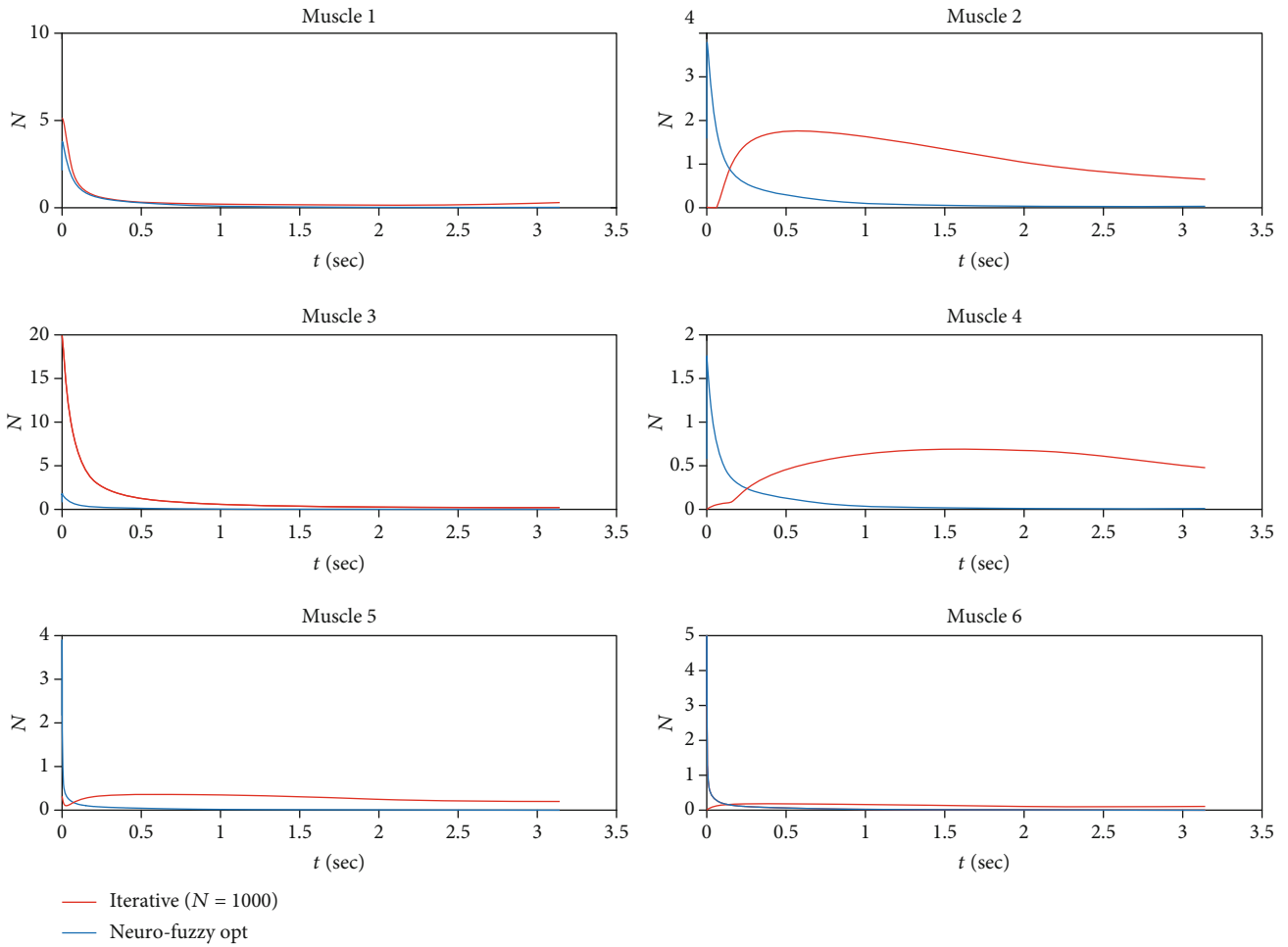


FIGURE 5: Comparison of the generated forces using the neuro-fuzzy adaptive controller and ILC.

the ILC controller has a more smooth bell-shaped profile similar to the agonist-antagonist paired muscles involved in the natural movements of the human body. However, both con-

trollers have similar patterns. In addition, muscle forces in both methods are in the adequate range for the human body, where the ILC method has almost fewer force values.

5. Discussion

The controller has two pieces, the first of which is feedback input comprising task space variables, while the other part is the feedforward input, which is made up of muscle space parameters gained through the iterative learning algorithm. Although all controller gains are tuned in all iterations, the simulation results demonstrated that the hand endpoint in the first iteration is significantly different from the 1000th iteration. Despite the nonlinear equations of muscles, the controller could pass through the desired trajectory after the 1000th iteration. In addition, the path tracking error has considerably been mitigated by increasing the repetition number. The use of different variable spaces in conjunction with the learning algorithm was the primary reason for the sufficient accuracy of path tracking in the proposed controller.

Furthermore, the travel time of the simulation was reduced to half using the ILC controller, compared to the neuro-fuzzy one following the 1000th repetition of the desired path. Force values for the given path were also in the sufficient force ranges of the human hand muscles. In most muscles, the endpoint passed through the given trajectory with a much lower force than the neuro-fuzzy controller. These results characterize the efficiency of this controller for musculoskeletal modeling in the human body. As a future study, we intend to conduct movement trials in actual and uncontrolled conditions using EMG signals plus effective technologies such as user-friendly contactless path recognition to increase the method's productivity [5].

6. Conclusion

The controller's performance was improved by iterating learning, and subsequently, the related error was reduced so that the final trajectory that has gone through simulation is practically adjusted to the desired trajectory. Such a process is similar to learning and muscle memory that can lead to perfectly going through a trajectory with practice and repetition. The quantitative comparison between the iterative learning controller and neuro-fuzzy controller results suggested that the proposed controller has a better performance. In other words, if we exceed 1000 iterations in training the controller, we will observe a further improvement in the controller results. However, it should be noted that the time required for solving the problem increases by increasing the number of iterations. By comparing the forces generated in the muscles for both controllers, it was observed that the maximum value of these forces for the current controller was less than that of the adaptive controller, although the average of generated force is higher for the current controller. Considering that the muscle forces' optimization is one of the design indicators in adaptive controllers, it was not considered in the proposed method. Here, it was important that the controller can successfully guide the model on the desired trajectory in the presence of system uncertainties, and the forces applied to the muscles are in the desired range.

Data Availability

The data is extracted from our other paper entitled "The Optimal Adaptive-Based Neuro-Fuzzy Control of the 3-DOF Musculoskeletal System of Human Arm in A 2D Plane."

Conflicts of Interest

The authors declare that they have no conflicts of interest in relation to this research and its publication.

References

- [1] W. He, C. Xue, X. Yu, Z. Li, and C. Yang, "Admittance-based controller design for physical human-robot interaction in the constrained task space," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 4, pp. 1937–1949, 2020.
- [2] J. Chen and H. Qiao, "Muscle-synergies-based neuromuscular control for motion learning and generalization of a musculoskeletal system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3993–4006, 2020.
- [3] J. Shi, Y. Mao, P. Li et al., "Hybrid mutation fruit fly optimization algorithm for solving the inverse kinematics of a redundant robot manipulator," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6315675, 2020.
- [4] K. Stollenmaier, W. Ilg, and D. F. Haeufle, "Predicting perturbed human arm movements in a neuro-musculoskeletal model to investigate the muscular force response," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 308, 2020.
- [5] W. Qiao, M. Khishe, and S. Ravakhah, "Underwater targets classification using local wavelet acoustic pattern and multi-layer perceptron neural network optimized by modified whale optimization algorithm," *Ocean Engineering*, vol. 219, article 108415, 2021.
- [6] Y. Xu, H. Chen, J. Luo, Q. Zhang, S. Jiao, and X. Zhang, "Enhanced moth-flame optimizer with mutation strategy for global optimization," *Information Sciences*, vol. 492, pp. 181–203, 2019.
- [7] L. Shen, H. Chen, Z. Yu et al., "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Systems*, vol. 96, pp. 61–75, 2016.
- [8] J. Hu, H. Chen, A. A. Heidari et al., "Orthogonal learning covariance matrix for defects of grey wolf optimizer: insights, balance, diversity, and feature selection," *Knowledge-Based Systems*, vol. 213, article 106684, 2021.
- [9] A. Sharifi, "On primitives in motor control," *Motor Control*, vol. 24, no. 2, pp. 318–346, 2020.
- [10] S. Arimoto and M. Sekimoto, "Human-like movements of robotic arms with redundant DOFs: virtual spring-damper hypothesis to tackle the Bernstein problem.," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*, pp. 1860–1866, IEEE, Orlando, FL, USA, 2006, May.
- [11] H. Suetani and J. Morimoto, "Canonical correlation analysis for muscle synergies organized by sensory-motor interactions in musculoskeletal arm movements," in *2013 IEEE International Conference on Robotics and Automation*, pp. 2606–2611, IEEE, Karlsruhe, Germany, 2013.
- [12] K. Tahara, S. Arimoto, M. Sekimoto, and Z. W. Luo, "On control of reaching movements for musculo-skeletal redundant

- arm model,” *Applied Bionics and Biomechanics*, vol. 6, no. 1, 26 pages, 2009.
- [13] S. Kawamura, F. Miyazaki, and S. Arimoto, “Realization of robot motion based on a learning method,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 126–134, 1988.
- [14] K. Tahara, Z. W. Luo, S. Arimoto, and H. Kino, “Task-space feedback control for a two-link arm driven by six muscles with variable damping and elastic properties,” in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pp. 223–228, Barcelona, Spain, 2005.
- [15] K. Tahara, Y. Kuboyama, and R. Kurazume, “Iterative learning control for a musculoskeletal arm: utilizing multiple space variables to improve the robustness,” in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4620–4625, Vilamoura-Algarve, Portugal, 2012.
- [16] A. Valizadeh and A. Akbari, “The Optimal Adaptive-Based Neurofuzzy Control of the 3-DOF Musculoskeletal System of Human Arm in a 2D Plane,” *Applied Bionics and Biomechanics*, vol. 2021, Article ID 5514693, 2021.
- [17] Q. Zhao, X. Wu, and W. Bu, “Contactless palmprint verification based on SIFT and iterative RANSAC,” in *2013 IEEE International Conference on Image Processing*, pp. 4186–4189, Melbourne, VIC, Australia, 2013.
- [18] R. VanRullen and C. Koch, “Visual selective behavior can be triggered by a feed-forward process,” *Journal of Cognitive Neuroscience*, vol. 15, no. 2, pp. 209–217, 2003.
- [19] H. Dong and N. Mavridis, “Adaptive biarticular muscle force control for humanoid robot arms,” in *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, pp. 284–290, IEEE, Osaka, Japan, 2012.

Review Article

A Review of Methods of Diagnosis and Complexity Analysis of Alzheimer's Disease Using EEG Signals

Mahshad Ouchani ¹, Shahriar Gharibzadeh ¹, Mahdieh Jamshidi ¹,
and Morteza Amini ^{2,3}

¹*Institute for Cognitive and Brain Sciences, Shahid Beheshti University, Tehran, Iran*

²*Shahid Beheshti University, Tehran, Iran*

³*Institute for Cognitive Science Studies (ICSS), Tehran, Iran*

Correspondence should be addressed to Morteza Amini; mor_amini@sbu.ac.ir

Received 21 April 2021; Revised 20 June 2021; Accepted 18 October 2021; Published 27 October 2021

Academic Editor: Qiushi Zhao

Copyright © 2021 Mahshad Ouchani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study will concentrate on recent research on EEG signals for Alzheimer's diagnosis, identifying and comparing key steps of EEG-based Alzheimer's disease (AD) detection, such as EEG signal acquisition, preprocessing function extraction, and classification methods. Furthermore, highlighting general approaches, variations, and agreement in the use of EEG identified shortcomings and guidelines for multiple experimental stages ranging from demographic characteristics to outcomes monitoring for future research. Two main targets have been defined based on the article's purpose: (1) discriminative (or detection), i.e., look for differences in EEG-based features across groups, such as MCI, moderate Alzheimer's disease, extreme Alzheimer's disease, other forms of dementia, and stable normal elderly controls; and (2) progression determination, i.e., look for correlations between EEG-based features and clinical markers linked to MCI-to-AD conversion and Alzheimer's disease intensity progression. Limitations mentioned in the reviewed papers were also gathered and explored in this study, with the goal of gaining a better understanding of the problems that need to be addressed in order to advance the use of EEG in Alzheimer's disease science.

1. Introduction

Alzheimer's disease (AD) is a neurological disease and is also the most common form of age-related dementia in today's culture. In 2018, it was reported that 50 million people globally have Alzheimer's disease. In 2030, this figure will be around 82 million, and in 2050, it will be around 152 million [1]. In recent decades, there has been a growing focus on using advanced electroencephalography (EEG) signal processing to predict or differentiate Alzheimer's disease. Neuroimaging studies have been extensively used to investigate the causes of AD and to increase the accuracy of AD diagnosis [2]. Since the brain is such a complex structure with complex nonlinear dynamics, complexity studies utilizing data from brain imaging such as EEG, magneto-encephalograms, and functional magnetic resonance imaging (fMRI) are becoming more common fMRI [3]. In several experi-

ments, brain impulses from just one channel, such as an electrode in EEG, a channel in magneto-encephalograms, or a voxel in fMRI, were studied. The brain complexity waves have recently been utilized to help explain the complexity of AD disorders [4]. Sufficient research into brain imaging modalities may help to describe the pathways underlying AD and to provide valuable evidence for the diagnosis. Recently, some research has shown that degrees of difficulty may be used as biomarkers in the early detection of AD. There is currently no systematic study that reviews the various modulation techniques and discusses the complexity of AD brain disorders. Optimizing EEG analysis is important for designing low-cost, noninvasive wearable applications to screen Alzheimer's patients [5]. The choice of critical EEG channels may also aid in the development of new wearable technologies and the optimization of computing resources. Many experiments have looked at multiscale entropy

(MSE) and MSE-based measurements of EEG signals from Alzheimer's patients [6].

MSE represents the degree of healthfulness of a biological process by its production physiologic signals through measuring the complexity of a physiologic time series over various time scales [5]. Many experiments have shown that the MSE values of EEG signals from normal persons are higher on small scales than those from AD patients, but lower on large scales than someone from AD patients [5, 7]. Furthermore, at broad scales, the slope of the MSE vs. size plot was observed to be higher for AD patients than for healthy controls. The EEG is a noninvasive experimental tool that shows how brain synapses work in real time. Quantitative EEG (qEEG) research offers many perspectives on EEG signals, including frequency, dynamic alterations, and source imaging. Various researches have explained that qEEG can diagnose the foregoing anomalies in AD patients [8]: (1) changes in EEG patterns, (2) decreased coordination, (3) diminished sophistication, and (4) neuromodulator defects as potential markers for brain activity assessment. Furthermore, qEEG offers objective and quantifiable data that can be replicated in subsequent trials, as well as the benefits of having less laboratory protocols and lower costs [8]. This makes it ideal for screening large-scale and early detection of AD. For the purpose, EEG has been thoroughly researched as a useful instrument for analyzing AD over the past few decades. Nonetheless, as far as we can see, few of the study findings assist physicians with their daily work or decision-making. The concern is that the EEG signal is noise-sensitive, with nonstationary properties, which makes detection difficult [9]. Furthermore, since there is so much variation between subjects, it is difficult to distinguish objects and patterns from natural brain function. Reliable biomarkers and rigorous diagnostic techniques that can derive valuable knowledge from jumbled EEG signals are also required urgently [9, 10].

For EEG signal processing, the wavelet transform has been suggested as an efficient tool for analyzing time and frequency. It entails convolving the EEG signal with a variable-width time window, and higher frequencies have narrower window widths, whereas lower frequencies have wider window widths. This adjusts well to the features of EEG signals, which are made up of short-duration high-frequency incidents and long-duration low-frequency incidents [11]. EEG pulse time-frequency measurement combined with machine learning (ML) methods could help with diagnosis and understanding of AD. Overfitting could be avoided by using machine learning algorithms like feature selection, which exclude data that is redundant from high-dimensional data [12]. The thesis is aimed at investigating robust functional biomarkers dependent on time-frequency features of qEEG and developing a computer-aided discriminant scheme for automatically classification EEG signals of AD and normal elderly controls (NC) as a result of the promising results obtained with the wavelet transform analysis and machine learning methods [13].

Many experiments have looked at multiscale entropy and MSE-based measurements of EEG signals from Alzheimer's patients. MSE represents the degree of healthfulness

of a living process by its production physiologic signals through measuring the complexity of a physiologic time series over various time scales. Many experiments have shown that the MSE values of EEG signals from healthy controls are higher on small scales than those from AD patients, but lower on large scales than those from AD patients. Furthermore, at broad scales, the slope of the MSE vs. size plot was observed to be higher for AD patients than for healthy controls. Lately, machine learning techniques have been introduced to EEG research in order to enhance the recognition accuracy at AD patients of various severity levels, as well as stable subjects. Any MSE time scale could be used as a function in a machine learning algorithm [14]. For each EEG channel, Fan et al. [15] used 38 features for machine learning, including MSE features and other spectral and temporal features derived from the EEG data. A total of 24 EEG recordings were obtained from stable, mild, and severe AD patients. There were five binary and one ternary classification problem to solve. Fan et al. used 19 EEG channels to remove 380 MSE functionality [15]. Each channel's EEG signals contributed a series of 20 distinct MSE values calculated at scales 1–20. A total of 123 EEG recordings were obtained from stable people, people with very minor AD, people with mild AD, and people with moderate to serious AD.

2. Literature Review

During rest, generalized EEG slowdown has been noticed in a variety of AD researches. This slowdown can be seen visually as a reduction in the dominant baseline rhythm's tempo, or spectrally as a rise in the strength of slow rhythms and a reduction in the power of quicker rhythms [16]. Indeed, in AD, the power spectrum's peak frequency is usually among 8-12 Hz variations to a lower range of 6–8 Hz. Just a few research has looked at EEG shifts in people with frontotemporal dementia. In frontotemporal dementia patients, qualitative examination of EEG recordings normally reveals no irregular slowing [17]. We would like to find out that pathological EEG slowdown is a more serious version of the general slowdown of the frame rhythm that occurs with healthy aging. As a result, age-matched control groups are needed in these studies; otherwise, the EEG-slowness effect would be exaggerated. A visual grand complete EEG score and the coordination probability as an indicator of functional connectivity were used to equate mild to moderate frontotemporal dementia and patients of Alzheimer's disease to healthy controls (HC) [18]. The complete EEG score in visual form did not vary significantly between frontotemporal dementia and HC. Using the visual grand total EEG, patients of Alzheimer's disease display substantial EEG slowing and lack of reactivity as compared to frontotemporal dementia and HC patients [19]. In high rates of frequency, AD patients have a lower chance of synchronization than frontotemporal dementia and HC patients, but there are no variations between frontotemporal dementia and HC patients (Figure 1). As a result, shifts in synchronization are likely to follow the slowing trend. Higher frequency

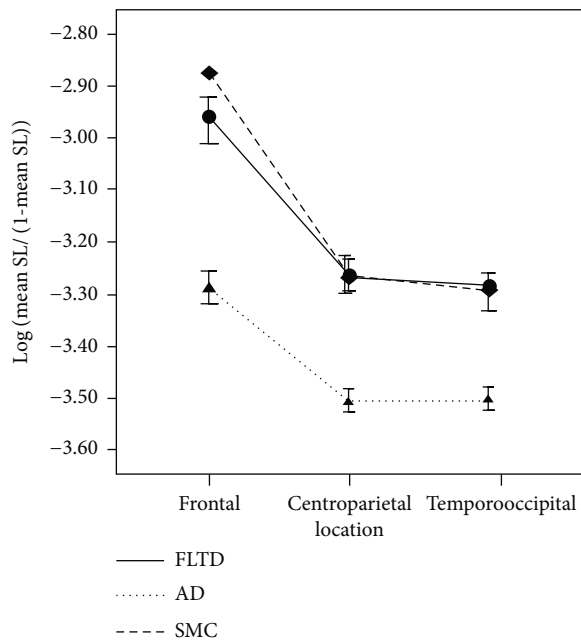


FIGURE 1: Electrode interaction effects caused significant group multiplication in the 8–10 Hz frequency range [20].

features, such as strength and synchronization, are diminished in AD but not in frontotemporal dementia.

The qEEG variations are between people with frontal lobe dementia and others with Alzheimer's disease, Parkinson's disease dementia [21]. Lewy body disease has been studied in some trials. The global field power for six frequency bands was measured for the qEEG: δ (1 to 3.5 Hz), θ (4 to 7.5 Hz), α (8 to 11 Hz), β_1 (12 to 15.5 Hz), β_2 (16 to 19.5 Hz), and β_3 (20 to 23.5 Hz). The number of quick frequency bands was used to measure the spectral ratio $\alpha + \beta_1 + \beta_2 + \beta_3$ and bands of low frequency $\delta + \theta$. Patients with likely frontotemporal dementia were similar to AD patients and healthy controls on cortical EEG sources' spectral profile [22]. The authors of this study used EEG band forces, coherence, dominant frequency, peak frequency, and cortical sources to distinguish sixteen patients with AD from nineteen patients by frontotemporal dementia. The most accurate predictors of frontotemporal dementia and AD were identified in a model using logistic regression analysis. Activities such as elevated levels of visuospatial capacity and episodic memory were among the predictors. The model's classification accuracy was 93.3 percent.

As a result, combining qEEG and neuropsychological assessments substantially improves classification performance and can be used for frontotemporal dementia and AD differential diagnoses [23]. Using power spectral analysis and uniform standardized low-resolution brain electromagnetic tomography within δ , θ , α_1 , α_2 , β_1 , β_2 , and β_3 , Caso et al. distinguished 39 Alzheimer's disease patients from among the frontotemporal dementia patients. As a result, the sensitivity is at the degree of chance. In comparison to HC, both studies showed higher expression of diffuse δ/θ and lower central/posterior quicker frequency bands in AD

patients. Patients with frontotemporal dementia had diffusely higher θ capacity than HC patients and lower δ than AD patients. In comparison to frontotemporal dementia patients, AD patients had diffusely higher θ power in the power spectrum and reduced α_2 and β_1 in central/temporal areas using standardized low-resolution brain electromagnetic tomography. Slower frequencies are becoming more important, whereas higher frequencies are becoming less relevant. In patients with moderate levels of frontotemporal dementia and in HC, studies of global field force, which is a metric for the electric field pressure in the entire brain, were combined with EEG neuroimaging observations with low-resolution standardized brain electromagnetic tomography (sLORETA) [24]. Important group effects were found in the global field power in the δ (1.5 to 6 Hz), α_1 (8.5 to 10 Hz), and β_1 (12.5 to 18 Hz) bands. Differences in activation were seen in the 1 band (health control > frontotemporal dementia) in the orbital frontal and temporal lobes, the band (Alzheimer's disease > health control) in widespread areas like the frontal lobe, and the δ band (frontal lobe dementia > Alzheimer's disease) in the parietal lobe and sensorimotor region in low-resolution standardized brain electromagnetic tomography research (Figure 2). As a result, it does not appear that a particular brain area is essential in identifying these types.

Snaedal et al. used qEEG to see whether they could tell the difference between 239 patients with AD, 52 patients to Parkinson disease, and 14 patients to FDT [26]. For grouping, the authors of this Icelandic analysis used θ , α_2 , and β_1 , as well as peak frequency. When utilizing a SVM method to classify cases of degenerative diseases from HC, a good-to-excellent distinction was observed, but this was less so when the risk of comorbidity was high [27]. The investigators were able to distinguish AD from Parkinson's disease dementia with 91 percent accuracy, 93 percent for Parkinson's disease dementia-frontotemporal dementia, and 88 percent for AD-frontotemporal dementia. Given the limited sample size of frontotemporal dementia patients, the precision of these statistical figures must be viewed with caution. In general, experiments including frontotemporal dementia face challenges in attracting volunteers, so the significance of this research should not be overlooked. Nonetheless, adequate feature subset selection is required for classification analysis, particularly in experiments with long vectors with features, such as this one, which included 1120 entries. It is unclear if the genetic algorithm's 10-fold cross-validation used a different preparation, assessment, and research collection in this analysis. This study reduced the original count of 382 studies to 126 studies after eliminating unqualified studies, as seen in Figures 3(a) and 3(b). EEG (64 percent), magneto-encephalograms (28 percent), and fMRI and practical near-infrared spectroscopy were the three types of studies (7%).

3. Preprocessing of EEG Signals

Hans Berger invented EEG, a noninvasive technique of functional imaging for studying the brain, in 1923. EEG

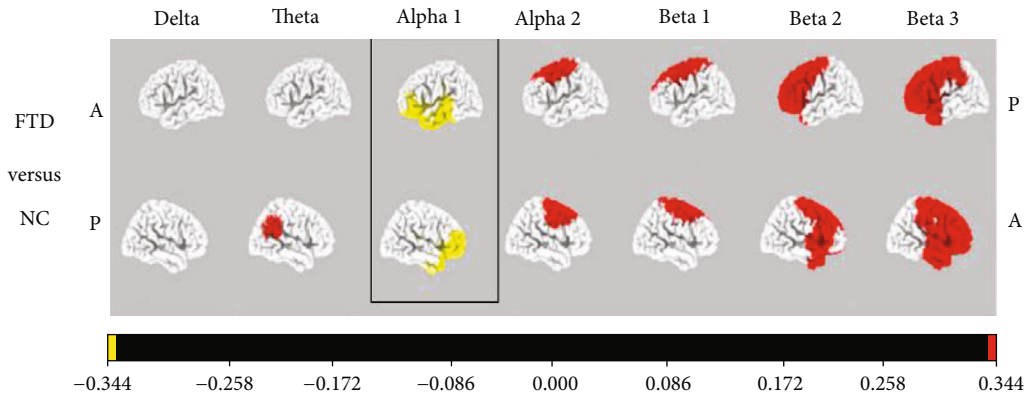


FIGURE 2: For the frontotemporal dementia and control classes, current density images in Talairach space collected by sLORETA were compared [25].

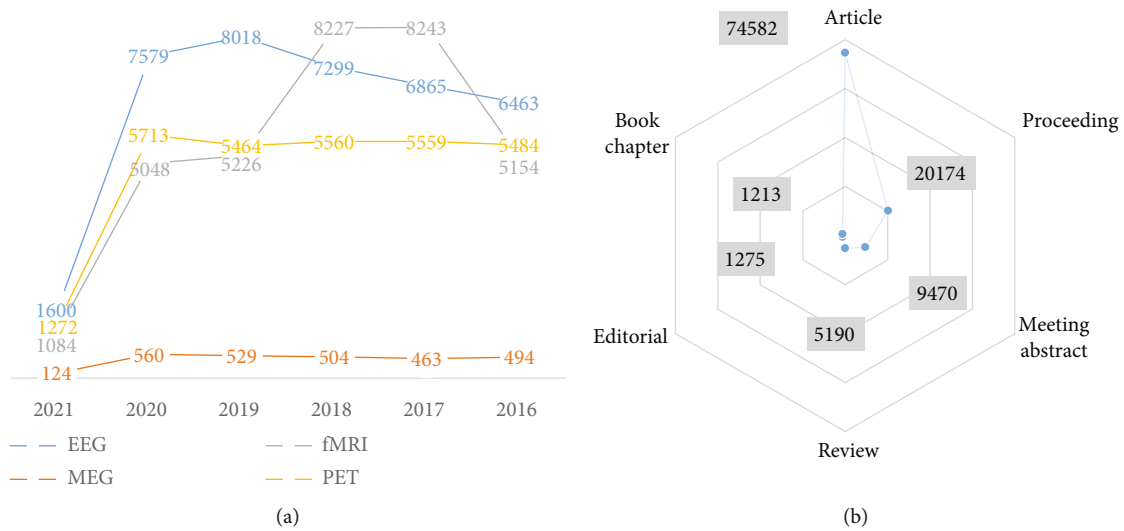


FIGURE 3: (a) Number of publications about EEG, EMG, fMRI, and PET between 2016 and 2021. (b) Types of published papers about EEG.

measures the electrical output of a community of neurons to capture electric signals of the brain from the cerebral mantle [28]. EEG has a poorer spatial resolution than functional MRI but has a better temporal view into neuronal activity. Until operation EEG signals, five frequency bands are usually examined, δ (up to 4 Hz), θ (4 to 8 Hz), α (8 to 12 Hz), β (12 to 26 Hz), and γ (26 to 100 Hz) [29]. Table 1 summarizes these frequency bands and their associations with human activity.

EEG has a frequency range of 1–100 Hz and a voltage range of 10–100 μ V. To detect a disorder or decipher brain function using EEG data, utilizing the Fourier transform or wavelet transform, extract features and utilize spectral information from raw EEG dataset [29]. After that, the extracted features or transformed raw data are utilize to train a ML-based classifier, with deep learning algorithms proving to be effective at automated feature extraction for testing. Centered on the location of the reference electrode, EEG recording can be done in two ways [29].

Experts believe that picture preprocessing is a bad idea since it affects or alters the raw data’s actual nature. Intelligent picture preprocessing, on the other hand, can give

benefits and address issues, resulting in enhanced locally and globally feature recognition. Image preprocessing may have a significant beneficial impact on the quality of feature extraction and machine vision findings. The statistical normalizing of a data collection, which is a frequent step in many visual feature techniques, is comparable to image preprocessing [30]. This is why a thorough study of picture preprocessing is important. A local binary encoder utilizing gray scale data, for instance, will involve different preprocessing than a color SIFT method; moreover, some investigative effort may be necessary to fine-tune the picture preprocessing stage for optimal results. The pixel intensity measurements of point pairs are dealt with using local binary features. As a consequence, the evaluations are highly insensitive to lighting, brightness, and contrast, and picture preprocessing may not be required to get satisfactory findings. Current literature-based local binary pattern techniques do not generally require significant picture preprocessing; instead, they depend on a simple matching criterion that can be modified to accommodate for lighting or contrast [30]. A Fourier transform calculated across the whole picture or block is an example of a global or regional

TABLE 1: Frequency bands in EEG and associated studies of brain control [29].

Bands	Range (Hz)	Human nature and the relationship
δ	1-4	Infants and average adults' deep sleep periods are the most common places to see it.
θ	4-8	A high θ rhyme meaning in awake adults indicates irregular cognitive function.
	8-12	In normal relaxed people, it is usually found in the posterior area of the brain.
β	12-26	Present in the frontal lobe of the brain and in nervous people who are conscious.
γ	26-1000	Predominantly present in people who are anxious, satisfied, or conscious.

basis space feature that spans a regular-shaped polygon. However, basis space characteristics, such as the Fourier spectrum of the LBP summary, which may be computed over histogram bin values of a local identifier to give rotational normalization, may be part of the local features. Another case is the Fourier descriptor, which is used to construct polygon factors for radial line segment lengths to offer rotational invariance by displaying the roundness of a feature. Rather than fixing issues, enhancements are utilized to optimize for certain feature measuring techniques. Sharpening and color balance are two common picture processing improvements [30].

Harris hawk's optimization [31], multiswarm whale [32], Moth-flame optimizer [33–35], gray wolf [36, 37], fruit fly [38, 39], bacterial foraging optimization [40], boosted binary Harris hawk's optimizer [41], anIt colony [42, 43], biogeography-based whale optimization [44], and grasshopper optimizer [45] are some optimization methods based on metaheuristic algorithms. Furthermore, biological applications of machine learning are common, such as tuberculosis [46], thyroid nodules [47], Parkinson's disease [48], and paraquat-poisoned individuals [49, 50]. The reference electrode is located on an electrically inactive region, and the active electrode is located on an electrically active area (e.g., an ear lobe). Scalp EEG is the standard technique for capturing EEG signals, which involves placing electrodes on the surface of the skull [51]. The biggest disadvantage of scalp EEG is that due to the vast spacing among neurons within the skull and the electrodes, the captured signals become blurred. Intracranial electroencephalography signals are recorded by inserting electrodes on the exposed region of the brain to improve signal strength in terms of interference and amplitude [51].

4. Feature Extraction of EEG Signals

Every prediction models must have consistent features that are well associated with the preictal and interictal levels. Those features may be classified as univariate (means that the measurements were taken separately on any EEG channel) or multivariate (means that the EEG measurements on two or up channels) on the basis of the amount of EEG channels. Of these may be further classified as linear or nonlinear elements. For ES estimation, Waser et al. contrasted the efficiency of univariate and bivariate tests that included methods that are both linear and nonlinear [52]. They discovered that preictal deviations occurred 5-30 minutes before the start of ES by using univariate tests. Bivariate

tests, on the other hand, worked preferred by capturing preictal changes least 240 minutes afore the start of an ES. Figure 4 depicts some of the linear and nonlinear ES estimation measures utilized in the related work. Nonlinear tests worked better or were equivalent to linear measures in some cases. Machine learning algorithms, such as artificial neural networks, k -means clustering, decision trees, SVM, and fuzzy logic, are used to identify preictal and interictal patterns from EEG results [53]. To draw conclusions, most people use thresholds based on function values. Machine learning-based research, on the other hand, has mostly focused on the processing of optimized features for projection. At the clinical stage, the EEG signal is provided in the couple the time and frequency domains. Since EEG signals are nonstationery and brain rhythms occur in time domain, also, the signal must be interpreted in both time and frequency domains [53] (see Figure 4).

The calculation of relative EEG power in each EEG frequency band is performed to check the slowing result in the EEG signal of Alzheimer's disease patients. Low-frequency bands (δ and θ bands), i.e., frequency area among 0.5 to 8 Hz, have a high relative power. The normative measure of EEG signal irregularity, such as Lempel Ziv complexity [54, 55], is used to quantify this irregularity. The spectrum of EEG signal is resulted by neurodegenerative disorders like mild cognitive impairment (MCI) and AD. Alzheimer's disease and mild cognitive impairment allow the EEG signal to slow down, according to recent research. The power in low-frequency bands (δ and θ bands, 0.5–8 Hz) is increased in EEG signals from Alzheimer's patients, while power in high-frequency bands (α and β bands, 8–30 Hz) is reduced. The power spectral density function aids in the evaluation of each epoch's spectral characteristics [56]. To achieve a normalized Postsynaptic density, also, the postsynaptic density is multiplied with the overall power in the frequency range of 0.1 to 40 Hz [56]. To acquire data from the EEG, good signal processing methods are needed because the data recorded by the EEG is a complex waveform. Doma and Pirouz [57] explained why the EEG signals are not stored in their normal state and why the captured data is not used for study in its original form. It is preferable to preprocess the EEG signals before beginning the process of extracting indications. The Fast Fourier Transform method is the most widely used signal processing method. Spectral, mapping, morphological localization, time metric, correlation, auxiliary, segment analysis, and other signal processing approaches should be noted. Figure 5 depicts the use of neural networks in the area of EEG signal processing in this study.

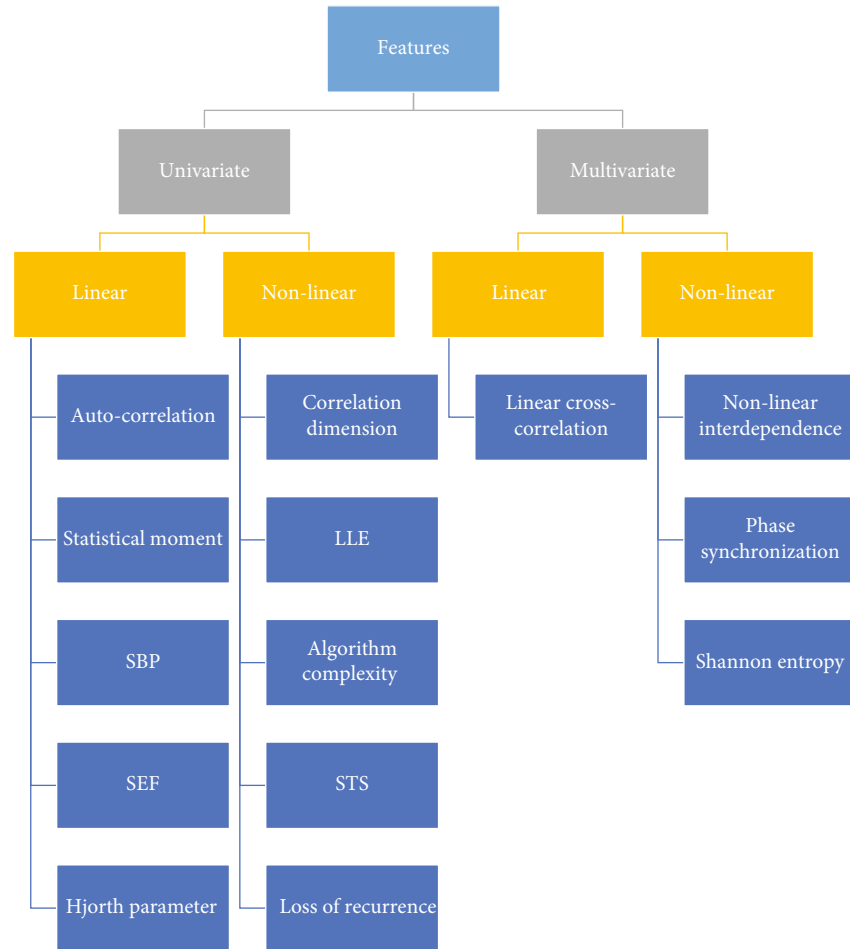


FIGURE 4: The number of channels in EEG data is used to categorize features.

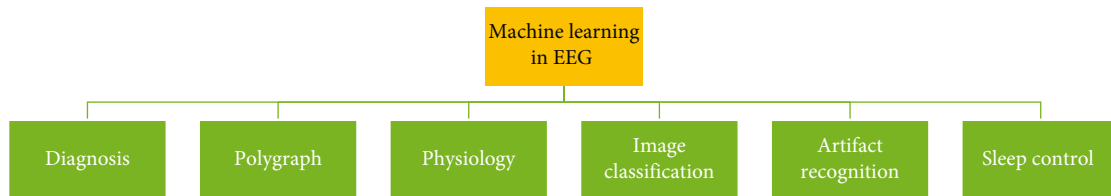


FIGURE 5: The process of EEG signal based on machine learning classifier.

Sadati et al. [58] used an adaptive diffusion neural network to diagnose epilepsy. Use DWT subband energy to extract features. However, their proposed method achieved an accuracy of about 85.9%. Ocak [59] proposed a method for feature extraction and DWT using approximate entropy and achieved an accuracy of more than 96% when using DWT and not using DWT. Nunes et al. did not just classify sentences A and E [60], but checked the complete data set of the University of Bonn (data sets A, B, C, D, and E) and checked various combinations of feature extraction and classification methods. The average accuracy of wood as a classifier is 89.2%. Subasi and Gursay [61] studied various analysis methods to reduce the size of EEG data and com-

bined EEG data with principal component analysis (PCA), linear discriminant analysis (LDA), and independent component analysis (ICA). Subasi [62] uses wavelet transform for feature extraction and expert model for classification. The overall accuracy of this method has reached 94.5%. Recently, Chen [63] introduced the double Fourier tree of complex waveforms as a feature extraction method and used the nearest neighbor classifier for classification. The proposed method achieved the ideal classification accuracy (100%). Djemili et al. used another newer method, which also achieved the desired classification speed. [64] uses empirical mode decomposition for feature extraction, and uses a multilayer perceptual neural network as a classifier.

5. Classification on Alzheimer Disease

EEG data are utilized to detect human brain diseases, as well as mental and emotional states, using a variety of deep learning architectures. The electroencephalogram (EEG) monitors the brain's neuro-activities, also known as brainwaves. Alpha waves, theta waves, beta waves, gamma waves, and delta waves are five distinct frequency waves. The neuroscience community has used several deep learning algorithms to analyze these brain waves in order to diagnose brain diseases and recognize human feelings [65]. Convolutional neural network (CNN), auto encoder (AE), recurrent neural network (RNN), deep belief network (DBN), restricted Boltzmann machine (RBM), multilayer perceptron neural network (MLPNN), optimized deep neural network, and EEG-functional magnetic resonance imaging- (EEGfMRI-) based deep learning are some of the most common learning algorithms.

Most classification trials, to our knowledge, have used data obtained from healthy people [66]. EEG data was used in 27 of these experiments to characterize feelings. Since our method uses a four-electrode EEG sensor, we will concentrate on studies that have used a small number of electrodes. Past experiments that used EEG data from up to four electrodes connected to healthy individuals are summarized in Table 1. Seo et al. [67, 68] and Kim et al. [69] used EEG data obtained from two electrodes. Lee et al. [70], for example, learned an SVM model but did not disclose the model's accuracy. Furthermore, rather than teaching a model, Kim et al. used general research to examine the association between EEG and eye-tracking data [69]. Lee et al. [70] did not go into depth about the methods they used for model output validation (if any). Finally, 5-fold cross-validation and leave-one-out cross-validation were used by Seo et al. [67, 68] to test their models. The experiments in Table 2 are aimed at classifying the feelings of healthy people; as a result, their findings may not be specific to patients with neurological disorders. However, based on these findings, we can infer that EEG data collected from electrodes on the forehead has the capacity to distinguish human emotions. As a result, classifying emotions using EEG data taken from an AD patient's forehead may be a promising avenue to pursue.

The association among signals x and y as a frequency structure, varying from 0 to 1, is known as coherence. Volume conduction through the scalp can have an effect on this measurement. In two trials, θ range coherence was found to be stronger in than in AD [54, 55]. In one study [74], α and β coherence were shown to be lower in dementia with Lewy body disease relative to AD, whereas other studies [75] found higher α and β coherence in dementia with Lewy body disease. Granger causality is also utilizing to describe how the time course of the EEG in channel X could be used to estimate possible EEG signal values in channel Y . According to one study, parietal area Granger causality is slightly greater in dementia with Lewy body disease than in AD, with a high precision of $\sim 100\%$. The PLI calculates a stable causal delay among two signal sources and is slowly influenced with volume conduction on the scalp. PLI ratings range from 0 to 1, with 0 indicating no causal synchroniza-

tion and 1 indicating complete causal synchronization. Dementia with Lewy body disease had a lower PLI within the α spectrum than AD, suggesting more extreme improvements in connectivity in dementia with Lewy body disease. The changes in α network connectivity are consistent with another analysis that found lower mean α band guided phase shift entropy in dementia with Lewy body disease relative to AD, which tests posterior-to-anterior connectivity [76].

Weighted phase lag index (PLI) is a variation of phase lag index that entails weighting the PLI rates by the imaginary portion of the cross-spectrum between the two time-series [77]; the latter part of the cross-spectrum is related to the phase difference, or delay, between the signals. The two signs are nearly overlapping if the imaginary component is close to 0. One advantage of weighted PLI can be significantly raised by loud conducting sources, although this effect is less pronounced in weighted PLI [78]. Only one study used this method and found that dementia with Lewy body disease had a lower weighted PLI in the β band than Alzheimer's disease [79]. LLC is a connectivity metric that is calculated with the aid of precise low-resolution brain electromagnetic tomography tools. LLC is less affected by volume conduction and calculates functional cortical source connectivity by eliminating zero-lag instantaneous step coupling among cortical sources of resting state EEG rhythms. When comparing AD to dementia with Lewy body disease, LLC in the α and δ levels was lower in AD, which Babiloni et al. [22] speculated may indicate that AD had more cortical disconnection as both disorders progressed to dementia [22]. To test functional network connectivity, a graph technique focused on weighted network, and least spanning tree (MST) processes was used. The one study that looked at weighted PLI found that dementia with Lewy body disease (LBD) had lower connectivity and more network segregation in the β network than AD [80]. MST was used in four experiments [81, 82], all of which found that dementia with LBD had a less degree, less Euclidean distance, upper diameter, higher eccentricity, and less leaf-fraction than AD [82], implying a less-efficient network. Lewy body disease tends to have a randomized sequence consistent with decreased performance and synchronization [82]. EEG connectivity results in dementia with Lewy body disease (LBD) are summarized in Table 3.

6. EEG Signal Complexity Analysis of AD

A variety of nonlinear approaches have been used to investigate the features of brain function in Alzheimer's patients, yielding a host of intriguing findings. Resting-state recordings offer more accurate estimates of brain adaptability because they are not affected with task-specific arousal or discrepancy in impetus or success [84]. Resting brain function records and task-related observations show network dynamics that are close [85, 86] and also represent the influence of metabolically active networks. The time resolution of the EEG signal is very high, and it has been discovered that the signals have been studied mostly in various frequency bands and using from electrodes to show the diversity in signal rates.

TABLE 2: EEG data are used to classify the emotions of healthier individuals (up to four electrodes).

Investigation	Emotional responses to be targeted	Method	Accuracy	Test
[71]	Happiness, rage, sorrow, fear, relaxation	Support vector machine (SVM)	73.32	Leave-one-out cross-testing
[72]	Engagement, perplexity, dissatisfaction, positive attitude	SVM, k -nearest neighbors (KNN)	95.69	—
[73]	Sorrow, displeasure	Multiclass support vector machine classifier	84.83	—
[70]	Arousal, sensitivity	SVM, K -means	—	—
[67]	Dissatisfaction, satisfaction	KNN	86.73	5-fold cross-testing
[68]	Dissatisfaction, satisfaction	Multilayer perceptron (MLP)	79.98	5-fold cross-testing
[69]	Boredom, frustration	Analysis	—	—

TABLE 3: Lewy body disease, a review of studies on EEG connectivity controls.

Author	Subband	Metrics	Outcome
[82]	β	Phase transfer entropy	AD > LBD
[80]	β	Weighted phase lag index	AD > LBD
[81]	α	Phase lag index	AD > LBD
[83]	α	Phase lag index	AD > LBD
[81]	α	Phase lag index	AD > LBD
[83]	α	Phase transfer entropy	AD < LBD
[22]	α	Lagged linear connectivity	AD < LBD/Parkinson's disease dementia
[22]	δ	Lagged linear connectivity	AD < LBD/Parkinson's disease dementia

6.1. *The Signal Complexity Analysis in EEG.* The signal complexity of the resting-state EEG in spinal cord injury, MCI, and AD patients is compared to standard controls in this segment. As applied to EEG signals, multiplex complexity technique, such as LZC, entropy complexity, and another complexity characteristics, has been shown to vary between spinal cord injury, MCI, AD, and control subjects in many experiments. Hogan et al. [87] discovered that MCI patients had a low entropy. According to a new analysis, the difficulty rates of EEG signals from AD patients are lower than those of spinal cord injury patients in all channels. ApEn and SampEn [3] in EEG signals have been seen to be slightly lower in healthy control and Alzheimer's disease patients relative to healthy individuals [88]; Garn et al. used various approaches [89] to investigate the complexity of EEG signals from AD patients and maturity clinical trial. In the EEGs of patients with AD, consistent findings were observed, including a substantial decrease in complexity at electrodes P₃, P₄, O₁, and O₂ positioned over the parietal, occipital, and temporal areas as compared with the healthy people. The medial temporal lobe, which is linked to short-term memory, is impaired during the MCI stage, as are the lateral temporal lobe and parietal lobe. The frontal lobe is compromised in the moderate stage of Alzheimer's disease. The occipital lobe is compromised during the acute stage of Alzheimer's disease [90]. The brain states that form during the transition from safe to AD have been studied using a variety of entropy approaches. The majority of the research has concentrated

on specific regions of the brain. Patients with Alzheimer's disease and healthy control have less En values in all five areas (EnAD EnMCI EnControl), with major variations in the frontal, temporal, and central areas. These findings indicate that the frontal, temporal, and central EEG impulses in AD and MCI patients' brains were slightly less complex than those in HC. Furthermore, AD patients have the least difficulty and the most consistency. The complexity of EEG signals declines with disease progression, as predicted, particularly for comparing HC issues to Alzheimer's disease patients [91].

6.2. *Conditions of EEG Recording and Symptoms of AD.* Many experiments have looked at the impact of AD and its development on EEG signals over the past few decades. EEG signals have been used in studies under a variety of recording environments, which can be divided into two categories:

6.2.1. *Resting State.* The brain background activation is measured by recording spontaneous EEG activity in the absence of some sort of stimuli. The acquisition of EEG data becomes less difficult, rather relaxed, and less stressful for the user, particularly for aged people [92], since the person is not expected to perform any particular task. A condition of rest EEG records include recordings made while resting-awake as well as recordings made while sleeping. AD has been shown to have four distinct impacts on resting-state EEG signals:

- (1) *Slowing*. In AD patients, power spectrum transitions from up-frequency ingredients (α , β , and γ) to low-frequency components (δ and θ) are normal [14, 93, 94]. The lack of cholinergic innervations in AD patients is believed to be the cause of this transition, which is proportional to the progression of the disease. The slowing of the EEG has been quantified using features obtained from the power curve, power spectrogram, and wavelet analysis
- (2) *Reduced Complexity*. By comparing AD patients to healthy controls, a reduction in the complexity of brain electrical activity has been found [14, 94–96]. Massive neuronal death and decreased interactions in cortical regions are likely to blame for this decline, which results in simplified EEG dynamics. Entropy metrics, auto mutual detail, Lempel-Ziv complexity, fractal dimension, and the Lyapunov exponent are some of the signal processing techniques used to investigate the complexity of EEG signals [92]
- (3) *Synchronization Declines*. This has been seen in many AD patients as a decrease in communication between cortical regions. While the cause of this syndrome is unknown, it is believed to be linked to atrophy in neural network connectivity
- (4) *Deficiencies in Neuromodulation*. Via cross-frequency interaction effects, the utilization of amplitude modulation to test EEG rhythms and brain neuromodulatory acting has lately been proposed [92]

6.3. *EEG Recordings Associated with a Particular Event*. The signals from the EEG are time-blocked, meaning they are captured in response to the occurrence of a single event. EEG operation is step locked as well as time-blocked, earning it the term event-related potentials. Induced activity is described as EEG activity that is not phase locked and can be examined using either event-related (de) synchronization [23] or event-related oscillations. Sensorial perceptive, motor, and cognitive functions can all be linked to events. Recent studies of the utilized event-connected EEG for Alzheimer's disease detection have been published in the AD literature. Although event-connected EEG studying enables researchers to investigate the impact of AD on individual brain circuits, these monitoring environments are not suitable for most AD patients, who experience a rise in anxiety and frustration, as well as a decline in their ability to do new things, even in the early stages of the disease. As a result, even completing a basic memory task may cause the patient pain and anxiety; they can become disoriented or unable to accomplish it [97]. Resting-state protocols, also, do not include extraneous stimulation, making them more straightforward and convenient for patients. Furthermore, these protocols produce less artifacts.

Some new articles on resting-state research for Alzheimer's disease detection have also been published. None of them, however, have focused solely on the subject of EEG-regarding Alzheimer's disease detection. Some reviews, for example, do not look at EEG as a primary diagnostic tool

[93, 94, 98], whereas others are solely concerned with EEG signal synchronization [95, 96]. Furthermore, other publications [99, 100] offer a wider overview of the entire dementia continuum, not just AD. In revisions [101, 102], the key function types for Alzheimer's disease detection are thoroughly explored. As a result, the current research adds to previous studies of EEG-regarding Alzheimer's disease detection by regularly and exclusively analyzing papers on resting-state EEG to offer a comprehensive overview of the current state of the subjects.

7. Datasets

Three styles of datasets had been used for enforcing and verifying our methods. The first datasets are used for epilepsy diagnosis, and the 1/3 is used for autism diagnosis. The first dataset is provided through the Bonn University, Germany, and protected 5 units, named A, B, C, D, and E. Each set includes precisely a hundred single-channel EEG signals. Sets A and B had been accrued from scalp EEGs of neurotypical persons, while units C, D, and E had been accrued the use of intracranial EEGs from epileptic persons. The general duration of every sign is about 23.6 s. The records had been accrued with a sampling frequency of 173.61 Hz. The reference furnished in [103] indicates an extra specified description of this dataset. The study crew from MIT, USA [104], affords the second one dataset, which incorporates 906 h of EEG records accrued from 23 epileptic patients. In this study, handiest records for the primary twelve epileptic topics had been used, in conjunction with the ones of 11 neurotypical topics. This record consists of 23 EEG channels with a sampling frequency of 256 Hz [105].

The 1/3 dataset become furnished with the aid of using King Abdulaziz University (KAU) Brain-Computer Interface (BCI) Group, Jeddah, Saudi Arabia. The dataset become accumulated in a comfortable kingdom and cut up into groups: the primary institution become named the neurotypical institution and protected information from ten healthful volunteer subjects (all men, age 9–sixteen years) with common intelligence and with none intellectual disorders. The 2nd institution become classified the autistic institution and protected 9 subjects (six men and 3 females, elderly 10–sixteen years) with ASD. The EEG indicators had been accumulated from the subjects' scalps in a comfortable kingdom the usage of an EEG information-acquisition machine that protected the subsequent components: a g.tec EEG cap with excessive accuracy, sixteen Ag/AgCl sensors (electrodes) primarily based totally at the 10–20 global acquisition machine, g.tec USB amplifiers (gtec scientific engineering company, Schiedberg, Austria), and BCI2000 software (The Brain-Computer Interface R&D Program on the Wadsworth Center of the New York State Department of Health in Albany, NY, USA). The dataset become filtered with the aid of using a band-byskip clear out with a passband of 0.1–60 Hz, and a notch clear out become used with a stopband frequency of 60 Hz. All EEG indicators had been digitized at a sampling frequency of 256 Hz. The EEG series time ranged from 12 to forty min for autistic sufferers with a complete of as much as 173 min. For neurotypical sufferers,

TABLE 4: Basic EEG features' classification accuracy.

EEG features	Studies	TPR	FPR	ACC	AUC
Dementia with Lewy bodies vs. AD	[81]	97%	100%	99%	—
EEG severity grade	[109]	72–79%	76–85%	—	0.78–0.90
Grand total EEG	[22] [100]	65–78%	67–74%	70–73%	0.72–0.75
Occipital α power	[22]	78%	67%	73%	0.72
δ standard deviation	[110]	92%	83%	—	0.94
θ FP + θ power + $\theta - \alpha$ DFV	[111]	~100%	~100%	~100%	—
Combined spectral array pattern	[83]	93%	97%	95%	—
Phase lag index β band	[82]	80%	85%	—	0.86
Minimum spanning tree-phase lag index	[80]	47%	100%	66%	0.78
P300- reversed amplitude distribution gradients	[21] [80] [112] [113]	76–100%	77–100%	66–100%	0.78–0.93
Machine learning algorithms	[106]	—	—	—	—
EEG severity grade > 2	[106]	—	—	—	0.76
Diffuse abnormalities	[106] [108]	51%	86%	—	0.84
Peak/dominant frequency	[106] [108]	61%	81%	—	0.70–0.89
β power	[106] [108]	41%	97%	—	0.71–0.91
α power	[108]	56%	83%	—	0.66–0.85
Pre- α power	[106] [108]	33%	89%	—	0.68
θ power	[106] [108]	23%	89%	—	0.60–0.94
δ power	[106] [108]	49%	83%	—	0.54–0.55
θ/α ratio	[106]	—	—	—	0.64–0.92

the recording is various among five and 27 min with a complete of as much as 148 min.

8. Prodromal Dementia with LBD

Even at the mild cognitive impairment level, EEG anomalies on visual rating have been noted to be more frequent in dementia with LBD. When comparing MCI with Lewy bodies (MCI-LB) to MCI due to AD (MCI-AD), MCI-LB had more diffused anomalies (76 percent vs. 8%) and FIRDA (22 percent vs. 0%) [106]. EEG severity ratings were also slightly lower in MCI-LB, with just 16 percent of MCI-LBs having regular EEGs compared to 49 percent of MCI-ADs [106]. MCI-LB empirical EEG results have been compared to those published in dementia with Lewy body disease and Alzheimer's disease, with MCI-LB having a lower dominant frequency than MCI-AD [106, 107]. This results in a higher θ/α ratio and higher pre- α power, as well as lower α and β power and a lower θ/α ratio [106–108]. In identifying MCI-LB (Table 4), Schumacher et al. 2020b found that spectral strength tests had sensitivities of 23 to 51 percent, precision of 81 to 97 percent, and a region under receiver

operating specification curve of 0.54 to 0.71 [108]. Van der Zande et al. 2020, but on the other hand, registered an AUROC of 0.76 to 0.97, but no sensitivities or specificities [106]. The MCI-LB connection was only studied in one study, which showed that LLC was lower in MCI-LB and MCI-AD as compared to age-matched controls, but no difference between the MCI groups [22] (see Table 4).

Several research looked at whether EEG features would predict dementia development in MCI patients. In one study, MCI patients who progressed to dementia with Lewy body disease (MCI-LB) had a lower mean frequency and α/θ ratio than those who suggested MCI-AD [107]. Other research utilized CSA to assess progression from MCI to dementia with Lewy body disease, AD, or no progression at 3 years in patients by MCI, with an average accuracy of 76%. Both patients with MCI who progressed to dementia with Lewy body disease had a CSA pattern of >1 (1-5) at baseline, while 93 percent of patients who improved to Alzheimer's disease had a CSA pattern of 1 (stable α) at baseline [111]. However, in 75% of patients with MCI, the involvement of one or more central or positive clinical characteristics of dementia with Lewy body disease predicted

progression to dementia with Lewy body disease. The dominant frequency variability was comparable when MCI patients who advanced to dementia with Lewy body disease (MCI-LB) were compared to dementia with Lewy body disease patients, despite dementia with Lewy body disease patients having lower mean dominant frequencies [111]. After having comparable MMSE scores at baseline and a similar decrease on follow-up (20.6 in dementia with Lewy body disease and 20.5 in Alzheimer's disease), follow-up EEG of MCI-dementia with LBD patient's demonstrated improvement, with all patients with CSA 1 plus progressing to CSA 2 or 3. In contrast, considering the cognitive impairment, follow-up EEG of patients with MCI-AD revealed no progression (93 percent with CSA trend 1).

9. Conclusions

Alzheimer's disease is a sophisticated brain disease with massive financial, social, and medical consequences. It is recognized as the leading cause of dementia, characterized by amyloid peptide and phosphorylated tau (p-tau) protein accumulation and aggregation, as well as dementia, neuron loss, and brain atrophy. Despite decades of study, no acceptable medication exists that will stop the progression of Alzheimer's disease by acting on the illness's root cause, whereas currently existing therapies merely give symptomatic relief and do not provide a definitive cure or protection. Clinical signs, health information, family consultations, and current screening procedures such as clinical, neurological, and psychiatric examinations are used to diagnose Alzheimer's disease, whereas neuropsychological testing can be acknowledged as a tool for detecting unbiased signs of memory disturbances in the early stages, and laboratory studies such as thyroid function tests and serum vitamin B12 are used. To wrap up this report, we will discuss some of the remaining problems and study topics. Obtaining EEG data from MCI or AD patients is currently very complicated. In comparison to ECG and other biomedical records, such databases are not open to the public. As a consequence, consistently benchmarking and evaluating the latest approaches for the detection of Alzheimer's disease from EEG signals are difficult. Furthermore, almost none of those techniques integrate biophysical information about AD; comprehensive mathematical models of AD pathology combined with EEG data analysis can aid in improving AD diagnosis. Combining EEG with another signal and imaging methods, such as MRI dMRI, TMS, and SPECT, may yield even better results. The relationship among AD risk criteria (e.g., elevated homocysteine levels in the blood) and EEG characteristics needs to be studied further. Furthermore, the exact relationship between cognitive and memory loss and EEG disorders in Alzheimer's patients is still largely unknown. It is also crucial to see how EEG can help differentiate between MCI and various phases of AD, as well as between AD and other dementias. The EEG monitoring state is an important degree of freedom: it may be recorded (i) when the subject is at rest; (ii) when the subject is performing working-memory or other tasks; and (iii) when the subject is being activated with auditory, visual, tactile, or other cues.

EEG signals can be more or less discriminative for MCI and AD depending on the recording situation; a thorough exploration of various recording situations with the goal of detecting MCI and AD is needed. In the future, it is also essential to evaluate the EEG in clinical studies of Alzheimer's disease, where the disease's development can be closely monitored; such studies may help us relate EEG abnormalities to AD neuropathology. Another intriguing line of investigation is the effect of treatment and therapy on the EEG of AD patients.

Data Availability

This is a review paper and data is not applicable.

Disclosure

The funding sources had no involvement support in the study design, collection, analysis, or interpretation of data, writing of the manuscript, or in the decision to submit the manuscript for publication.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] R. H. Blank, "Alzheimer's Disease and Other Dementias: An Introduction," in *In Social & Public Policy of Alzheimer's Disease in the United States*, Palgrave Pivot, Singapore, 2019.
- [2] M. Amini, M. Pedram, A. Moradi, and M. Ouchani, "Diagnosis of Alzheimer's disease severity with fMRI images using robust multitask feature extraction method and convolutional neural network (CNN)," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 5514839, 2021.
- [3] J. Sun, B. Wang, Y. Niu et al., "Complexity analysis of EEG, MEG, and fMRI in mild cognitive impairment and Alzheimer's disease: a review," *Entropy*, vol. 22, no. 2, p. 239, 2020.
- [4] M. Amini, M. M. Pedram, A. Moradi, M. Jamshidi, and M. Ouchani, "Single and combined neuroimaging techniques for Alzheimer's disease detection," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 9523039, 2021.
- [5] M. Amini, M. Pedram, A. Moradi, and M. Ouchani, "Diagnosis of Alzheimer's disease by time-dependent power spectrum descriptors and convolutional neural network using EEG signal," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 5511922, 2021.
- [6] M. Fraiwan, M. Alafeef, and F. Almomani, "Gauging human visual interest using multiscale entropy analysis of EEG signals," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2435–2447, 2021.
- [7] Y. Niu, B. Wang, M. Zhou et al., "Dynamic complexity of spontaneous bold activity in Alzheimer's disease and mild cognitive impairment using multiscale entropy analysis," *Frontiers in Neuroscience*, vol. 12, p. 677, 2018.
- [8] R. P. Turner, "Clinical application of combined EEG-qEEG functional neuroimaging in the practice of pediatric neuroscience: a personal perspective," *Clinical EEG and Neuroscience*, vol. 52, no. 2, pp. 126–135, 2020.

- [9] G. R. Iannotti, *Combined EEG-fMRI in Epilepsy: Methodological Improvements and Application to Functional Connectivity*, Doctoral dissertation, University of Geneva, 2018.
- [10] M. Ahmadi, A. Sharifi, M. Jafarian Fard, and N. Soleimani, "Detection of brain lesion location in MRI images using convolutional neural network and robust PCA," *International journal of neuroscience*, pp. 1–12, 2021.
- [11] Z. Gao, X. Cui, W. Wan, and Z. Gu, "Recognition of emotional states using multiscale information analysis of high frequency EEG oscillations," *Entropy*, vol. 21, no. 6, p. 609, 2019.
- [12] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Computational Statistics & Data Analysis*, vol. 143, p. 106839, 2020.
- [13] A. S. Malik and W. Mumtaz, *EEG-Based Experiment Design for Major Depressive Disorder: Machine Learning and Psychiatric Diagnosis*, Academic Press, 2019.
- [14] M. Ahmadi, A. Sharifi, S. Hassantabar, and S. Enayati, "QAIS-DSNN: tumor area segmentation of MRI image with optimized quantum matched-filter technique and deep spiking neural network," *BioMed Research international*, vol. 2021, Article ID 6653879, 2021.
- [15] M. Fan, A. C. Yang, J. L. Fuh, and C. A. Chou, "Topological Pattern Recognition of Severe Alzheimer's Disease via Regularized Supervised Learning of EEG Complexity," *Frontiers in Neuroscience*, vol. 12, no. 685, 2018.
- [16] R. Franciotti, A. Pilotto, D. V. Moretti et al., "Anterior EEG slowing in dementia with Lewy bodies: a multicenter European cohort study," *Neurobiology of Aging*, vol. 93, pp. 55–60, 2020.
- [17] R. Nardone, L. Sebastianelli, V. Versace et al., "Usefulness of EEG techniques in distinguishing frontotemporal dementia from Alzheimer's disease and other dementias," *Disease Markers*, vol. 2018, Article ID 6581490, 2018.
- [18] S. Nobukawa, T. Yamanishi, S. Kasakawa, H. Nishimura, M. Kikuchi, and T. Takahashi, "Classification methods based on complexity and synchronization of electroencephalography signals in Alzheimer's disease," *Frontiers in psychiatry*, vol. 11, p. 11, 2020.
- [19] Z. K. Law, C. Todd, R. Mehraram et al., "The role of EEG in the diagnosis, prognosis and clinical correlations of dementia with Lewy bodies—a systematic review," *Diagnostics*, vol. 10, no. 9, p. 616, 2020.
- [20] Y. A. Pijnenburg, R. L. M. Strijers, Y. . Made, W. M. van der Flier, P. Scheltens, and C. J. Stam, "Investigation of resting-state EEG functional connectivity in frontotemporal lobar degeneration," *Clinical Neurophysiology*, vol. 119, no. 8, pp. 1732–1738, 2008.
- [21] H. Garn, C. Coronel, M. Waser, G. Caravias, and G. Ransmayr, "Differential diagnosis between patients with probable Alzheimer's disease, Parkinson's disease dementia, or dementia with Lewy bodies and frontotemporal dementia, behavioral variant, using quantitative electroencephalographic features," *Journal of Neural Transmission*, vol. 124, no. 5, pp. 569–581, 2017.
- [22] C. Babiloni, C. Del Percio, R. Lizio et al., "Abnormalities of resting-state functional cortical connectivity in patients with dementia due to Alzheimer's and Lewy body diseases: an EEG study," *Neurobiology of Aging*, vol. 65, pp. 18–40, 2018.
- [23] R. Cassani, M. Estarellas, R. San-Martin, F. J. Fraga, and T. H. Falk, "Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment," *Disease Markers*, vol. 2018, Article ID 5174815, 2018.
- [24] Y. S. Shim and H. E. Shin, "Analysis of neuropsychiatric symptoms in patients with Alzheimer's disease using quantitative EEG and sLORETA," *Neurodegenerative Diseases*, vol. 20, no. 1, pp. 12–19, 2020.
- [25] K. Nishida, M. Yoshimura, T. Isotani et al., "Differences in quantitative EEG between frontotemporal dementia and Alzheimer's disease as revealed by LORETA," *Clinical Neurophysiology*, vol. 122, no. 9, pp. 1718–1725, 2011.
- [26] J. Snaedal, G. H. Johannesson, T. E. Gudmundsson et al., "Diagnostic accuracy of statistical pattern recognition of electroencephalogram registration in evaluation of cognitive impairment and dementia," *Dementia and Geriatric Cognitive Disorders*, vol. 34, no. 1, pp. 51–60, 2012.
- [27] A. Sharifi, M. Ahmadi, M. A. Mehni, S. Jafarzadeh Ghouschi, and Y. Pourasad, "Experimental and numerical diagnosis of fatigue foot using convolutional neural network," *Computer Methods in Biomechanics and Biomedical Engineering*, pp. 1–13, 2021.
- [28] D. Millett, "Hans Berger: from psychic energy to the EEG," *Perspectives in Biology and Medicine*, vol. 44, no. 4, pp. 522–542, 2001.
- [29] Z. Y. Ong, A. Saidatul, and Z. Ibrahim, "Power spectral density analysis for human EEG-based biometric identification," in *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, pp. 1–6, Kuching, Malaysia, 2018, August.
- [30] S. Krig, *Computer vision metrics: survey, taxonomy, and analysis*, Springer nature, 2014.
- [31] H. Chen, A. A. Heidari, H. Chen, M. Wang, Z. Pan, and A. H. Gandomi, "Multi-population differential evolution-assisted Harris hawks optimization: framework and case studies," *Future Generation Computer Systems*, vol. 111, pp. 175–198, 2020.
- [32] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Applied Soft Computing*, vol. 88, article 105946, 2020.
- [33] Y. Xu, H. Chen, J. Luo, Q. Zhang, S. Jiao, and X. Zhang, "Enhanced Moth-flame optimizer with mutation strategy for global optimization," *Information Sciences*, vol. 492, pp. 181–203, 2019.
- [34] M. Wang, H. Chen, B. Yang et al., "Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses," *Neurocomputing*, vol. 267, pp. 69–84, 2017.
- [35] W. Shan, Z. Qiao, A. A. Heidari, H. Chen, H. Turabieh, and Y. Teng, "Double adaptive weights for stabilization of moth flame optimizer: balance analysis, engineering cases, and medical diagnosis," *Knowledge-Based Systems*, vol. 214, article 106728, 2021.
- [36] X. Zhao, X. Zhang, Z. Cai et al., "Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients," *Computational Biology and Chemistry*, vol. 78, pp. 481–490, 2019.
- [37] J. Hu, H. Chen, A. A. Heidari et al., "Orthogonal learning covariance matrix for defects of grey wolf optimizer: insights, balance, diversity, and feature selection," *Knowledge-Based Systems*, vol. 213, article 106684, 2021.
- [38] L. Shen, H. Chen, Z. Yu et al., "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Systems*, vol. 96, pp. 61–75, 2016.

- [39] H. Yu, W. Li, C. Chen et al., "Dynamic Gaussian bare-bones fruit fly optimizers with abandonment mechanism: method and analysis," *Engineering with Computers*, pp. 1–29, 2020.
- [40] X. Xu and H. L. Chen, "Adaptive computational chemotaxis based on field in bacterial foraging optimization," *Soft Computing*, vol. 18, no. 4, pp. 797–807, 2014.
- [41] Y. Zhang, R. Liu, X. Wang, H. Chen, and C. Li, "Boosted binary Harris hawks optimizer and feature selection," *Engineering with Computers*, vol. 37, no. 4, pp. 3741–3770, 2021.
- [42] D. Zhao, L. Liu, F. Yu et al., "Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy," *Knowledge-Based Systems*, vol. 216, p. 106510, 2021.
- [43] X. Zhao, D. Li, B. Yang, C. Ma, Y. Zhu, and H. Chen, "Feature selection based on improved ant colony optimization for online detection of foreign fiber in cotton," *Applied Soft Computing*, vol. 24, pp. 585–596, 2014.
- [44] J. Tu, H. Chen, J. Liu et al., "Evolutionary biogeography-based whale optimization methods with communication structure: towards measuring the balance," *Knowledge-Based Systems*, vol. 212, article 106642, 2021.
- [45] C. Yu, M. Chen, K. Cheng et al., "SGOA: annealing-behaved grasshopper optimizer for global tasks," *Engineering with Computers*, pp. 1–28, 2021.
- [46] C. Li, L. Hou, B. Y. Sharma et al., "Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 211–225, 2018.
- [47] J. Xia, H. Chen, Q. Li et al., "Ultrasound-based differentiation of malignant and benign thyroid nodules: an extreme learning machine approach," *Computer Methods and Programs in Biomedicine*, vol. 147, pp. 37–49, 2017.
- [48] H.-L. Chen, G. Wang, C. Ma, Z. N. Cai, W. B. Liu, and S. J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131–144, 2016.
- [49] L. Hu, G. Hong, J. Ma, X. Wang, and H. Chen, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Computers in Biology and Medicine*, vol. 59, pp. 116–124, 2015.
- [50] Y. Zhang, R. Liu, A. A. Heidari et al., "Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis," *Neurocomputing*, vol. 430, pp. 185–212, 2021.
- [51] G. Alarcón, D. Jiménez-Jiménez, A. Valentín, and D. Martín-López, "Characterizing EEG cortical dynamics and connectivity with responses to single pulse electrical stimulation (SPES)," *International Journal of Neural Systems*, vol. 28, no. 6, p. 1750057, 2018.
- [52] M. Waser, H. Garn, R. Schmidt et al., "Quantifying synchrony patterns in the EEG of Alzheimer's patients with linear and non-linear connectivity markers," *Journal of Neural Transmission*, vol. 123, no. 3, pp. 297–316, 2016.
- [53] K. S. Hong, M. J. Khan, and M. J. Hong, "Feature extraction and classification methods for hybrid fNIRS-EEG brain-computer interfaces," *Frontiers in Human Neuroscience*, vol. 12, p. 246, 2018.
- [54] Y. Li, S. Wang, and Z. Deng, "Intelligent fault identification of rotary machinery using refined composite multi-scale Lempel-Ziv complexity," *Journal of Manufacturing Systems*, 2020.
- [55] T. V. Yakovleva, I. E. Kutepov, A. Y. Karas et al., "EEG analysis in structural focal epilepsy using the methods of nonlinear dynamics (Lyapunov exponents, Lempel–Ziv complexity, and multiscale entropy)," *The Scientific World Journal*, vol. 2020, Article ID 8407872, 2020.
- [56] Q. Zhang, A. Nicolson, M. Wang, K. K. Paliwal, and C. Wang, "Deepmmse: a deep learning approach to mmse-based noise power spectral density estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1404–1415, 2020.
- [57] V. Doma and M. Pirouz, "A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals," *Journal of Big Data*, vol. 7, no. 1, pp. 1–21, 2020.
- [58] N. Sadati, H. R. Mohseni, and A. Maghsoudi, "Epileptic seizure detection using neural fuzzy networks," in *proceedings of the IEEE international conference on fuzzy systems*, pp. 596–600, Vancouver, BC, Canada, July 2006.
- [59] H. Ocak, "Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2027–2036, 2009.
- [60] T. M. Nunes, A. L. Coelho, C. A. Lima, J. P. Papa, and V. H. C. De Albuquerque, "EEG signal classification for epilepsy diagnosis via optimum path forest - A systematic assessment," *Neurocomputing*, vol. 136, pp. 103–123, 2014.
- [61] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659–8666, 2010.
- [62] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.
- [63] G. Chen, "Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2391–2394, 2014.
- [64] R. Djemili, H. Bourouba, and M. A. Korba, "Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals," *Biocybernetics and Biomedical Engineering*, vol. 36, no. 1, pp. 285–291, 2016.
- [65] D. Merlin Praveena, D. Angelin Sarah, and S. Thomas George, "Deep learning techniques for EEG signal applications—a review," *IETE Journal of Research*, pp. 1–8, 2020.
- [66] Y. P. Lin and T. P. Jung, "Improving EEG-based emotion classification using conditional transfer learning," *Frontiers in Human Neuroscience*, vol. 11, p. 334, 2017.
- [67] J. Seo, T. H. Laine, and K. A. Sohn, "Machine learning approaches for boredom classification using EEG," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 10, pp. 3831–3846, 2019.
- [68] J. Seo, T. H. Laine, and K. A. Sohn, "An exploration of machine learning methods for robust boredom classification using EEG and GSR data," *Sensors*, vol. 19, no. 20, p. 4561, 2019.
- [69] J. Kim, J. Seo, and T. H. Laine, "Detecting boredom from eye gaze and EEG," *Biomedical Signal Processing and Control*, vol. 46, pp. 302–313, 2018.
- [70] H. J. Lee, D. I. K. Shin, and D. K. Shin, "A study on an emotion-classification algorithm of users adapting brain-wave," in *In proceedings of symposium of the Korean Institute*

- of Communications and Information Sciences; Korea institute of communication sciences*, pp. 786–787, Seoul, Korea, 2013.
- [71] K. Guo, R. Chai, H. Candra et al., “A hybrid fuzzy cognitive map/support vector machine approach for EEG-based emotion classification using compressed sensing,” *International Journal of Fuzzy Systems*, vol. 21, no. 1, pp. 263–273, 2019.
- [72] M. Li, H. Xu, X. Liu, and S. Lu, “Emotion recognition from multichannel EEG signals using K-nearest neighbor classification,” *Technology and Health Care*, vol. 26, Supplement 1, pp. 509–519, 2018.
- [73] V. Vanitha and P. Krishnan, “Time-frequency analysis of EEG for improved classification of emotion,” *International Journal of Biomedical Engineering and Technology*, vol. 23, no. 2/3/4, pp. 191–212, 2017.
- [74] W. Peelaerts, L. Bousset, V. Baekelandt, and R. Melki, “ α -Synuclein strains and seeding in Parkinson’s disease, incidental Lewy body disease, dementia with Lewy bodies and multiple system atrophy: similarities and differences,” *Cell and Tissue Research*, vol. 373, no. 1, pp. 195–212, 2018.
- [75] K. A. Jellinger and A. D. Korczyn, “Are dementia with Lewy bodies and Parkinson’s disease dementia the same disease?,” *BMC Medicine*, vol. 16, no. 1, pp. 1–16, 2018.
- [76] M. Sezgin, B. Bilgic, S. Tinaz, and M. Emre, “Parkinson’s disease dementia and Lewy body disease,” *In seminars in neurology*, vol. 39, no. 2, pp. 274–282, 2019.
- [77] L. Zhang, B. Shi, M. Cao, S. Zhang, Y. Dai, and Y. Zhu, “Identifying EEG responses modulated by working memory loads from weighted phase lag index based functional connectivity microstates,” in *In international conference on neural information processing*, pp. 441–449, Springer, Cham, 2019, December.
- [78] G. Fraga González, D. J. Smit, M. J. W. van der Molen et al., “EEG resting state functional connectivity in adult dyslexics using phase lag index and graph analysis,” *Frontiers in Human Neuroscience*, vol. 12, p. 341, 2018.
- [79] D. A. Blanco and A. Díaz-Méndez, “Pain detection with EEG using phase indexes,” in *2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*, pp. 48–51, IEEE, Bethesda, MD, USA, 2017, November.
- [80] R. Mehraram, M. Kaiser, R. Cromarty et al., “Weighted network measures reveal differences between dementia types: an EEG study,” *Human Brain Mapping*, vol. 41, no. 6, pp. 1573–1590, 2020.
- [81] J. J. van der Zande, A. A. Gouw, I. van Steenoven, P. Scheltens, C. J. Stam, and A. W. Lemstra, “EEG characteristics of dementia with Lewy bodies, Alzheimer’s disease and mixed pathology,” *Frontiers in Aging Neuroscience*, vol. 10, p. 190, 2018.
- [82] L. R. Peraza, R. Cromarty, X. Kobeleva et al., “Electroencephalographic derived network differences in Lewy body dementia compared to Alzheimer’s disease patients,” *Scientific Reports*, vol. 8, no. 1, pp. 1–13, 2018.
- [83] M. Dauwan, M. M. Linszen, A. W. Lemstra, P. Scheltens, C. J. Stam, and I. E. Sommer, “EEG-based neurophysiological indicators of hallucinations in Alzheimer’s disease: comparison with dementia with Lewy bodies,” *Neurobiology of Aging*, vol. 67, pp. 75–83, 2018.
- [84] H. Hampel, N. Toschi, C. Babiloni et al., “Revolution of Alzheimer precision neurology. Passageway of systems biology and neurophysiology,” *Journal of Alzheimer’s Disease*, vol. 64, supplement 1, pp. S47–S105, 2018.
- [85] J. K. Grooms, G. J. Thompson, W. J. Pan et al., “Infraslow electroencephalographic and dynamic resting state network activity,” *Brain Connectivity*, vol. 7, no. 5, pp. 265–280, 2017.
- [86] A. Damborská, M. I. Tomescu, E. Honzirková et al., “EEG resting-state large-scale brain network dynamics are related to depressive symptoms,” *Frontiers in Psychiatry*, vol. 10, p. 548, 2019.
- [87] M. J. Hogan, L. Kilmartin, M. Keane et al., “Electrophysiological entropy in younger adults, older controls and older cognitively declined adults,” *Brain Research*, vol. 1445, pp. 1–10, 2012.
- [88] N. Houmani, F. Vialatte, E. Gallego-Jutglà et al., “Diagnosis of Alzheimer’s disease with electroencephalography in a differential framework,” *PLoS One*, vol. 13, no. 3, article e0193607, 2018.
- [89] H. Garn, M. Waser, M. Deistler et al., “Quantitative EEG in Alzheimer’s disease: cognitive state, resting state and association with disease severity,” *International Journal of Psychophysiology*, vol. 93, no. 3, pp. 390–397, 2014.
- [90] J. N. Ianof, F. J. Fraga, L. A. Ferreira et al., “Comparative analysis of the electroencephalogram in patients with Alzheimer’s disease, diffuse axonal injury patients and healthy controls using LORETA analysis,” *Dementia & neuropsychologia*, vol. 11, no. 2, pp. 176–185, 2017.
- [91] E. Perez-Valero, M. A. Lopez-Gordo, C. Morillas, F. Pelayo, and M. A. Vaquero-Blasco, “A review of automated techniques for assisting the early detection of Alzheimer’s disease with a focus on EEG,” *Journal of Alzheimer’s disease*, pp. 1–14, 2021.
- [92] A. Wolff, D. A. Di Giovanni, J. Gómez-Pilar et al., “The temporal signature of self: temporal measures of resting-state EEG predict self-consciousness,” *Human Brain Mapping*, vol. 40, no. 3, pp. 789–803, 2019.
- [93] J. Wang, Y. Fang, X. Wang, H. Yang, X. Yu, and H. Wang, “Enhanced gamma activity and cross-frequency interaction of resting-state electroencephalographic oscillations in patients with Alzheimer’s disease,” *Frontiers in Aging Neuroscience*, vol. 9, p. 243, 2017.
- [94] Q. Lin, M. D. Rosenberg, K. Yoo, T. W. Hsu, T. P. O’Connell, and M. M. Chun, “Resting-state functional connectivity predicts cognitive impairment related to Alzheimer’s disease,” *Frontiers in Aging Neuroscience*, vol. 10, p. 94, 2018.
- [95] D. Blackburn, Y. Zhao, M. De Marco et al., “A pilot study investigating a novel non-linear measure of eyes open versus eyes closed eeg synchronization in people with Alzheimer’s disease and healthy controls,” *Brain Sciences*, vol. 8, no. 7, p. 134, 2018.
- [96] L. Cai, X. Wei, J. Wang, H. Yu, B. Deng, and R. Wang, “Reconstruction of functional brain network in Alzheimer’s disease via cross-frequency phase synchronization,” *Neurocomputing*, vol. 314, pp. 490–500, 2018.
- [97] S. Mazza, M. Frot, and A. E. Rey, “A comprehensive literature review of chronic pain and memory,” *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 87, no. - Part B, pp. 183–192, 2018.
- [98] F. Ramzan, M. U. G. Khan, A. Rehmat et al., “A deep learning approach for automated diagnosis and multi-class classification of Alzheimer’s disease stages using resting-state fMRI

- and residual neural networks,” *Journal of Medical Systems*, vol. 44, no. 2, pp. 1–16, 2020.
- [99] V. Karami, F. Amenta, G. Noce et al., “P68-F abnormalities of cortical neural synchronization mechanisms in patients with Alzheimer’s diseases dementia: an EEG study,” *Clinical Neurophysiology*, vol. 130, no. 7, pp. e86–e87, 2019.
- [100] C. Babiloni, C. del Percio, R. Lizio et al., “Abnormalities of cortical neural synchronization mechanisms in patients with dementia due to Alzheimer’s and Lewy body diseases: an EEG study,” *Neurobiology of Aging*, vol. 55, pp. 143–158, 2017.
- [101] T. Y. Chang, Y. Yamauchi, M. T. Hasan, and C. Chang, “Cellular cholesterol homeostasis and Alzheimer’s disease: thematic review series: ApoE and lipid homeostasis in Alzheimer’s disease,” *Journal of Lipid Research*, vol. 58, no. 12, pp. 2239–2254, 2017.
- [102] D. A. Butterfield and D. Boyd-Kimball, “Oxidative stress, amyloid- β peptide, and altered key molecular pathways in the pathogenesis and progression of Alzheimer’s disease,” *Journal of Alzheimer’s Disease*, vol. 62, no. 3, pp. 1345–1367, 2018.
- [103] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, “Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state,” *Physical Review E*, vol. 64, no. 6, article 061907, 2001.
- [104] A. L. Goldberger, L. A. N. Amaral, L. Glass et al., “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [105] A. Shoeb, *Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment [Ph.D. Thesis]*, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
- [106] J. J. van der Zande, A. A. Gouw, I. van Steenoven et al., “Diagnostic and prognostic value of EEG in prodromal dementia with Lewy bodies,” *Neurology*, vol. 95, no. 6, pp. e662–e670, 2020.
- [107] F. Massa, R. Meli, M. Grazzini et al., “Utility of quantitative EEG in early Lewy body disease,” *Parkinsonism & Related Disorders*, vol. 75, pp. 70–75, 2020.
- [108] J. Schumacher, J. P. Taylor, C. A. Hamilton et al., “Quantitative EEG as a biomarker in mild cognitive impairment with Lewy bodies,” *Alzheimer’s Research & Therapy*, vol. 12, no. 1, pp. 1–12, 2020.
- [109] H. Lee, G. J. Brekelmans, and G. Roks, “The EEG as a diagnostic tool in distinguishing between dementia with Lewy bodies and Alzheimer’s disease,” *Clinical Neurophysiology*, vol. 126, no. 9, pp. 1735–1739, 2015.
- [110] M. Stylianou, N. Murphy, L. R. Peraza et al., “Quantitative electroencephalography as a marker of cognitive fluctuations in dementia with Lewy bodies and an aid to differential diagnosis,” *Clinical Neurophysiology*, vol. 129, no. 6, pp. 1209–1220, 2018.
- [111] L. Bonanni, B. Perfetti, S. Bifulchetti et al., “Quantitative electroencephalogram utility in predicting conversion of mild cognitive impairment to dementia with Lewy bodies,” *Neurobiology of Aging*, vol. 36, no. 1, pp. 434–445, 2015.
- [112] K. Engedal, J. Snaedal, P. Hoegh et al., “Quantitative EEG applying the statistical recognition pattern method: a useful tool in dementia diagnostic workup,” *Dementia and Geriatric Cognitive Disorders*, vol. 40, no. 1-2, pp. 1–12, 2015.
- [113] S. J. Colloby, R. A. Cromarty, L. R. Peraza et al., “Multimodal EEG-MRI in the differential diagnosis of Alzheimer’s disease and dementia with Lewy bodies,” *Journal of Psychiatric Research*, vol. 78, pp. 48–55, 2016.

Research Article

CST: A Multitask Learning Framework for Colorectal Cancer Region Mining Based on Transformer

Dong Sui ¹, Kang Zhang,¹ Weifeng Liu,¹ Jing Chen,² Xiaoxuan Ma,¹ and Zhaofeng Tian ²

¹School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

²Department of Laboratory and Diagnosis, Changhai Hospital, Navy Medical University, Shanghai 200433, China

Correspondence should be addressed to Dong Sui; suidong@bucea.edu.cn and Zhaofeng Tian; tian_zhao_feng@163.com

Received 30 July 2021; Accepted 30 August 2021; Published 11 October 2021

Academic Editor: Qiushi Zhao

Copyright © 2021 Dong Sui 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.

Colorectal cancer is a high death rate cancer until now; from the clinical view, the diagnosis of the tumour region is critical for the doctors. But with data accumulation, this task takes lots of time and labor with large variances between different doctors. With the development of computer vision, detection and segmentation of the colorectal cancer region from CT or MRI image series are a great challenge in the past decades, and there still have great demands on automatic diagnosis. In this paper, we proposed a novel transfer learning protocol, called CST, that is, a union framework for colorectal cancer region detection and segmentation task based on the transformer model, which effectively constructs the cancer region detection and its segmentation jointly. To make a higher detection accuracy, we incorporate an autoencoder-based image-level decision approach that leverages the image-level decision of a cancer slice. We also compared our framework with one-stage and two-stage object detection methods; the results show that our proposed method achieves better results on detection and segmentation tasks. And this proposed framework will give another pathway for colorectal cancer screen by way of artificial intelligence.

1. Introduction

Colorectal cancer is a common malignancy tumour worldwide, which has ranked the third position as the most common cancer and the second cause of cancer-related deaths worldwide. Also, the 2021 analysis observes that the diagnosed patients are rising in the crowd younger than 50 years old [1]. In China, there are more than 480,000 new cases with a higher than 30% death percentage in 2020, which increases the incidence and mortality rates rank following lung cancer [2]. In the early stages, occult blood examination and medical images were employed for clinical detection and diagnosis. These methods exhibited a productive approach for the early colorectal cancer diagnosis and can improve the survival of these patients [3]. However, mere blood examination and colonoscopy inspection could not reveal the biological morphology and tumour statutes [4, 5]. In the past decades, imaging approaches such as computed tomography (CT) and magnetic resonance imaging (MRI)

have become an effective way for colorectal cancer diagnosis; doctors can get an overall scheme of the tumour region in a comprehensive way without invasion [2, 6]. In addition, medical imaging can get detailed information of the tumour region without any physical cathartic cleansing, which has become a prevalent screening guideline [7, 8].

Medical image processing and analysis have achieved remarkable progress in the past several years, especially the use of deep learning, but this field is still challenging for the difficulty of acquisition and annotation in medical imaging datasets [9, 10]. In this context, transfer learning is another pathway for handling the lack of annotated medical data with small-scale data training and becomes a common protocol instead of traditional supervised learning [9, 10]. As an effective way, the transfer learning protocol in medical image processing usually employed the ImageNet pretrained deep architecture, e.g., ResNet, DenseNet, FCN, and U-Net family, and then, these models are fine-tuned on small-scale medical images to fit some certain tasks [11–15]. These

fine-tuned models yield much better results than bottom-up training strategy, especially when it confronts a small set of image samples.

However, despite 3D or 2D medicine having the same image structures, it is worth emphasizing that especially in tasks of computer vision, these medical images have distinct interest with the natural image benchmark datasets. In the case of medical data, most of the ROIs took part in a small region in the whole image, resulting in sophisticated yet hard example problems [16, 17]. Learning a medical image analysis network transferred from the natural scene usually leads to strong bias without considering the characteristic of the medical image. Thus, all of the above formations motivate developing a transfer learning protocol directly by a medical imaging dataset that can handle subtle variances in these two datasets. In this situation, ConvNets become a popular backbone in many medical image and nature image processing protocols. But for the 3D image tasks, the traditional ConvNets cannot perform long-range dependency modelling. In addition, applications of pyramid ConvNets and some attention mechanisms can facilitate the processing of these sequential image datasets [18]. However, most of these methods have not focused on long-range medical image pipeline on multitask tumour region analysis.

Transformer models such as BERT and DERT have successfully achieved the state of the art in nature language processing and computer vision fields [19–22]. Due to its ability to learn long-range dependencies from input tokens, the self-attention mechanism can model the dependency among the input tokens. The famous vision transformer (ViT) have achieved comparable performance with the traditional deep learning model such as the CNN model on image recognition tasks [23]. But all of these models have to be trained on a large-scale dataset. DeiT (Data-efficient Image Transformers) is the first transformer-based model adapted by mid-sized datasets [24].

In this work, we propose a novel transfer learning protocol, called *CST*, which is a union framework for colorectal cancer region detection and segmentation task based on the transformer model, which effectively constructs the cancer region detection and its segmentation jointly. To make a higher detection accuracy, we incorporate an autoencoder-based image-level decision approach that leverages the image-level decision of a cancer slice. First of all, we pretrain an encoder-decoder architecture for cancer/normal image slice representation that generates the encoding vectors of the original input image slices as the image-level label. Then, another transformer-based global to local architecture is pre-trained by our colorectal medical image datasets for tumour region detection and segmentation. To validate the effectiveness of our proposed framework, we test the model output on the collected colorectal cancer MRI image series and achieved remarkable performances compared with other traditional methods.

In summary, our main contributions are as follows:

- (i) We propose a novel framework for colorectal cancer region detection and segmentation. Our framework

provides a more flexible pathway for tumour region mining

- (ii) We combine the traditional autoencoder and transformer architecture together for the multitask framework for the final decision
- (iii) We evaluate the proposed method on the colorectal cancer MRI image dataset, and our method has achieved a better result on tumour region detection and segmentation

2. Related Works

Convolution neural networks with their excellent feature representation ability have raised a revolution in the nature language processing field, as well as the computer vision and signal processing fields [25]. Position-sensitive tasks such as semantic segmentation that contains several parts of ROIs have been well represented by using convolutional encoder-decoder architectures [26, 27]. The main aim of convolutional operation is to catch local texture feature information by the convolution kernel, and more layers and stride kernel in the receptive field can extend the capture range during downsampling; in this way, the model can capture global to local information explicitly. However, the size and shape of these kernels are usually of fixed size and cannot adapt to all the input range [28–30].

Recent advance in transformer-based architecture with a self-attention mechanism and the ability of long-range modelling has achieved the state of the art in natural language processing and computer vision [31]. Vision transformer (ViT) can treat the whole image into several patches and feed into the transformer pipeline as tokens. The simple application of the transformer has shown excellent results compared with the traditional CNN model [23]. However, the computation cost and large-scale dataset are the fatal drawbacks to competing with the convolutional neuronal network.

In the medical image processing paradigm, few annotated clinical data cannot generate efficient models and have to use the ImageNet pretrained model for the downstream tasks [32]. In practice, most of the prevalent methods use their pretrained weights for the medical ROI detection, such as ResNet and DenseNet, and fine-tune the higher layers on some special tasks. But most of these strategies is limited to applying for new datasets [11–15]. In a word, the pretrained model on ImageNet and other datasets by fully supervised learning paradigm have to be severed with massive annotated datasets to fit the downstream transfer learning. In another way, a self-supervised learning framework can get a suitable result by using few or no need of labelled datasets; this has gained great attention in medical image analysis recently [33, 34]. Furthermore, the self-supervised learning paradigm has attracted great attention in the medical image analysis field [35]. The critical challenge for self-supervised learning is how to define a suitable proxy task from the unlabelled data. But most of these proxy tasks have exhibited less use on medical image-related tasks.

3. Material and Methods

3.1. Dataset. We construct a novel dataset for this research, which contains 375 cases of colorectal cancer tumour MRI image datasets from 2013 to 2020, which contains 289 CRM negatives and 86 CRM positives. For segmentation and detection tasks, we also collect 375 cases of colorectal cancer negative samples as the negative samples for deep architecture training. Our collaborator labels the image slices with mask and bounding box separately as the final ground truth. Then, the dataset was divided into training, testing, and validation sets for the network training and evaluation. The main aims of this dataset are to collect for colorectal cancer region detection and segmentation, and to follow this aim, we construct the framework in this manuscript to perform them and prepared for the clinical applications. Figure 1 shows the details of the labelled tumour region about CRM negative and positive, respectively.

3.2. Multitask Framework. The motivation of our work is to construct a multitask framework that combines tumour region detection and segmentation tasks. In this section, we illustrate the overall framework of our proposed CST framework as shown in Figure 2. Our framework is divided into two pipelines, the tumour region detector and the tumour segmentation pipeline. In the detection pipeline, we first generate the region proposal of the input images, and an encoder-decoder model is used for the position encoder as the DETR input. In the segmentation pipeline, we use image patches as the input and project to a sequence of embedding for the transformer, and the class embedding is used for the final mask prediction.

3.2.1. Detector Pipeline. In this part, we start from the region proposals generated from an input medical image with H and W in height and width as the initial image $x_{\text{rpn}} \in \mathbb{R}^{3 \times H_0 \times W_0}$ with 3 channels, H_0 in height and W_0 width of the RPN. We choose a conventional CNN backbone to generate the lower resolution activation map $f \in \mathbb{R}^{C \times H' \times W'}$; typically, the values of $C = 2048$; H' and W' are resized as the initial input of the setting $H' = H/32$, $W' = W/32$. Unfortunately, the position encoding method in the original image only reflects the location of the pixels in the column and row, but the input position in our pipeline is from the random selected RPNs, so in this part, we pretrained an autoencoder for the position representation and we added this coding to the traditional position encoding with the anchor position together.

For the transformer encoder, we use a 1×1 convolution to reduce the dimension of the activation map into a d -dimension vector; for the input of the transformer, the feature map is collapsed into a 1-dimension vector with a $d \times HW$ feature map. Each encoder layer is adopting the standard setting as is stated in the DETR [19–22]. For the transformer decoder, it follows the standard architecture of the transformer, and the model can decode the tumour region at each decoder layer. Each object/RPN is transformed into an output embedding by the decoder. They are decoded into bounding box coordinates and tumour/nontumour class

labels by the following feed forward network (FFN). The FFN is a 3-layer perceptron with ReLU, hidden dimension d , and a linear projection layer. It can predict the normalized centre, height, and width of the tumour bounding box. In addition, the tumour and nontumour class of the detected bounding box is predicted by a SoftMax function.

The loss function of the tumour detection part is to optimize the lowest cost:

$$\hat{\sigma} = \arg \min_{\sigma \in \mathcal{P}_N} \sum_i^N L_{\text{match}}(y_i, \hat{y}_{\sigma(i)}), \quad (1)$$

$$L_{\text{match}}(y_i, \hat{y}_{\sigma(i)}) = L_{\text{box}}(b_i, \hat{b}_{\sigma(i)}) - \hat{p}_{\sigma(i)}(c_i). \quad (2)$$

For efficient computing, we choose optimal assignment with the Hungarian algorithm to accelerate the training process. Here, we use Hungarian loss for all pairs matched, and the object detector loss is defined like similarly loss; the total Hungarian loss is defined as follows:

$$L_{\text{Hungarian}}(y, \hat{y}) = \sum_{i=1}^N \left[-\log \hat{p}_{\sigma(i)}(c_i) + L_{\text{box}}(b_i, \hat{b}_{\sigma(i)}) \right]. \quad (3)$$

$$L_{\text{box}}(b_i, \hat{b}_{\sigma(i)}) = \lambda_{\text{iou}} L_{\text{iou}}(b_i, \hat{b}_{\sigma(i)}) + \lambda_{L1} \|b_i - \hat{b}_{\sigma(i)}\|_1, \quad (4)$$

where for the prediction with index $\sigma(i)$, $y_i = (c_i, b_i)$ is the labelled ground truth, and we define the probability of the tumour region; we define c_i as $\hat{p}_{\sigma(i)}(c_i)$ and the predicted bounding box as $\hat{b}_{\sigma(i)}$, $\lambda_{\text{iou}} \in \mathbb{R}$ and $\lambda_{L1} \in \mathbb{R}$.

3.2.2. Segmenter Pipeline. The segmenter part is based on full transformer-based architecture for pixel-level class annotation. As shown in the upper part of Figure 2, we model the sequence of patches by using a transformer encoder and a point-wise linear mapping or a mask transformer. The whole pipeline is trained end-to-end with cross-entropy loss per pixel.

In the encoder part, we first split the input image into a sequence of identical size patches, and each patch is flattened into a 1-dimension vector to produce a sequence of patch embeddings. For the position information encoding, we treat each patch as a separated part from the whole image and finally add position information to the original patch position. After that, the traditional transformer encoder is employed for the sequential information encoding with a multihead self-attention block. For the decoder part, it first learns to map patch-level encodings from the encoder to patch-level class scores; following that, these scores are unsampled to pixel-level scores by bilinear interpolation. The whole mask transformer is illustrated in the lower part of Figure 2.

For the mask transformer, we use a set of K learnable class embeddings in the decoder; in our pipeline, K is 2. Each class embedding is randomly initialized and assigned to a single class so as to generate the class mask. At last, the class

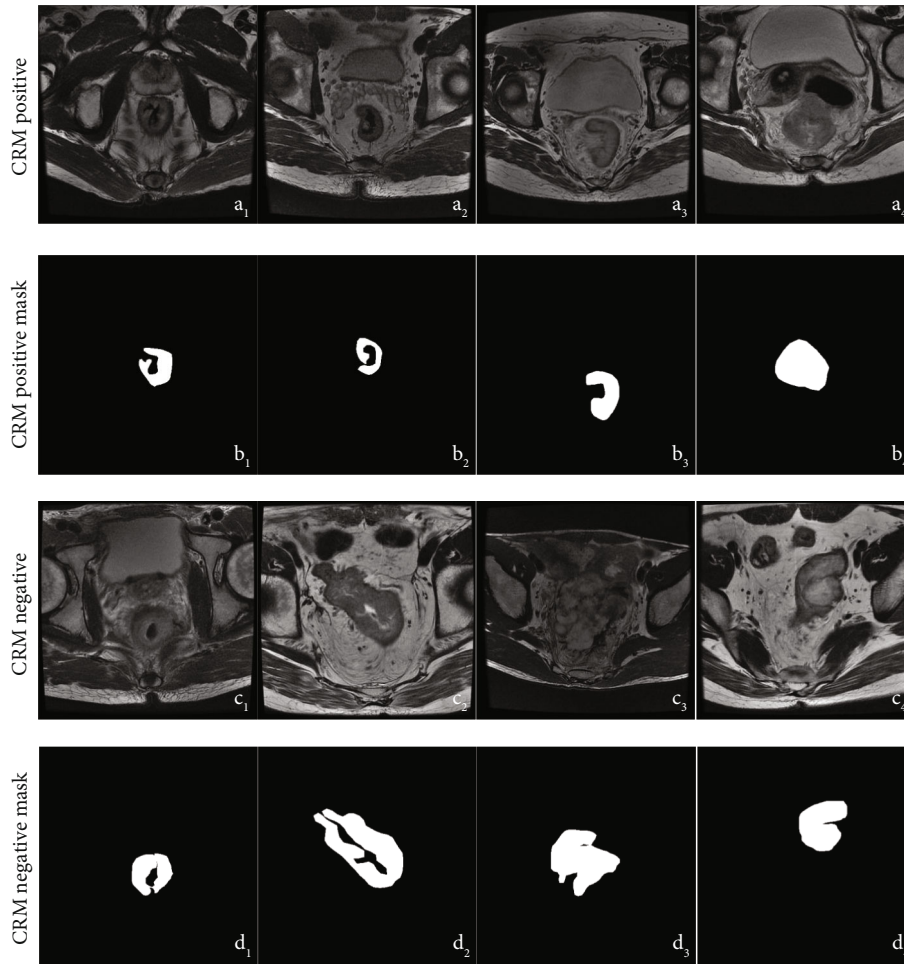


FIGURE 1: Examples of colorectal cancer in MRI images. (a1–a4, c1–c4) Are the original image slices from the MRI DICOM series; (b1–b4, d1–d4) are the tumour region mask labelled by doctors.

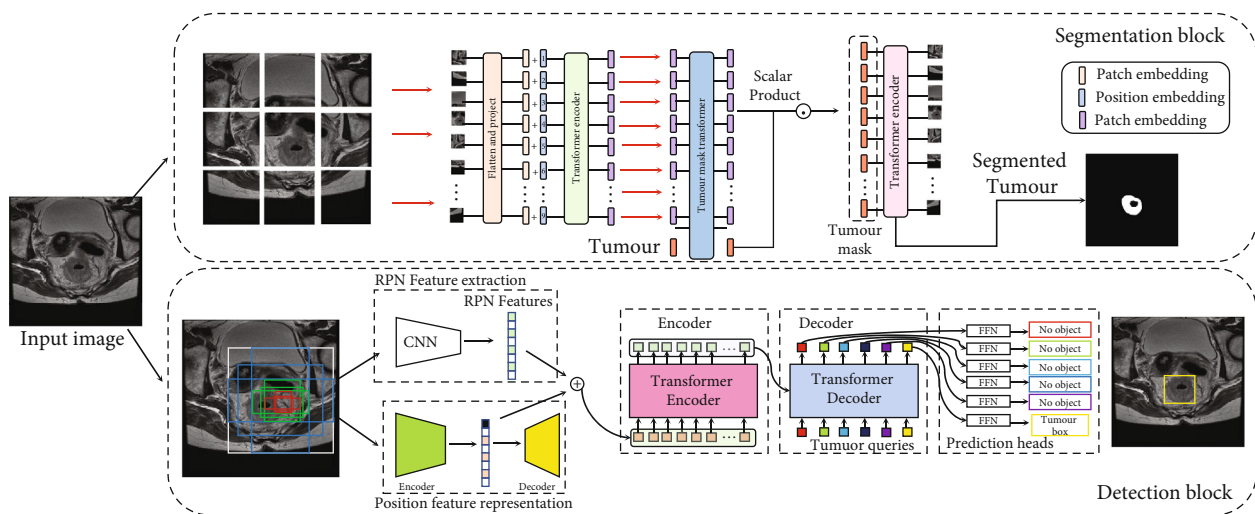


FIGURE 2: Schematic diagram of the proposed multitask learning framework for colorectal cancer region mining frame.

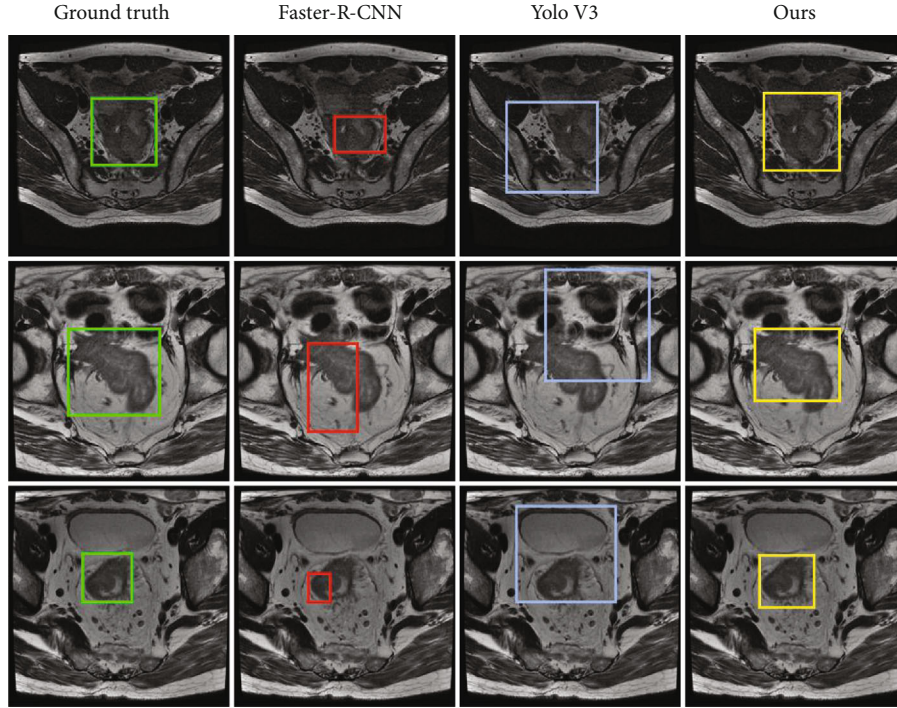


FIGURE 3: Tumour region detection results and its comparison results.

embeddings are processed jointly with patch encodings by the decoder as depicted in Figure 2. The total loss function is defined as follows:

$$L_{\text{tumour}} = - \sum_{i=1}^n t_i \log (p_i), \quad \text{for } k \text{ classes,} \quad (5)$$

where t_i is the ground truth label and p_i is the softmax probability for the i^{th} class.

In this way, we combine these loss functions to form an end-to-end train protocol, for the total loss is defined as follows:

$$L_{\text{total}} = L_{\text{match}} + L_{\text{tumour}}. \quad (6)$$

Our proposed method employs a simple process to treat the patch and tumour region segmentation jointly during the decoding phase in the segmentator; in the whole framework, we address tumour region detection and segmentation jointly and combine them into a whole framework.

4. Results and Discussion

4.1. Implementation Details. In the tumour region detection pipeline, we train DETR with AdamW optimizer with the initial transformer’s learning rate to 10^{-4} , the backbone’s to 10^{-5} , and the weight decay to 10^{-4} . We choose the transformer weights with Xavier in it and all backbone is ImageNet pretrained ResNet50 model for the basic architecture [36, 37].

In the tumour segmentation pipeline, the architecture is based on the vision transformer (ViT), and the head size of the multihead self-attention block is fixed to 64, other parameters are set as the default of the ViT model, and the input patches are with the same size [23, 24]. The segmentation model is pretrained on ImageNet; ViT is pretrained on ImageNet with random cropping. Following that, we fine-tune the pretrained models for the tumour region segmentation task and the pixel-wise cross-entropy loss without weight rebalancing. In the training phase, the SGD optimizer with a base learning rate 0.0001 and weight decay 0 is set in the initial training paradigm.

Here, we choose the standard evaluation method of tumour detection and segmentation. The Jaccard index is used for evaluating the ground truth bounding box and the predicted bounding box variances, and formally, the IoU measures the overlap between the ground truth box and the predicted box over their union. The total IoU is defined as follows:

$$\text{IoU}_{\text{pred}}^{\text{truth}} = \frac{\text{truth} \cap \text{pred}}{\text{truth} \cup \text{pred}}. \quad (7)$$

For comparisons with other methods, the results of our framework and other methods are reported in terms of recall, precision, and f1-measure values as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (9)$$

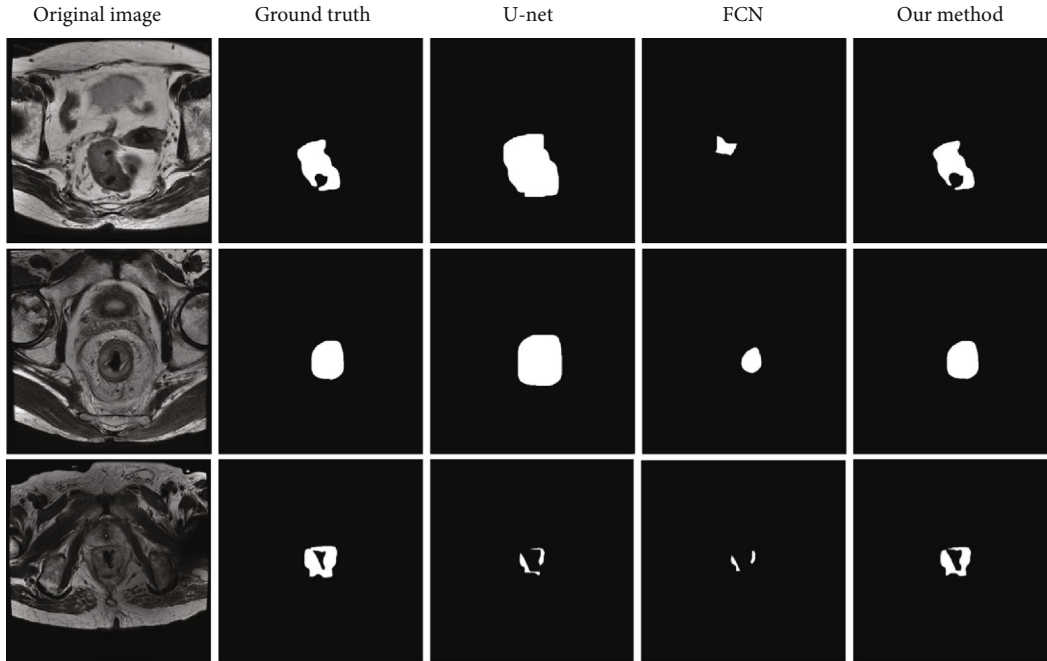


FIGURE 4: Basic rocket ship design. The rocket ship is propelled with three thrusters and features a single viewing window. The nose cone is detachable upon impact.

$$F1 = \frac{2PR}{P+R}. \quad (10)$$

Our framework is performed on the environment of Ubuntu 14.04 with Inter Core i9 platform, 32GB RAM, and 2080Ti GPU×2 and is based on the PyTorch platform, CUDA v10.0, cuDNN v8.0.

4.2. Experiment Results. For the model training, we first divide the two pipelines separated, and training them individually, it takes about 72 hours and 103 hours for the detector and segmentator. For the restriction of our GPUs, we have not extended other comparisons with other transformer-based methods.

For tumour detection, we divide the dataset into training and validation sets. For the same baseline, these methods cover the one-stage and two-stage object detection pipelines. The most prevalent method Faster-RCNN and Yolo-V3 are chosen as the test bed for the final comparisons. Figure 3 shows the detection results of the three methods, our proposed pipeline can cover most of the tumour regions, and the bounding box can converge to the tumour boundary accurately. The Faster-RCNN model is a popular two-stage object detection method, which can catch the tumour region, but it is affected by the changes of the background; this is largely because the tumour regions have the same texture as other organs. The Yolo-v3 is a popular two-stage object detection method and has been employed for many object detection and location tasks. We have evaluated this method on this dataset, the results are shown in column 3 of Figure 3, it shows that Yolo-v3 has detected the tumour region, but the result exhibits that this method usually covers the tumour region and the neighbour organs together and

TABLE 1: Tumour region detection results covered by the study.

Methods	CRM+ (%)	CRM- (%)	Average (%)	Total (%)
Faster-RCNN	67.1	62.3	64.7	65.6
Yolo-v3	43.4	37.6	40.5	41.2
Ours	87.5	89.1	88.3	88.6

TABLE 2: Tumour region segmentation accuracy covered by the study.

Methods	CRM+ (%)	CRM- (%)	Average (%)	Total (%)
U-Net	82.1	81.7	81.9	81.8
FCN	67.2	66.4	66.8	66.5
Ours	91.2	90.6	90.9	91.1

this is largely affected by the distinct boundary of these regions.

For the tumour segmentation, we have the same separation of dataset like the detection pipeline. We set the U-Net and FCN as the baseline for the comparison, the U-Net is a specific method for medical image analysis, and FCN have been greatly used in the nature image segmentation; for this reason, we have listed the comparison result with these two methods and to exhibit our framework's advantages. In this part, the U-Net model has better results than the FCN model, and it can catch the tumour region in high contract images but less in low contract slices. The FCN model usually needs an intensive training protocol on large image data, but in this program, the dataset is less than those, so it cannot get better results. Compared with these two methods, our proposed method has achieved an excellent result on the image segmentation tasks. And the results are shown in Figure 4.

For a better comparison, we also compared the accuracy of these two methods, and our proposed framework also achieves a better result on the dataset (see Tables 1 and 2).

5. Conclusions

In this paper, we propose a novel transfer learning framework, CST. We combine the colorectal cancer region detection and segmentation task jointly and fine-tuned a transformer-based model to perform these tasks. For higher accuracy, we incorporate image-level information into the final cancer region detection, the results demonstrate that the proposed framework can handle these tasks well, and the comparison results have shown that our method has achieved better accuracy than the traditional methods such as CNN. In this way, the proposed framework explores a new protocol for colorectal cancer information mining. In future works, we mainly focus on how to use few samples to achieve a better result.

Data Availability

The DICOM data used to support the findings of this study were supplied by Zhaofeng Tian and Jing Chen under license and so cannot be made freely available. Requests for access to these data should be made to Dong Sui.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Dong Sui and Jing Chen have equal contributions to this manuscript.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant Nos. 61702026 and 62031003), the Pyramid Talent Training Project of Beijing University of Civil Engineering and Architecture (Grant No. JDYC20200318), the National Key Research and Development Program of China (Grant No. 2020YFF0305504), and the Doctoral Research Initiation Fund of Beijing University of Civil Engineering and Architecture (No. X20040).

References

- [1] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] J. Jin, "Screening for colorectal cancer," *JAMA*, vol. 325, no. 19, p. 2026, 2021.
- [3] J. S. Lin, L. A. Perdue, N. B. Henrikson, S. I. Bean, and P. R. Blasi, "Screening for colorectal cancer: updated evidence report and systematic review for the US Preventive Services Task Force," *JAMA*, vol. 325, no. 19, pp. 1978–1998, 2021.
- [4] S. J. Mehta, A. M. Morris, and S. S. Kupfer, "Colorectal cancer screening starting at age 45 years—ensuring benefits are realized by all," *JAMA Netw Open*, vol. 4, no. 5, p. e2112593, 2021.
- [5] K. Ng, F. P. May, and D. Schrag, "US Preventive Services Task Force recommendations for colorectal cancer screening: forty-five is the new fifty," *JAMA*, vol. 325, no. 19, pp. 1943–1945, 2021.
- [6] A. B. Knudsen, C. M. Rutter, E. F. P. Peterse et al., "Colorectal cancer screening: an updated modeling study for the US Preventive Services Task Force," *JAMA*, vol. 325, no. 19, pp. 1998–2011, 2021.
- [7] S. Jensch, S. Bipat, J. Peringa et al., "CT colonography with limited bowel preparation: prospective assessment of patient experience and preference in comparison to optical colonoscopy with cathartic bowel preparation," *European Radiology*, vol. 20, no. 1, pp. 146–156, 2010.
- [8] S. Wilson and J. D. Thompson, "Comparison of two Meglumine-Diatrizoate based bowel preparations for computed tomography colonography: comparison of patient symptoms and bowel preparation quality," *Radiography (Lond)*, vol. 26, no. 4, pp. e290–e296, 2020.
- [9] S. B. Desai, A. Pareek, and M. P. Lungren, "Deep learning and its role in COVID-19 medical imaging," *Intelligence-Based Medicine*, vol. 3-4, p. 100013, 2020.
- [10] H. M. Ahmad, M. J. Khan, A. Yousaf, S. Ghuffar, and K. Khurshid, "Deep learning: a breakthrough in medical imaging," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 16, no. 8, pp. 946–956, 2020.
- [11] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: a novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustainable Cities and Society*, vol. 65, p. 102600, 2021.
- [12] Y. D. Zhang, S. C. Satapathy, X. Zhang, and S. H. Wang, "COVID-19 diagnosis via DenseNet and optimization of transfer learning setting," *Cognitive Computation*, pp. 1–17, 2021.
- [13] X. Li, Q. Dou, H. Chen et al., "3D multi-scale FCN with random modality voxel dropout learning for intervertebral disc localization and segmentation from multi-modality MR images," *Medical Image Analysis*, vol. 45, pp. 41–54, 2018.
- [14] Y. Ma, P. Feng, P. He et al., "Segmenting lung lesions of COVID-19 from CT images via pyramid pooling improved Unet," *Biomedical Physics & Engineering Express*, vol. 7, no. 4, 2021.
- [15] M. Lei, J. Li, M. Li, L. Zou, and H. Yu, "An improved UNet++ model for congestive heart failure diagnosis using short-term RR intervals," *Diagnostics (Basel)*, vol. 11, no. 3, 2021.
- [16] T. Hassanzadeh, D. Essam, and R. Sarker, "2D to 3D evolutionary deep convolutional neural networks for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 2, pp. 712–721, 2021.
- [17] V. Gupta, M. Demirel, M. Bigelow et al., "Performance of a deep neural network algorithm based on a small medical image dataset: incremental impact of 3D-to-2D reformation combined with novel data augmentation, photometric conversion, or transfer learning," *Journal of Digital Imaging*, vol. 33, no. 2, pp. 431–438, 2020.
- [18] O. Petit, N. Thome, and L. Soler, "Iterative confidence relabeling with deep ConvNets for organ segmentation with partial labels," *Computerized Medical Imaging and Graphics*, vol. 91, p. 101938, 2021.

- [19] D. Song, B. Fu, F. Li et al., “Deep relation transformer for diagnosing glaucoma with optical coherence tomography and visual field function,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2392–2402, 2021.
- [20] Z. Li, Z. Zhang, H. Zhao et al., “Text compression-aided transformer encoding,” *IEEE Trans Pattern Anal Mach Intell*, p. 1, 2021.
- [21] J. Wang, R. Chakraborty, and S. X. Yu, “Spatial transformer for 3D point clouds,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2021.
- [22] Z. Zhang, Y. Wu, J. Zhou, S. Duan, H. Zhao, and R. Wang, “SG-Net: syntax guided transformer for language representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2020.
- [23] A. Srinivas, T. Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, “Bottleneck transformers for visual recognition,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16519–16529, 2021.
- [24] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, “Training data-efficient image transformers & distillation through attention,” *In International Conference on Machine Learning*, pp. 10347–10357, 2021.
- [25] Y. Gao, C. Liu, and L. Zhao, “Multi-resolution path cnn with deep supervision for intervertebral disc localization and segmentation,” *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 2019.
- [26] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, Cham, October 2015.
- [27] L. Yang, Y. Zhang, Z. Zhao et al., “Boxnet: deep learning based biomedical image segmentation using boxes only annotation,” *ArXiv abs/1806.00593*, 2018.
- [28] Y. Gao, R. Huang, Y. Yang et al., “FocusNetv2: imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images,” *Medical Image Analysis*, vol. 67, p. 101831, 2021.
- [29] D. Lin, Y. Ji, D. Lischinski, D. Cohen-Or, and H. Huang, “Multi-scale context intertwining for semantic segmentation,” *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
- [30] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890, Honolulu, HI, USA, July 2017.
- [31] A. Vaswani, N. Shazeer, N. Parmar et al., “Attention is all you need,” *In Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.
- [32] H. C. Shin, H. R. Roth, M. Gao et al., “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [33] P. Liu, M. R. Lyu, I. King, and J. Xu, “Learning by distillation: a self-supervised learning framework for optical flow estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [34] H. Xie, Y. Lei, T. Wang et al., “High through-plane resolution CT imaging with self-supervised deep learning,” *Physics in Medicine & Biology*, vol. 66, no. 14, p. 145013, 2021.
- [35] Q. Lu, Y. Li, and C. Ye, “Volumetric white matter tract segmentation with nested self-supervised learning using sequential pretext tasks,” *Medical Image Analysis*, vol. 72, p. 102094, 2021.
- [36] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” *Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings*, 2010.
- [37] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” *Proceedings of the International Conference on Learning Representations (ICLR) 2019*, 2019.

Review Article

Privacy Protection and Secondary Use of Health Data: Strategies and Methods

Dingyi Xiang^{1,2} and Wei Cai³

¹Internet Rule of Law Institute, East China University of Political Science and Law, Shanghai, China

²Humanities and Law School, Northeast Forest University, Harbin, Heilongjiang, China

³Beidahuang Information Company, Harbin, Heilongjiang, China

Correspondence should be addressed to Dingyi Xiang; 147689865@qq.com

Received 24 July 2021; Revised 16 September 2021; Accepted 18 September 2021; Published 7 October 2021

Academic Editor: Lei Zhang

Copyright © 2021 Dingyi Xiang and Wei Cai. 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.

Health big data has already been the most important big data for its serious privacy disclosure concerns and huge potential value of secondary use. Measurements must be taken to balance and compromise both the two serious challenges. One holistic solution or strategy is regarded as the preferred direction, by which the risk of reidentification from records should be kept as low as possible and data be shared with the principle of minimum necessary. In this article, we present a comprehensive review about privacy protection of health data from four aspects: health data, related regulations, three strategies for data sharing, and three types of methods with progressive levels. Finally, we summarize this review and identify future research directions.

1. Introduction

The rapid development and application of multiple health information technologies enabled medical organizations to store, share, and analyze a large amount of personal medical/health and biomedical data, of which the majority are electronic health records (EHR) and genomic data. Meanwhile, the emerging technologies, such as smart phones and wearable devices, also enabled third-party firms to provide many kinds of complementary mHealth services and collect huge tons of consumer health data. Health big data has already been the most important big data for its serious privacy disclosure concerns and huge potential value of secondary use.

Health big data stimulated the development of personalized medicine or precision medicine. Empowered by health informatics and analytic techniques, secondary use of health data can support clinical decision making; extract knowledge about diseases, genetics, and medicine; improve patients' healthcare experiences; reduce healthcare costs; and support public health policies [1–3]. On the other side of the coin, health data contains much personal privacy and confidential

information. For the guidance of protecting health-related privacy, the Health Insurance Portability and Accountability Act (HIPAA) of the US specifies 18 categories of protected health information (PHI) [4]. The heavy concerns about privacy disclosure much hinder secondary use of health big data. Much efforts tried to balance between privacy management and health data secondary use from both the legislation side [5] and the technology side [6, 7]. But for much more circumstances, a perfect balance is difficult to achieve; instead, a certain tradeoff or compromise must always be made. Recently, COVID-19 may perfectly illustrate the conundrum between protecting health information and ensuring its availability to meet the challenges posed by a significant global pandemic. In this ongoing battle, China and South Korea have mandated public use of contact tracing technologies, with few privacy controls; other countries are also adopting contact tracing technologies [7].

The direct and also important strategy to balance both issues is reusing health data under the premise of protecting privacy. The most primary idea is to share deidentified health data by removing 18 specified PHI. Based on deidentified health data, machine learning and data mining can be

used for knowledge extraction or learning health system building for the purpose of analyzing and improving care, whereby treatment is tailored to the clinical or genetic features of the patient [8]. However, transforming data or anonymizing individuals may minimize the utility of the transferred data and lead to inaccurate knowledge [9]. This tradeoff between privacy and utility, also accuracy, is the center issue of sensitive data secondary usage [10]. Deidentification refers to a collection of techniques devised for removing or transforming identifiable information into nonidentifiable information and also introducing random noise into the dataset. By deidentification, privacy protection will be leveraged, but the outcome of analysis may be not exact, rather an approximation. To reconcile this conflict, the privacy loss parameter, also called privacy budget, was proposed to tune the tradeoff between privacy and accuracy: by changing the value of this parameter, more or less privacy resulting in less or more accuracy, respectively [11]. Furthermore, deidentified data may become reidentifiable through data triangulation from other datasets, which means that the privacy harms of big health data arise not merely in the collection of data but in their eventual use [12]. Just deidentification is far from needed. Instead, a holistic solution is the right direction, by which the risk of reidentification from records should be kept as low as possible and data be shared with the principle of minimum necessary [13]. For the minimum necessary, user-controlled access [6, 14] and secure network architecture [15] can be a practical implementation. For effective reusing health data while reducing the risk of reidentification, attempts in three aspects can be applicable references, that is, risk-mitigation methods, privacy-preserving data mining, and distributed data mining without sharing out data.

The remainder of this paper is organized as follows. Section 2 describes the scope of health data and its corresponding category. Section 3 summarizes regulations about privacy protection of health data in several countries. Section 4 concisely reviews two strategies for privacy protection and secondary use of health data. Section 5 reviews three aspects of tasks and methods for privacy preservation and data mining the primary tasks of data mining. Section 6 concludes this study.

2. Health Data and Its Category

Generally speaking, any data associated with users' health conditions can be viewed as health data. The most important health data is clinical data, especially electronic medical records (EMR), produced by different level hospitals. With the development of health information technology and the popularization of wearable health device, vast amounts of health-relevant data, such as monitored physiological data and diet or exercise data, are collected from individuals and entities elsewhere, both passively and actively. According to the review article by Deven McGraw and Kenneth D. Mandl, health-relevant data can be classified into four categories [7]. In this research, we focus on the first two categories of data, which are directly related to users' health and privacy.

Category 1. Health data generated by healthcare system. This type of data is clinical data and is recorded by clinical professionals or medical equipment when a patient gets

healthcare service in a hospital or clinic. Clinical data includes EMR, prescriptions, laboratory data, pathology images, radiography, and payor claims data. Patients' historical condition and current condition are recorded for treatment requirement. For making better health service for patients, it is important to track patients' lifelong clinical data and make clinical data sharing among different healthcare providers. Personal health record (PHR) was proposed to integrate patients' cross-institutions and lifelong clinical data [16]. This type of health data is generated and collected routinely in the process of healthcare, with the explicit aim that those data be used for the purpose of analyzing and improving care. For the purpose of clinical treatment, and also because of consumers' firm trust on healthcare experts and institutions, clinical data contains a high degree of health-related privacy. Therefore, the majority of health privacy laws mainly cover the privacy protection of clinical data [7]. Under the constraints of health privacy laws, tons of clinical data have been restricted only for internal use in medical institutions. Meanwhile, the clinical data is also extremely valuable for secondary usage since the data is created by professional experts and is direct description of consumers' health conditions. The tradeoff between utility and privacy of this type of health data has been one of the most important issues in the age of medical big data.

Category 2. Health data generated by consumer health and wellness industry. This type of health data is an important complementation to clinical data. With the widespread application of new-generation information technology, such as IoT, mHealth, smart phone, and wearable device, consumers' health attitude has greatly changed from passive treatment to active health. Consumers' health data can be generated through wearable fitness tracking devices, medical wearables such as insulin pumps and pacemakers, medical or health monitoring apps, and online health service. These health data can include breath, heart rate, blood pressure, blood glucose, walking, weight, diet preference, position, and online health consultation. These products or services and health data play important role in consumers' daily health management, especially for chronic disease patients. This area has gained more and more focus from industry and academia. Consumer health informatics is the representative direction [17]. This type of nontraditional health-relevant data, often equally revealing of health status, is in widespread commercial use and, in the hands of commercial companies, yet often less accessible by providers, patients, and public health for improving individual and population health [18]. These big health data are scattered across institutions and intentionally isolated to protect patient privacy. For this type of health data, integration and linking at individual level are an extra challenge except for the utility-privacy tradeoff.

Table 1 summarizes the two categories of health data and their comparative features.

3. Regulations about Privacy Protection of Health Data

Personal information and health-relevant data are necessary to record in order to provide regular health service.

TABLE 1: Summarization of clinical data and consumer health data.

	Category 1: clinical data	Category 2: consumer health data
Generated/record by	Healthcare system Clinical professionals Medical equipment	Wearable device (wristband, watch) Medical wearable Health App
Data detail	Name, id, age, address, phone, medical history, family history, conditions, laboratory test, treatments, prescriptions, etc.	Name, id, phone, address, position, age, weight, heart rate, breath, blood pressure, blood glucose, exercise data, diet preference, online health consultation, etc.
Data characteristics	Discrete but more professional, more clinical information and more privacy, stored in healthcare system, passive	Continuous but less standardization, more health information, privacy tend to be ignored, stored by different providers, active, vast amounts

Meanwhile, personal information and health-relevant data are closely associated with user privacy and confidential information. Therefore, several important privacy protection-related regulations or acts are published to guide health data protection and reuse. Modern data protection law is built on “fair information practice principles” (FIPPS) [19].

The most referenced regulation is Health Insurance Portability and Accountability Act (HIPAA) [4]. HIPAA was created primarily to modernize the flow of healthcare information, stipulate how personally identifiable information maintained by the healthcare and healthcare insurance industries should be protected from fraud and theft, and address limitations on healthcare insurance coverage. The HIPAA Safe Harbor (SH) rule specifies 18 categories of explicitly or potentially identifying attributes, called protected health information (PHI), that must be removed before the health data is released to a third party. HIPAA also covers electronic PHI, ePHI. This includes medical scans and electronic health records. A full list of PHI elements is provided in Table 2. PHI elements in Table 2 only cover identity information and do not include any sensitive attribute. That is, HIPAA does not provide guidelines on how to protect sensitive attribute data; instead, the basic idea of the HIPAA SH rule is to protect privacy by preventing identity disclosure. However, other sensitive attributes may still uniquely combine into a quasi-identifier (QI), which can allow data recipients to reidentify individuals to whom the data refer. Therefore, a strict implementation of the SH rule, however, may be inadequate for protecting privacy or preserving data quality. Recognizing this limitation, HIPAA also provides alternative guidelines that enable a statistical assessment of privacy disclosure risk to determine if the data are appropriate for release [20].

The Health Information Technology for Economic and Clinical Health (HITECH) Act [21] was enacted as part of the American Recovery and Reinvestment Act of 2009 to promote the adoption and meaningful use of health information technology. Subtitle D of the HITECH Act addresses the privacy and security concerns associated with the electronic transmission of health information, in part, through several provisions that strengthen the civil and criminal enforcement of the HIPAA rules. It is complimentary with HIPAA and strengthens HIPAA’s privacy regulations.

HITECH has also widened the scope of HIPAA through the Omnibus Rule. This extends the privacy and security reach of HIPAA/HITECH to business associates. According to HIPAA and HITECH Act, much of data beyond category 1 in Table 1 is outside of the scope of comprehensive health privacy laws in the U.S.

The Consumer Data Right (CDR) [22] is coregulated by the Office of the Australian Information Commissioner (OIA) and Australian Competition and Consumer Commission (ACCC). “My Health Record System” is run to track citizen medical conditions, test results, and so on. The OIA sets out controls on how health information in a My Health Record can be collected, used, and disclosed, which corresponds to PHR integration. The Personal Information Protection and Electronic Documents Act (PIPEDA) [23] of Canada applies to all personal health data. PIPEDA is stringent and although has many commonalities with HIPAA; it goes beyond HIPAA requirements in several areas. One such area is in the protection of data generated by mobile health apps which is not strictly covered by HIPAA. PIPEDA runs to protected consumer health data. Under PIPEDA, organizations can seek implied or explicit consent, which is based on the sensitivity of the personal information collected and the reasonable data processing consent expectations of the data subject. The General Data Protection Regulation (GDPR) is a wide-ranging data protection regulation in EU, which covering health data as well as all other personal data, even they contain sensitive attributes. GDPR also has data consent and breach notification expectations and contains several key provisions, including notification, right to access, right to be forgotten, and portability. Under GDPR, organizations are required to gain explicit consent from data subjects, and individuals have the right to restriction of processing and not to be subject to automated decision-making.

China has no specific regulations for health data privacy protection. Several restriction rules to prohibit privacy disclosure scatter in China Civil Code (CCC), Medical Practitioners Act of the PRC (MPAPRC), and Regulations on Medical Records Management in Medical Institutions (RMRMMMI), which make privacy disclosure restrictions to individuals, medical practitioners, and medical institutions, respectively. CCC specifies 9 categories of personal information to be protected, including name, birthday, ID

TABLE 2: Protected health information defined by HIPAA.

Category	Description
1	Names
2	Locations
3	Dates
4	Phone number
5	Fax numbers
6	E-mail addresses
7	Social security numbers
8	Medical record numbers
9	Health plan beneficiary numbers
10	Account numbers
11	Certificate/license numbers
12	Vehicle identifiers and serial numbers
13	Device identifiers and serial numbers
14	Web Universal Resource Locators (URLs)
15	Internet Protocol (IP) address numbers
16	Biometric identifiers, including finger and voice prints
17	Full face photographic images and any comparable images
18	Any other unique identifying number, characteristics, or code

number, biometric information, living address, phone number, email address, health condition information, and position tracking information. RMRMMMI only approves reuse of health data just for medical care, teaching, and academic research. Recently, the Personal Information Protection Law of the PRC (PIPIIRC) [24] is released and will come into force on November 1, 2021. This is the first complete and comprehensive regulation on personal information protection. In this regulation, the definition of sensitive personal information and automatic decision making both involve health data, so, this regulation is applicable to privacy protection of health data. According to this regulation, secondary use of deidentified or anonymized health data for automatic decision making is permitted, and data processing consent from consumers is also required. This regulation, so far as can be foreseen, will greatly stimulate the exploitation and exploration of health big data.

According to the comparison of these data privacy relevant regulations, shown in Table 3, PIPEDA and GDPR and the newly released PIPIIRC can cover both clinical data and consumer health data, and others pay the majority of attention to clinical data. Health data need to be reused for multiple important purposes. In fact, health data processing and reusing are never absolutely prohibited in the regulations mentioned above, as long as privacy protection is achieved as the important prerequisite. In this respect, HIPAA sets Safe Harbor rules to make sure PHI be removed before the health data is released to a third party. Furthermore, PIPEDA and GDPR require consumers' consent for data processing. Regulations from China also encourage health data to be reused in certain restricted areas. As the newcomer, PIPIIRC presents a more complete and comprehensive guidance to protect and process health data.

TABLE 3: Regulations and corresponding data category.

Regulations	Category 1: clinical data	Category 2: consumer health data
HIPAA & HITECH (USA)	✓	
CDR (Australia)	✓	
PIPEDA (Canada)	✓	✓
GDPR (EU)	✓	✓
MPAPRC & RMRMMMI (China)	✓	
CCC & PIPIIRC (China)	✓	✓

4. Strategies and Framework

The exploitation of health data can provide tremendous benefits for clinical research, but methods to protect patient privacy while using these data have many challenges. Some of these challenges arise from a misunderstanding that the problem should be solved by a foolproof solution. There exists a paradox: well deidentified and scrubbed data may lose much meaningful information results in low quality, maintaining much PHI may have high risk of privacy breach. Therefore, a holistic solution, or to say a unified strategy, is needed. Three strategies are summarized in this section. The first is for clinical data and provides a practical user access rating system, and the second is majority for genomic data and designs a network architecture to address both security access and potential risk of privacy disclosure and reidentification. From a more practical starting point, the third tries to share a model without exposing any data.

TABLE 4: Health data access level categories.

Privacy level of user	Data available	Trustworthiness of user	Technical security
Obfuscated data user	Users have access to data by client-side application only	Low: only obfuscated aggregate results are available	Low: only client-side application exposed to users
Aggregated data user	Users have access to HIPAA deidentified data by client-side application only	Low: users can get exact patient counts against deidentified data	Low: but data manager assumes burden of deidentifying data
LDS data user	HIPAA-defined LDS and deidentified structured data	Medium: users can see LDS as defined by HIPAA	Medium: requires user-facing direct access to the database
Notes-enabled LDS data user	HIPAA deidentified data and deidentified narrative text	Medium: users see both LDS and narrative text that is mostly deidentified	Medium: requires user-facing direct access to the database
PHI-viewable data user	All patient data may be accessed	High: users can see all protected health information on patients	High: requires management of encryption keys

The tree strategies present solutions from different perspectives, therefore can be complementary to each other.

4.1. Strategies for Clinical Data. As for clinical data, Murphy et al. proposed an effective strategy to build a clinical data sharing platform while protecting patient privacy [6]. The proposed approach to resolving the balance between privacy management and data secondary use is to match the level of data deidentification with the trustworthiness of the data recipients, in which the more identified the data, the more “trustworthy” the recipients are required to be, and vice versa. The level of trust for a data recipient becomes a critical factor in determining what data may be seen by that person. This type of hierarchical access rating is similar to the film rating, which can accommodate the requirement and appetites of different types of audiences. Murphy et al.’s strategy sets up five patient privacy levels with three aspects of requirements: availability of the data, trust in the researcher and the research, and the security of the technical platforms. Corresponding to the privacy levels are five user role levels.

The lowest level of user is “obfuscated data user.” For this user, data are obfuscated as it is served to a client machine with possibly low technical security. Obfuscation methods try to add a random number to the aggregated counts instead of providing accurate result [25, 26]. The second level of user is “aggregated data user,” to whom exact numbers from aggregate query results are permissible. The third is “LDS data user,” who is granted to access HIPAA-defined LDS (limited dataset) and structured patient data in which PHI must be removed. The fourth is “Notes-enabled LDS data user,” who is additionally allowed to view PHI scrubbed text notes (such as discharge summary). The final level of user is “PHI-viewable data user,” who has access to all patient data.

These access level categories are summarized in Table 4.

With the guidance of health data access level categories, Murphy et al. implemented five cases in clinical research. In a realistic project, multiple use role or different access privileges must be needed to reconcile different data access requirements. Murphy et al. also provided three exemplar projects and their possible privacy level user distributions. This proposed strategy gave a complete reference for data sensitive project and also implemented a holistic approach

to patient privacy solutions in Informatics for Integrating Biology and the Bedside (i2b2) research framework [27]. The i2b2 framework is the most widespread open-source framework for exploring clinical research data-warehouses and was jointly developed by the Harvard Medical School and Massachusetts Institute of Technology to enable clinical researchers to use existing deidentified clinical data and only IRB-approved genomic data for research aims. Yet, i2b2 does not provide any specific protection mechanism for genomic data.

4.2. Strategies for Genomic Data. As for genomic data, two potential privacy threats are loss of patients’ health data confidentiality due to illegitimate data access and patients’ re-identification and resulting sensitive attribute disclosure from legitimate data access. On the basis of the i2b2 framework, Raisaro et al. [15] proposed to apply homomorphic encryption [28] to the first threat and differential privacy [29] to the second threat. Furthermore, Raisaro et al. designed a system model, consisting of two physically separated networks, from the perspective of architecture. The network architecture is shown in Figure 1. This network architecture is aimed at isolating data that is used for clinical/medical care and that is used for research activities by a few trusted and authorized individuals.

The clinical network is used for hospital’s clinical daily activities, containing clinical and genomic data of patients. This network is very controlled and protected by a firewall that blocks all incoming network traffic. Authorized users are permitted to log in.

The research network hosts i2b2 service used by researchers in their research activities. The i2b2 service is composed of an i2b2 server and a proxy server, in which a homomorphic encryption method and a differential privacy method are implemented and deployed. The i2b2 server can receive deidentified clinical data and encrypted genomic data from the clinical network and perform security data query and computation. The proxy server is devoted to support the decryption phase and the storage of partial decryption keys for homomorphic encryption. Through the research network, researchers can get authorized data via query execution module by the sequential five steps: query generation, query processing, result perturbation, result

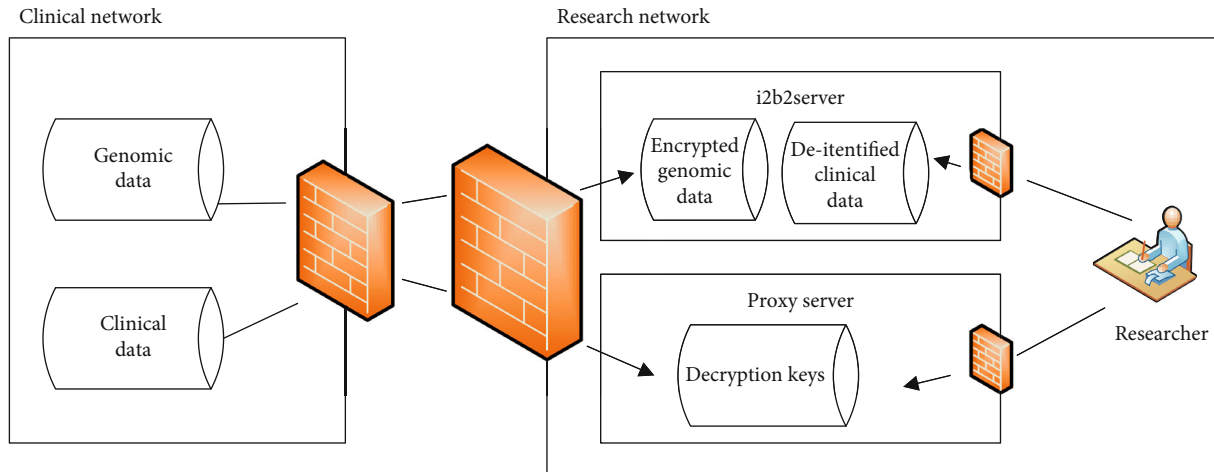


FIGURE 1: Network architecture of privacy protection for health data including genomic data.

partial decryption, and result decryption at the final user-client side.

This network architecture and its privacy-preserving solution have been successfully deployed and tested in Lausanne University Hospital and used for exploring genomic cohorts in a real operational scenario. This application is also a practicable demonstration for similar scenario. It is not a unique instance but has its counterpart. Azencott reviewed how breaches in patient privacy can occur, and recent developments in computational data protection also proposed a similar secure framework for genomic data sharing around three aspects, which includes algorithmic solutions to deidentification, database security, and user trustworthy access [3].

4.3. Strategies for Sharing Not Data but Models. Since the new paradigm of the machine learning method, namely, federated learning (FL), was first introduced in 2016 [30], has achieved a rapid development, and become a hot research topic in the field of artificial intelligence, its core idea is to train machine learning models on separate datasets that are distributed across different devices or parties, which can preserve the local data privacy to a certain extent. This development mainly benefits from the following three facts [31]: (1) the wide successful applications of machine learning technologies, (2) the explosive growth of big data, and (3) the legal regulations for data privacy protection worldwide.

The idea of federated learning is to only share the model parameters instead of the original data. By this way, many of these initiatives are based on federated models in which the actual data never leave the institution of origin, allowing researchers to share models without necessarily sharing patient data. Federated learning has inspired another important strategy to develop smart healthcare based on sensitive and private medical records which exist in isolated medical centers and hospitals. As shown in Figure 2, federated learning offers a framework to jointly train a global model using datasets stored in separate clients.

Model building of this kind has been used in real-world applications where user privacy is crucial, e.g., for hospital data or text predictions on mobile devices, and it has been stated that model updates are considered to contain less information than the original data, and through the aggregation of updates from multiple data points, original data is considered impossible to recover. Federated learning emphasizes the data privacy protection of the data owner during the model training process. Effective measures to protect data privacy can better cope with the increasingly stringent data privacy and data security regulatory environment in the future [32].

5. Tasks and Methods

Under the strategies of health data protection, specific tasks and methods about privacy and data processing can be employed and deployed. The tasks and methods can be viewed at three progressive levels. Methods in the first level are aimed at mitigating the risk of privacy disclosure, from four aspects. Methods in the second level target on data mining or knowledge extraction from deidentified or anonymized health data. No need to share health data, methods in the third level try to build a learning model or extract knowledge in a distributed manner, then share the model or knowledge.

5.1. Risk-Mitigation Methods. There are two widely recognized types of privacy disclosure [33]: identity disclosure (or reidentification) and attribute disclosure. The former occurs when illegitimate data users try to match a record in a dataset to an individual, and the latter occurs when illegitimate data users try to predict the sensitive value(s) of an individual record. According to Malin et al. [34], methods of mitigating the risk of two types of privacy disclosure can be divided into four classes: suppression, generalization, randomization, and synthetization. This perspective of method categories expects to well summarize the recent research on risk-mitigation methods.

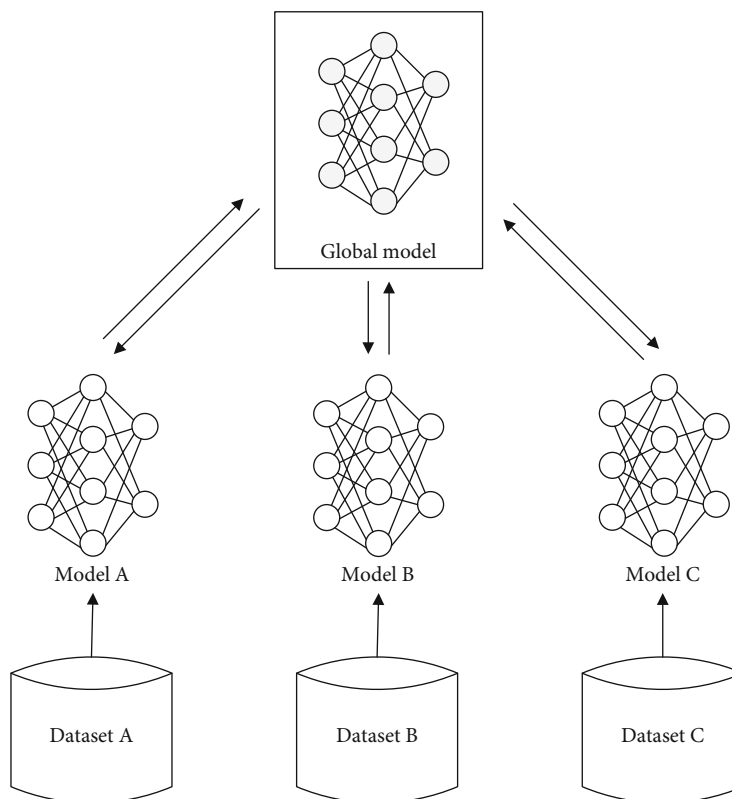


FIGURE 2: Architecture for a federated learning system.

5.1.1. Suppression Methods. Suppression methods are aimed at scrubbing (remove or mask) 18 PHI defined in HIPAA, which is the most important deidentification method. Before PHI scrubbing, the major task is to identify the PHI from health data. For structural data, PHI identification can be done easily according to data schema. For narrative data or free text, such as discharge summary or progress note, natural language processing (NLP) is the preferred technology for PHI identification. Specifically, named entity recognition (NER) is the mainstream technology used in clinical data for deidentification and medical knowledge extraction. The 18 PHI are regarded as predefined entity types, and machine learning is employed to annotate type tags for each word in a sentence, then those tags are merged, and finally, the position and type of PHI can be identified. Conditional random fields (CRFs) are the classic sequential tagging model for NER and are often applied for deidentification [35]. Meystre et al. made a systematic review of deidentification methods [36], and Uzuner et al. [37] and Deleger et al. [38] both conducted some evaluations on a certain human-annotated dataset. The identified PHI values are then simply removed from or replaced with a constant value in the released text documents, which may be inadequate for protecting privacy or preserving data quality. Li and Qin proposed a new systematic approach to integrate methods developed in both data privacy and health informatics fields. The key novel elements of the proposed approach include a recursive partitioning method to cluster medical text records and a value enumeration method to anonymize potentially

identifying information in the text data, which essentially masks the original values, to improve privacy protection and data utility [20].

For genomic data, homomorphic encryption [28] is applied to encrypting genomic data, and then, encrypted data can be shared for secondary use. Raisaro et al. employed homomorphic encryption to build a data warehouse for genomic data [15]. Kamm et al. [39] also proposed a framework for generating aggregated statistics on genomic data by using secure multiparty computation based on homomorphic secret sharing. Several other works [28, 40, 41] proposed using homomorphic encryption to protect genomic information in order to allow researchers to perform some statistics directly on the encrypted data and decrypt only the final result.

5.1.2. Generalization Methods. These methods transform data into more abstract representations. The much easier implementation is abbreviation. For instance, the age of a patient may be generalized from 1-year to 5-year age groups. Based on this type of generation, sensitive attributes can be generalized subgroup and be anonymized to some extent, which is the back idea of k -anonymity and its variations. k -anonymity seeks to prevent reidentification by stripping enough information from the released data that any individual record becomes indistinguishable from at least $(k - 1)$ other records [42]. The idea of k -anonymity is based on modifying the values of the QI attributes to make it difficult for an attacker to unravel the identity of persons in a

particular dataset while the released data remain as useful as possible. This modification is a sort of generalization, by which stored values can be replaced with semantically consistent but less precise alternatives [43]. For example, let us consider a dataset in which age is a quasi-identifier. While the three records {age = 30, gender = male}, {age = 35, gender = male}, and {age = 31, gender = female} are all distinct, releasing them as {age = 3 *, gender = male}, {age = 3 *, gender = male}, and {age = 3 *, gender = female} ensures they all belong to the same age category and the anonymity is 3-anonymity. Based on k -anonymity, l -diversity [44, 45] were proposed to address further disclosure issues of sensitive attributes.

5.1.3. Randomization Methods. Randomization can be used for attribute-level data. In this case, original sensitive values are replaced with similar but different values, with a certain probability. For example, a patient's name may be masked by a randomly selected made-up name. This basic approach may result in worse data quality. Li and Qin proposed to obtain value via a clustering method [20].

Randomization can further be used for aggregation operation. Obfuscation is a sort of such randomization. Numerous repetitions of a query by a single user must be detected and interrupted because they will converge on the true patient count making proper user identification absolutely necessary for the methods to function properly [6]. Aiming to deidentify aggregated data, obfuscation methods include the addition of a random number to the patient counts that has a distribution defined by a Gaussian function. Obfuscation is applied to aggregate patient counts that are reported as a result of ad hoc queries on the client machine [26]. Another protection model for preventing reidentification is differential privacy [10, 46]. In this model, reidentification is prevented by the addition of noise to the data. The model is based on the fact that auxiliary information will always make it easier to identify an individual in a dataset, even if anonymized. Instead, differential privacy seeks to guarantee that the information that is released when querying a dataset is nearly the same whether a specific person is included or not [46]. Unlike other methods, differential privacy provides formal statistical privacy guarantees.

5.1.4. Synthetization Methods. Synthetization is compelling for two main reasons: preserving confidentiality and valid inferences for various estimates [47]. In this case, the original data are never shared. Instead, general aggregate statistics about the data are computed, and new synthetic records are generated from the statistics to create fake, but realistic-like, data. Exploiting clinical data for building an intelligent system is one of the scenarios. Developing clinical natural language processing systems often requires access to many clinical documents, which are not widely available to the public due to privacy and security concerns. To address this challenge, Li et al. proposed to develop methods to generate synthetic clinical notes and evaluate their utility in real clinical natural language processing tasks. Thanks to the development of deep learning, recent advances in text generation have made it possible to generate synthetic clinical

notes that could be useful for training NER models for information extraction from natural clinical notes, thus lowering the privacy concern and increasing data availability [48].

5.2. Privacy-Preserving Data Mining. Data mining is also synonymously called knowledge discovery from data (KDD), which highlights the goal of the mining process. To obtain useful knowledge from data, the mining process can be divided into four iterative steps: data preprocessing, data transformation, data mining, and pattern evaluation and presentation. Based on the stage division in the process of KDD, Xu et al. developed a user-role-based methodology and identified four different types of users in a typical data mining scenario: data provider, data collector, data miner, and decision maker. By differentiating the four different user roles, privacy-preserving data mining (PPDM) can be explored in a principled way, by which all users care about the security of sensitive information but each user role views the security issue from its own perspective [49]. In this research, PPDM is explored from the view of a data miner role, that is, from the data mining stage of KDD.

Privacy-preserving data mining is aimed at mining or extracting information, via a certain machine learning-based model, from privacy-preserving data in which the values of individual records have been perturbed or masked [50]. The key challenge is that the privacy-preserving data look very different from the original records and the distribution of data values is also very different from the original distribution. Researches for this issue have started very early. Agrawal and Srikant proposed a reconstruction procedure to estimate the distribution of original data values and then built a decision-tree classifier [50]. Recent studies on PPDM include privacy-preserving association rule mining, privacy-preserving classification, and privacy-preserving cluster.

Association rule mining is aimed at finding interesting associations and correlation relationships among large sets of data items. For PPDM, some of the rules may be considered to be sensitive. For hiding these rules, the original data need to be modified to generate a sanitized dataset from which sensitive rules cannot be mined, while those nonsensitive ones can still be discovered [51]. Classification is a task of data analysis that learns models to automatically classify data into defined categories. Privacy-preserving classification evolves decision tree, Bayesian model, support vector machine, and neural classification. The strategies of adapting the classification method to a privacy-preserving scenario can simply be described as two aspects. The first is learning the classification model based on data transformation, since the transformed data is difficult to be recovered [52, 53]. The second is learning the classification model based on secure multiparty computation (SMC) [54], where multiparties collaborate to develop a classification model from vertically partitioned or horizontally partitioned data, but no one wants to disclose its data to others [55, 56]. Cluster analysis is the process of grouping a set of records into multiple groups or clusters so that objects within a cluster have high similarity but are very dissimilar to objects in other clusters. This process runs in an unsupervised manner. Similar to classification, current researches on privacy-preserving

clustering can be roughly categorized into two types, based on data transformation [57, 58] and based on secure multiparty computation [59, 60].

5.3. Federated Privacy-Preserving Data Mining. For the distributed or isolated data, distributed data mining is the research topic. Distributed data mining can be further categorized into data mining over horizontally partitioned data and data mining over vertically partitioned data. Research on distributed data mining attracts much attention. To overcome the difficulty of data integration and promote efficient information exchange without sharing sensitive raw data, Que et al. developed a Distributed Privacy-Preserving Support Vector Machine (DPP-SVM). The DPP-SVM enables privacy-preserving collaborative learning, in which a trusted server integrates “privacy-insensitive” intermediary results [61]. In medical domain, much raw data can hardly leave the institution of origin. Instead of bringing data to a central repository for computation, Wu et al. proposed a new algorithm, Grid Binary Logistic Regression (GLORE), to fit a LR model in a distributed fashion using information from locally hosted databases containing different observations that share the same attributes [62].

It is worth to note that learning (classification or clustering) on secure multiparty computation is an important distributed learning strategy, by which privacy disclosure concern can be much reduced since data need not to be shared out. This research topic probably inspired federated machine learning [30, 32]. Today’s AI still faces two major challenges. One is that data exists in the form of isolated islands. The other is the strengthening of data privacy and security. The two challenge is much severer in the healthcare domain. Federated machine learning is aimed at building a learning model from decentralized data [30]. Federated learning can be classified into horizontally federated learning, vertically federated learning, and federated transfer learning based on how data is distributed among various parties in the feature and sample ID space [32]. Horizontal federated learning, or sample-based federated learning, is introduced in the scenarios that datasets share the same feature space but different in samples. At the end of the learning, the universal model and the entire model parameters are exposed to all participants. Vertical federated learning or feature-based federated learning is applicable to the cases that two datasets share the same sample ID space but differ in feature space. At the end of learning, each party only holds the model parameters associated with its own features; therefore, at inference time, the two parties also need to collaborate to generate output. Federated transfer learning (FTL) applies to the scenarios that the two datasets differ not only in samples but also in feature space. FTL is an important extension to the existing federated learning systems and is more similar to vertical federated learning. The challenge of protecting data privacy while maintaining the data utility through machine learning still remains. For a comprehensive introduction of federated privacy-preserving data mining, please refer to the survey based on the proposed 5 W-scenario-based taxonomy [31].

5.4. Summary: Privacy vs. Accuracy. Privacy protection is the indispensable prerequisite of secondary use of health data.

As discussed above, risk-mitigation methods are aimed at anonymizing private or sensitive information so as to reduce the risk of reidentification. Methods about privacy-preserving data mining target to process the privacy-scrubbed data and extract knowledge and even build AI systems. If absolute privacy safe is pursued, the scrubbed data is definitely useless, since the data quality is severely corrupted. With the poor-quality data, accuracy and effectiveness of data utilization are extremely affected. Therefore, in a practical scenario, a certain tradeoff or compromise between privacy and accuracy must always be made. The tradeoff can be tuned to provide more or less privacy resulting in less or more accuracy, respectively, according to the requirements of privacy level and utility level. Federated privacy-preserving data mining sheds light on the new direction to compromise, even to balance, the privacy and accuracy. No need to share data out, federated privacy-preserving data mining first processes the original health data within institutions, and the conduct federated mining or learning. This type of method is expected to reconcile privacy and accuracy with more elegant style and more acceptable way.

6. Conclusions

Clinical data, genomic data, and consumer health data are the majority of health big data. Protection and reuse always gain much focused research topics. In this review article, the type and scope of health data are firstly discussed, followed by the related regulations for privacy protection. Then, strategies for user-controlled access and secure network architecture are presented. Sharing trained model without original data leaving out is a new important strategy and gains more and more focus. According to different data reuse scenarios, tasks and methods at three different levels are summarized. The strategies and methods can be combined to form a holistic solution.

With the rapid develop health information technology and artificial intelligence, the capability of privacy protection will impede the urgent demand of reusing health data. Some potential research directions may include (1) applying modern machine learning to deidentification and anonymization for multimodal health data while ensuring its data quality; (2) learning model construction and knowledge extraction based on anonymized data to leverage secondary use of health data; (3) federated learning on isolated health data can both protect privacy perfectly and improve the efficiency of data transferring and processing, being deserved more attention; (4) research on alleviating reidentification risk, such as linkage or inference, from a trained model.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was funded by the China Postdoctoral Science Foundation Grant (2020M671059) and the Fundamental Research Funds for the Central Universities (2572020BN02).

References

- [1] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care," *Nature Reviews Genetics*, vol. 13, no. 6, pp. 395–405, 2012.
- [2] M. Jiang, Y. Chen, M. Liu et al., "A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 601–606, 2011.
- [3] C.-A. Azencott, "Machine learning and genomics: precision medicine versus patient privacy," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2128, article 20170350, 2018.
- [4] DHHS, "Standards for privacy of individually identifiable health information. Office of the Assistant Secretary for Planning and Evaluation, DHHS. Final rule," *Federal Register*, vol. 65, no. 250, pp. 82462–82829, 2000.
- [5] Y. Joly, E. S. Dove, K. L. Kennedy et al., "Open science and community norms," *Medical Law International*, vol. 12, no. 2, pp. 92–120, 2012.
- [6] S. N. Murphy, V. Gainer, M. Mendis, S. Churchill, and I. Kohane, "Strategies for maintaining patient privacy in i2b2," *Journal of the American Medical Informatics Association*, vol. 18, Supplement 1, pp. i103–i108, 2011.
- [7] D. McGraw and K. D. Mandl, "Privacy protections to encourage use of health-relevant digital data in a learning health system," *npj Digital Medicine*, vol. 4, no. 1, p. 2, 2021.
- [8] W. N. Price and I. G. Cohen, "Privacy in the age of medical big data," *Nature Medicine*, vol. 25, no. 1, pp. 37–43, 2019.
- [9] R. Mendes and J. P. Vilela, "Privacy-preserving data mining: methods, metrics, and applications," *IEEE Access*, vol. 5, pp. 10562–10582, 2017.
- [10] A. Wood, M. Altman, A. Bembenek et al., "Differential privacy: a primer for a non-technical audience," *Vanderbilt Journal of Entertainment & Technology Law*, vol. 21, no. 1, p. 209, 2018.
- [11] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.
- [12] W. N. Price and I. G. Cohen, "Privacy in the age of medical big data," *Nature Medicine*, vol. 25, no. 1, pp. 37–43, 2019.
- [13] Y. Joly, S. O. M. Dyke, B. M. Knoppers, and T. Pastinen, "Are data sharing and privacy protection mutually exclusive?," *Cell*, vol. 167, no. 5, pp. 1150–1154, 2016.
- [14] D. Milius, E. S. Dove, D. Chalmers et al., "The International Cancer Genome Consortium's evolving data-protection policies," *Nature Biotechnology*, vol. 32, no. 6, pp. 519–523, 2014.
- [15] J. L. Raisaro, G. Choi, S. Pradervand et al., "Protecting privacy and security of genomic data in i2b2 with homomorphic encryption and differential privacy," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 5, pp. 1–1426, 2018.
- [16] N. Archer, U. Fevrier-Thomas, C. Lokker, K. A. McKibbin, and S. E. Straus, "Personal health records: a scoping review," *Journal of the American Medical Informatics Association*, vol. 18, no. 4, pp. 515–522, 2011.
- [17] L. Alpay, J. Verhoef, B. Xie, D. te'eni, and J. H. M. Zwetsloot-Schonk, "Current challenge in consumer health informatics: bridging the gap between access to information and information understanding," *Biomedical Informatics Insights*, vol. 2, no. 1, pp. 1–10, 2009.
- [18] G. M. Weber, K. D. Mandl, and I. S. Kohane, "Finding the missing link for big biomedical data," *JAMA*, vol. 311, no. 24, pp. 2479–2480, 2014.
- [19] Federal Trade Commission, "Fair Information Practice Principles," 2021, <https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf>.
- [20] X.-B. Li and J. Qin, "Anonymizing and sharing medical text records," *Information Systems Research*, vol. 28, no. 2, pp. 332–352, 2017.
- [21] "Health Information Technology for Economic and Clinical Health Act (HITECH Act)," 2021, <https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html>.
- [22] "Office of the Australian Information Commissioner (OIAC)," 2021, <https://www.cdr.gov.au/>.
- [23] "Personal Information Protection and Electronic Documents Act (PIPEDA)," 2021, <https://laws-lois.justice.gc.ca/eng/acts/p-8.6/>.
- [24] "Personal Information Protection Law of the PRC," 2021, <http://www.npc.gov.cn/npc/c30834/202108/a8c4e3672c74491a80b53a172bb753fe.shtml>.
- [25] S. N. Murphy and H. C. Chueh, "A security architecture for query tools used to access large biomedical databases," in *Proceedings of the AMIA Symposium 2002*, pp. 552–556, San Antonio, TX, USA, 2002.
- [26] S. N. Murphy, V. Gainer, and H. C. Chueh, "A visual interface designed for novice users to find research patient cohorts in a large biomedical database," in *AMIA Annual Symposium Proceedings 2003*, pp. 489–493, Bethesda, MD, USA, 2003.
- [27] S. N. Murphy, G. Weber, M. Mendis et al., "Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2)," *Journal of the American Medical Informatics Association*, vol. 17, no. 2, pp. 124–130, 2010.
- [28] M. Kantarcioglu, W. Jiang, Y. Liu, and B. Malin, "A cryptographic approach to securely share and query genomic sequences," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 5, pp. 606–617, 2008.
- [29] F. Tramèr, Z. Huang, J.-P. Hubaux, and E. Ayday, "Differential privacy with bounded priors: reconciling utility and privacy in genome-wide association studies," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 1286–1297, Denver Colorado USA, October 2015.
- [30] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54, pp. 1273–1282, 2017.
- [31] X. Yin, Y. Zhu, and J. Hu, "A comprehensive survey of privacy-preserving federated learning," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–36, 2021.
- [32] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.
- [33] G. Duncan and D. Lambert, "The risk of disclosure for microdata," *Journal of Business & Economic Statistics*, vol. 7, no. 2, p. 207, 1989.
- [34] B. Malin, K. Benitez, and D. Masys, "Never too old for anonymity: a statistical standard for demographic data sharing

- via the HIPAA Privacy Rule,” *Journal of the American Medical Informatics Association*, vol. 18, no. 1, pp. 3–10, 2011.
- [35] B. He, Y. Guan, J. Cheng, K. Cen, and W. Hua, “CRFs based de-identification of medical records,” *Journal of Biomedical Informatics*, vol. 58, pp. S39–S46, 2015.
- [36] S. M. Meystre, F. J. Friedlin, B. R. South, S. Shen, and M. H. Samore, “Automatic de-identification of textual documents in the electronic health record: a review of recent research,” *BMC Medical Research Methodology*, vol. 10, no. 1, p. 70, 2010.
- [37] O. Uzuner, Y. Luo, and P. Szolovits, “Evaluating the state-of-the-art in automatic de-identification,” *Journal of the American Medical Informatics Association*, vol. 14, no. 5, pp. 550–563, 2007.
- [38] L. Deleger, K. Molnar, G. Savova et al., “Large-scale evaluation of automated clinical note de-identification and its impact on information extraction,” *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 84–94, 2013.
- [39] L. Kamm, D. Bogdanov, S. Laur, and J. Vilo, “A new way to protect privacy in large-scale genome-wide association studies,” *Bioinformatics*, vol. 29, no. 7, pp. 886–893, 2013.
- [40] W. Lu, Y. Yamada, and J. Sakuma, “Efficient secure outsourcing of genome-wide association studies,” in *2015 IEEE Security and Privacy Workshops*, pp. 3–6, San Jose, CA, USA, May 2015.
- [41] S. Wang, X. Jiang, D. Fox, and L. Ohno-Machado, “Preserving genome privacy in research studies,” in *Medical Data Privacy Handbook*, pp. 425–441, Springer International Publishing, Cham, 2015.
- [42] P. Samarati, “Protecting respondents identities in microdata release,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 6, pp. 1010–1027, 2001.
- [43] P. Samarati and L. Sweeney, “Generalizing data to provide anonymity when disclosing information,” in *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems - PODS '98*, p. 188, Seattle Washington USA, 1998.
- [44] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkatasubramanian, “l-diversity: privacy beyond k-anonymity,” *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, p. 3, 2007.
- [45] N. Li, T. Li, and S. Venkatasubramanian, “t-closeness: privacy beyond k-anonymity and l-diversity,” in *2007 IEEE 23rd International Conference on Data Engineering*, pp. 106–115, Istanbul, Turkey, 2007.
- [46] C. Dwork, “Differential privacy,” in *Automata, Languages and Programming*, pp. 1–12, Springer, 2006.
- [47] J. P. Reiter, “Releasing multiply imputed, synthetic public use microdata: an illustration and empirical study,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 168, no. 1, pp. 185–205, 2005.
- [48] J. Li, Y. Zhou, X. Jiang et al., “Are synthetic clinical notes useful for real natural language processing tasks: a case study on clinical entity recognition,” *Journal of the American Medical Informatics Association*, vol. 28, no. 10, pp. 2193–2201, 2021.
- [49] Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, and Yong Ren, “Information security in big data: privacy and data mining,” *IEEE Access*, vol. 2, pp. 1149–1176, 2014.
- [50] R. Agrawal and R. Srikant, “Privacy-preserving data mining,” in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00*, pp. 439–450, New York, NY, United States, 2000.
- [51] S. K and G. S. Sadasivam, “A survey on privacy preserving association rule mining,” *International Journal of Data Mining & Knowledge Management Process*, vol. 3, no. 2, pp. 119–131, 2013.
- [52] P. K. Fong and J. H. Weber-Jahnke, “Privacy preserving decision tree learning using unrealized data sets,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 2, pp. 353–364, 2012.
- [53] K.-P. Lin and M.-S. Chen, “On the design and analysis of the privacy-preserving SVM classifier,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1704–1717, 2011.
- [54] C. Zhao, S. Zhao, M. Zhao et al., “Secure Multi-Party Computation: Theory, practice and applications,” *Information Sciences*, vol. 476, pp. 357–372, 2019.
- [55] J. Vaidya, H. Yu, and X. Jiang, “Privacy-preserving SVM classification,” *Knowledge and Information Systems*, vol. 14, no. 2, pp. 161–178, 2008.
- [56] J. Vaidya, M. Kantarcioğlu, and C. Clifton, “Privacy-preserving Naïve Bayes classification,” *VLDB Journal*, vol. 17, no. 4, pp. 879–898, 2008.
- [57] R. R. Rajalaxmi and A. M. Natarajan, “An effective data transformation approach for privacy preserving clustering,” *Journal of Computer Science*, vol. 4, no. 4, pp. 320–326, 2008.
- [58] M. NagaLakshmi and K. Sandhya Rani, “SVD based data transformation methods for privacy preserving clustering,” *International Journal of Computers and Applications*, vol. 78, no. 3, pp. 39–43, 2013.
- [59] J. Vaidya and C. Clifton, “Privacy-preserving k-means clustering over vertically partitioned data,” in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03*, p. 206, Washington, D.C., 2003.
- [60] R. Akhter, R. J. Chowdhury, K. Emura, T. Islam, M. S. Rahman, and N. Rubaiyat, “Privacy-preserving two-party k-means clustering in malicious model,” in *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*, pp. 121–126, Japan, July 2013.
- [61] J. Que, X. Jiang, and L. Ohno-Machado, “A collaborative framework for distributed privacy-preserving support vector machine learning,” *AMIA Annual Symposium Proceedings 2012*, vol. 2012, pp. 1350–1359, 2012.
- [62] Y. Wu, X. Jiang, J. Kim, and L. Ohno-Machado, “Grid Binary LOGistic REGression (GLORE): building shared models without sharing data,” *Journal of the American Medical Informatics Association*, vol. 19, no. 5, pp. 758–764, 2012.

Research Article

Identification of Specific Cell Subpopulations and Marker Genes in Ovarian Cancer Using Single-Cell RNA Sequencing

Yan Li ¹, Juan Wang,² Fang Wang,² Chengzhen Gao,¹ Yuanyuan Cao,¹
and Jianhua Wang ³

¹Department of Obstetrics and Gynecology, The Yancheng Clinical College of Xuzhou Medical University, The First People's Hospital of Yancheng, Yancheng, 224001 Jiangsu, China

²Department of Obstetrics and Gynecology, The First Affiliated Hospital of Soochow University, Suzhou, 215006 Jiangsu, China

³Department of Gastroenterology, The Yancheng Clinical College of Xuzhou Medical University, The First People's Hospital of Yancheng, Yancheng, 224001 Jiangsu, China

Correspondence should be addressed to Jianhua Wang; jiahuw@163.com

Received 11 May 2021; Accepted 24 August 2021; Published 7 October 2021

Academic Editor: Lei Zhang

Copyright © 2021 Yan Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. Ovarian cancer is the deadliest gynaecological cancer globally. In our study, we aimed to analyze specific cell subpopulations and marker genes among ovarian cancer cells by single-cell RNA sequencing (RNA-seq). **Methods.** Single-cell RNA-seq data of 66 high-grade serous ovarian cancer cells were employed from the Gene Expression Omnibus (GEO). Using the Seurat package, we performed quality control to remove cells with low quality. After normalization, we detected highly variable genes across the single cells. Then, principal component analysis (PCA) and cell clustering were performed. The marker genes in different cell clusters were detected. A total of 568 ovarian cancer samples and 8 normal ovarian samples were obtained from The Cancer Genome Atlas (TCGA) database. Differentially expressed genes were identified according to $|\log_2$ fold change (FC) > 1 and adjusted p value < 0.05 . To explore potential biological processes and pathways, functional enrichment analyses were performed. Furthermore, survival analyses of differentially expressed marker genes were performed. **Results.** After normalization, 6000 highly variable genes were identified across the single cells. The cells were divided into 3 cell populations, including G1, G2M, and S cell cycles. A total of 1,124 differentially expressed genes were identified in ovarian cancer samples. These differentially expressed genes were enriched in several pathways associated with cancer, such as metabolic pathways, pathways in cancer, and PI3K-Akt signaling pathway. Furthermore, marker genes, STAT1, ANP32E, GPRC5A, and EGFL6, were highly expressed in ovarian cancer, while PMP22, FBXO21, and CYB5R3 were lowly expressed in ovarian cancer. These marker genes were positively associated with prognosis of ovarian cancer. **Conclusion.** Our findings revealed specific cell subpopulations and marker genes in ovarian cancer using single-cell RNA-seq, which provided a novel insight into the heterogeneity of ovarian cancer.

1. Introduction

Ovarian cancer is one of the most common gynaecological cancers in the world, with high heterogeneity and poor prognosis [1]. High-grade serous ovarian cancer is the deadliest subtype of ovarian cancer, with up to 80% of patients recurring after initial treatment [2]. Despite advances in treatments such as surgery and chemotherapy, the 5-year survival rate of patients with advanced ovarian cancer remains around 30%

40% [3, 4]. Since ovarian cancer patients are usually diagnosed at an advanced stage, genetic risk prediction and prevention strategies will be an important way to reduce ovarian cancer mortality [5]. Targeted therapies significantly improve the therapeutic effects of patients with ovarian cancer [6]. However, ovarian cancer shows heterogeneity within the tumor that may affect the therapeutic outcomes of targeted therapies. Tumors including ovarian cancer usually consist of heterogeneous cells that are different in many biological features, like

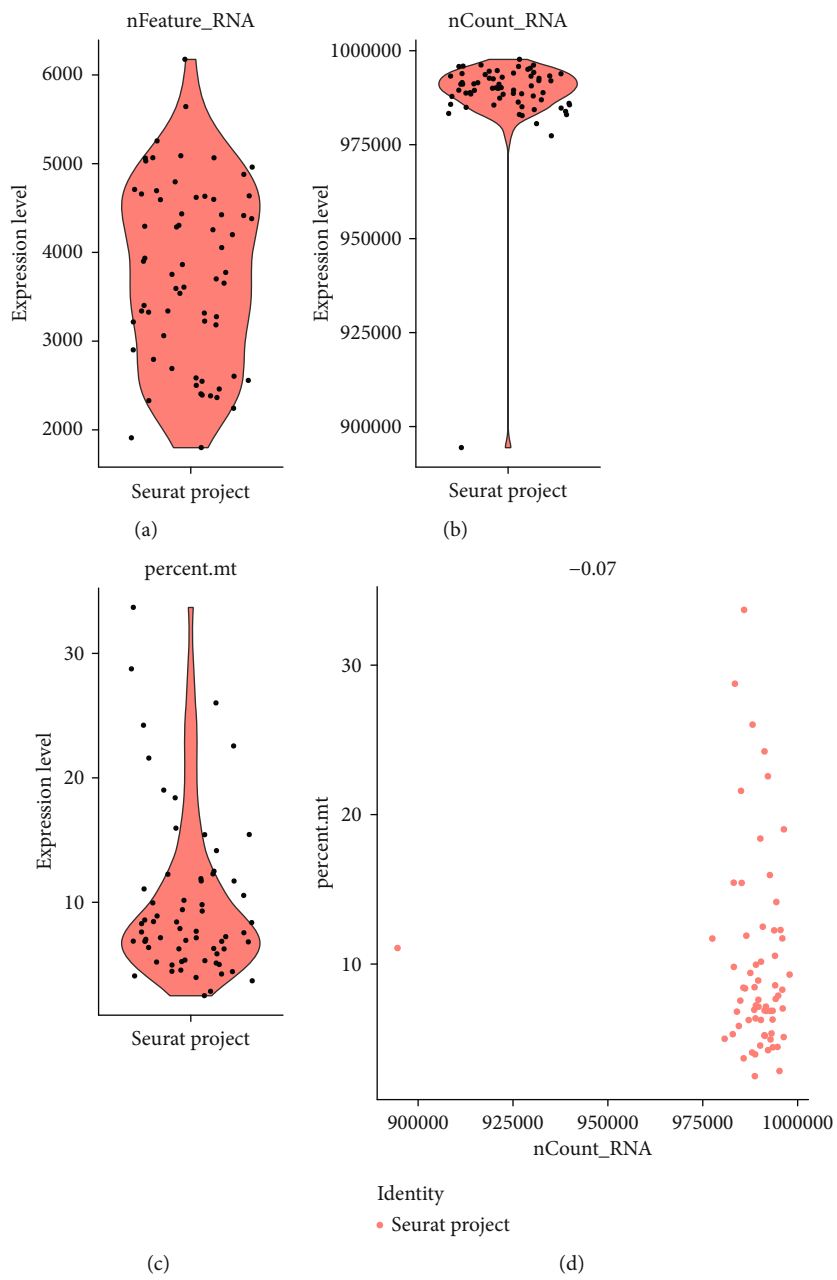


FIGURE 1: Continued.

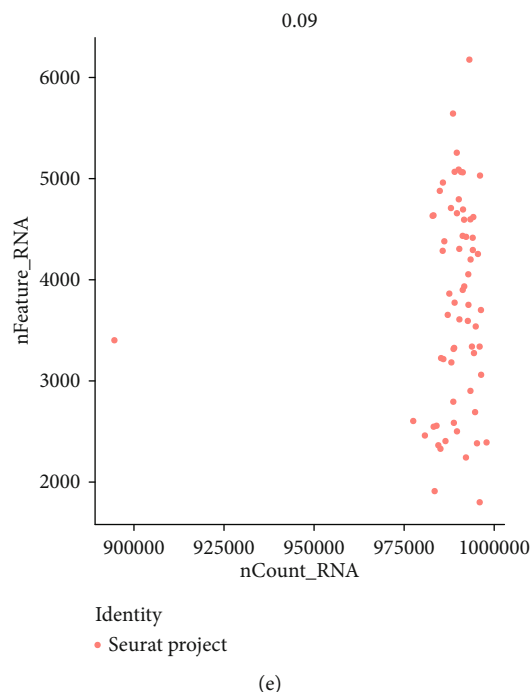


FIGURE 1: Quality control filtering to remove cells with low quality. (a) Violin plots showing the counts of genes in each cell. (b) Violin plots of the sum of the expression levels of all genes in each cell. (c) Violin plots of the percentage of mitochondrial genes. (d) Scatter plots for the percentage of mitochondrial genes in the sum of the expression levels of all genes in each cell. (e) Scatter plots for the counts of genes in the sum of the expression levels of all genes in each cell.

morphology, apoptosis, and invasion [7]. However, RNA-seq data reflect the average expression levels of different cells, not to reveal the intrinsic expression differences between different cell subpopulations. The genetic heterogeneity of ovarian cancer has been confirmed at single-cell resolution. The heterogeneity of gene expression levels greatly affects the patients' clinical outcomes [8]. Therefore, understanding the heterogeneity of tumors at the transcriptome level and the precise characterization of gene expression in tumors may help to identify better therapeutic molecular targets [9]. The characterization of heterogeneous tumor features will help to develop more effective molecular targeted therapeutics.

The basic unit of cancer is the innovative single cell along with genetics and epigenetics. Single-cell control determines the parameters of various aspects of cancer biology. Thus, single-cell analysis provides the ultimate resolution for us to understand the biology of various diseases [10]. Single-cell RNA-seq has become a promising approach for revealing the clonal genotype and population structure of human cancers. RNA-seq of the single cell can be used to analyze the cell type in the tumor microenvironment, the tumor heterogeneity, and its clinical significance [11]. Unlike traditional sequencing methods, single-cell sequencing methods provide different types of omics analysis for individual cells, such as genomics and transcriptomics [12]. Among them, single-cell RNA sequencing (scRNA-seq) is capable of measuring gene expression at the single-cell level. Based on classical markers, the scRNA-seq reveals the heterogeneity of gene expression in individual cells or cells with the same type [13], rather than

simply examining differential expression between two cells. In this study, we analyzed the heterogeneity among ovarian cancer cells and identified marker genes by scRNA-seq.

2. Materials and Methods

2.1. Ovarian Cancer Single-Cell RNA-seq Datasets. Single-cell RNA-seq gene expression data of ovarian cancer were employed from the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) database with accession number GSE123476. According to the study of Winterhoff et al., 19 cells were excluded due to poor cell morphology, extremely large or small cell size, or evidence of multiple cells in the well. Meanwhile, 7 cells that did not express at least 1,000 of the highly expressed genes were also removed [14]. As a result, 26 ovarian cancer cells with low quality were removed from 92 cells. The barcode information and single-cell RNA-seq gene expression matrix were extracted for further analyses [14].

2.2. Quality Control Filtering and Data Normalization. The gene expression matrix was imported into the Seurat package in R (version 3.1.0; <http://satijalab.org/seurat/>). Seurat, as a tool for single-cell genomics, is used for quality control, analysis, and exploration of single-cell RNA-seq data [15]. For single-cell RNA-seq data, there could be cells with low quality, probably due to the loss of cytoplasmic RNA when the cells were disrupted. Since mitochondria were much larger than single transcriptional molecules, they were not easily leaked out of the broken cell membrane, causing the

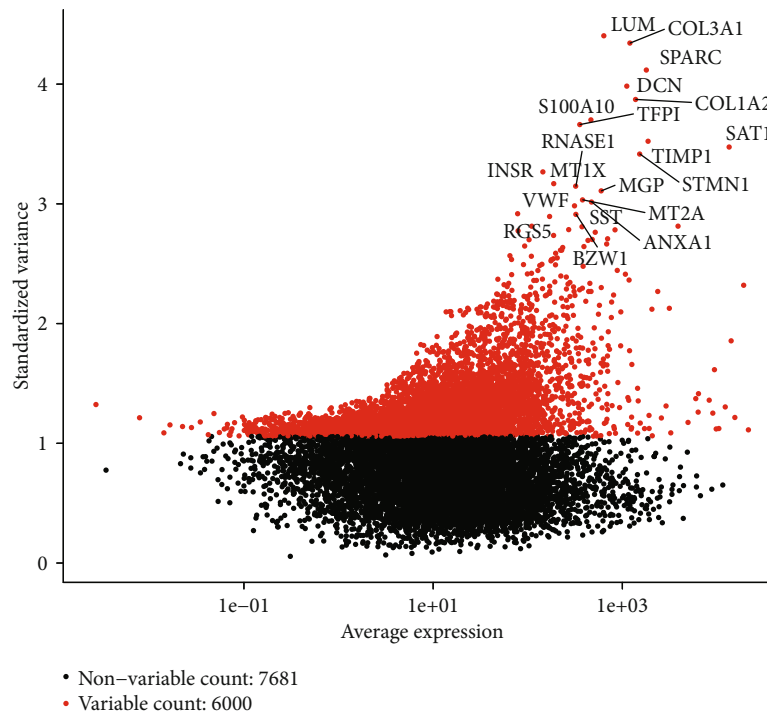


FIGURE 2: Detection of highly variable genes across the cells. x axis represents the average expression, and y axis represents standardized variance.

abnormally high proportion of mitochondrial genes among the cells in sequencing results. Thus, to remove cells with low quality, quality control was performed. After quality control, fragments per kilobase of transcript per million mapped read (FPKM) values were transformed into the log-space.

2.3. Detection of Highly Variable Genes across the Single Cells.

To eliminate the dimensional relationship between variable genes and make the data comparable, using the `NormalizeData` function of the Seurat package, data were normalized with the log-normalize method. For each gene, we calculated the standard variance in all cells using the `FindVariableFeature` function. Herein, mean-variance was calculated as 1. Standard variance cut-off of 1 was used to identify highly variable genes. The top 20 highly variable genes were identified.

2.4. Cell Clustering Analysis Using Seurat.

Principal component analysis (PCA) is a multivariate statistical method that examines the correlation between different variables. PCA was used to examine how to reveal the internal structure between multiple variables through a few principal components. That is, a few principal components were derived from the original variables while they retained the information of the original variables as much as possible and were not related to each other. In our study, PCA was carried out based on highly variable genes. Using the screened PCs as input, the cell clustering was visualized using Uniform Manifold Approximation and Projection (UMAP) via the `RunUMAP` function.

2.5. Gene Scoring. The `CellCycleScoring` function of the Seurat package was used to score the marker genes in the two cell cycles G2M and S based on the gene expression levels.

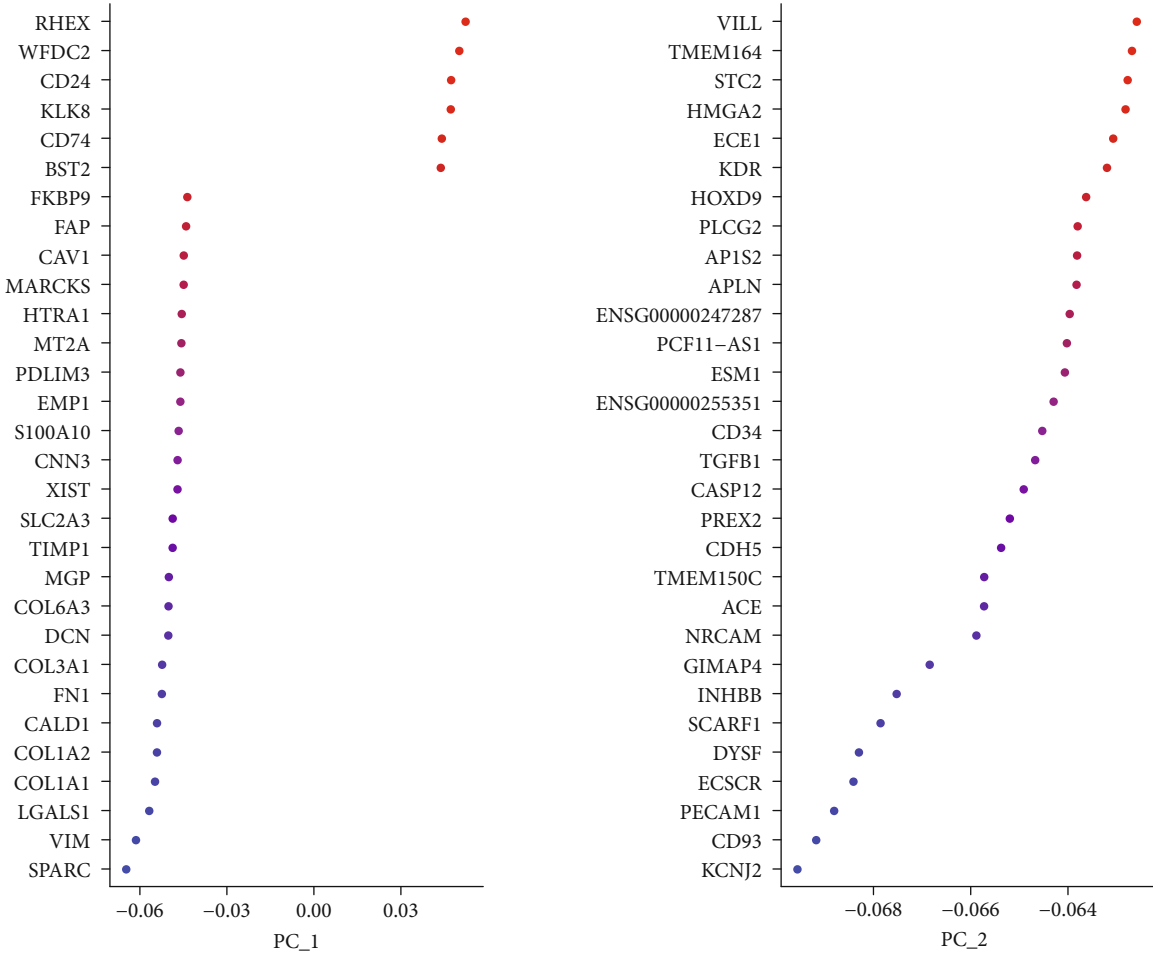
We calculated the average expression value of S phase genes and G2/M phase genes for each cell. All genes were divided into different bins based on the average expression levels, and then, the control genes were randomly selected as the background from each bin. The average expression levels of these control genes were calculated. The average expression levels of control genes were subtracted from the average expression levels of S phase genes and G2/M phase genes to obtain S.Score and G2M.Score. S.Score < 0 and G2M.Score < 0 were judged as G1 phase, otherwise, which phase was judged as which score was higher. The difference between the cell cluster and the cell cycle distribution was examined by Fisher's test. The top ten differentially expressed genes and the cell cycle were separately plotted, which were visualized into heatmaps.

2.6. Detection of Marker Genes and Functional Enrichment Analysis.

The cluster marker genes with $|\log_2 \text{fold change (FC)}| \geq 0.25$, the expression ratio of cell population ≤ 0.25 , and p value ≤ 0.05 were identified using the `FindAllMarkers` function in the Seurat package. An expression heatmap was generated for given cells and genes using the `DoHeatmap`. The expression level of markers in each cluster was calculated, and the putative identities of each cell clustering were identified. The top 20 markers were plotted for each cluster. To explore potential biological processes and pathways enriched by markers in each cluster, functional enrichment analyses were performed using the `gProfiler` package.

2.7. Reconstruction of Differentiation Trajectories Using Monocle Analysis.

The pseudotime estimation analysis of epithelial cancer cells and stromal cells was performed using



(a)

FIGURE 3: Continued.

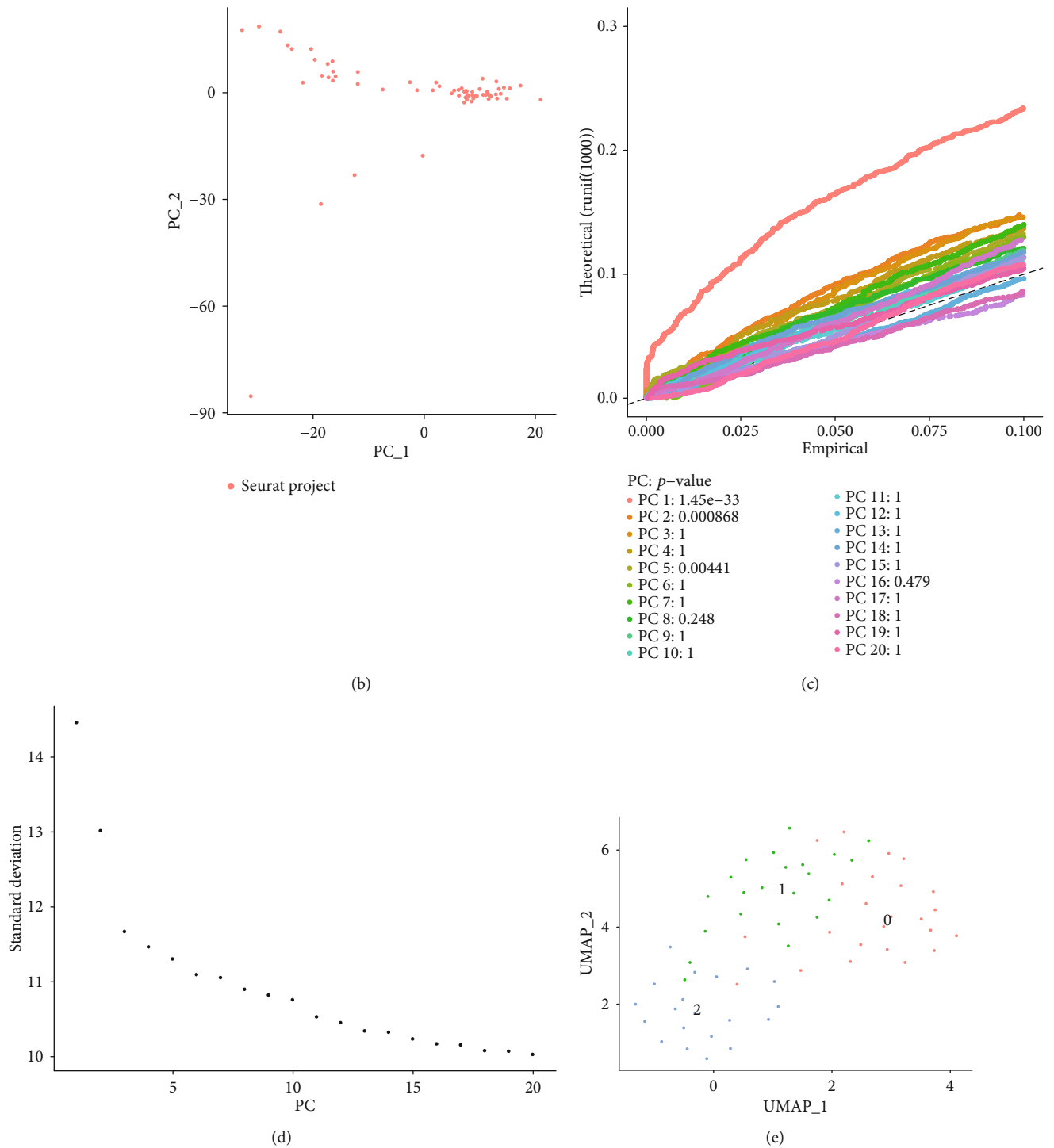
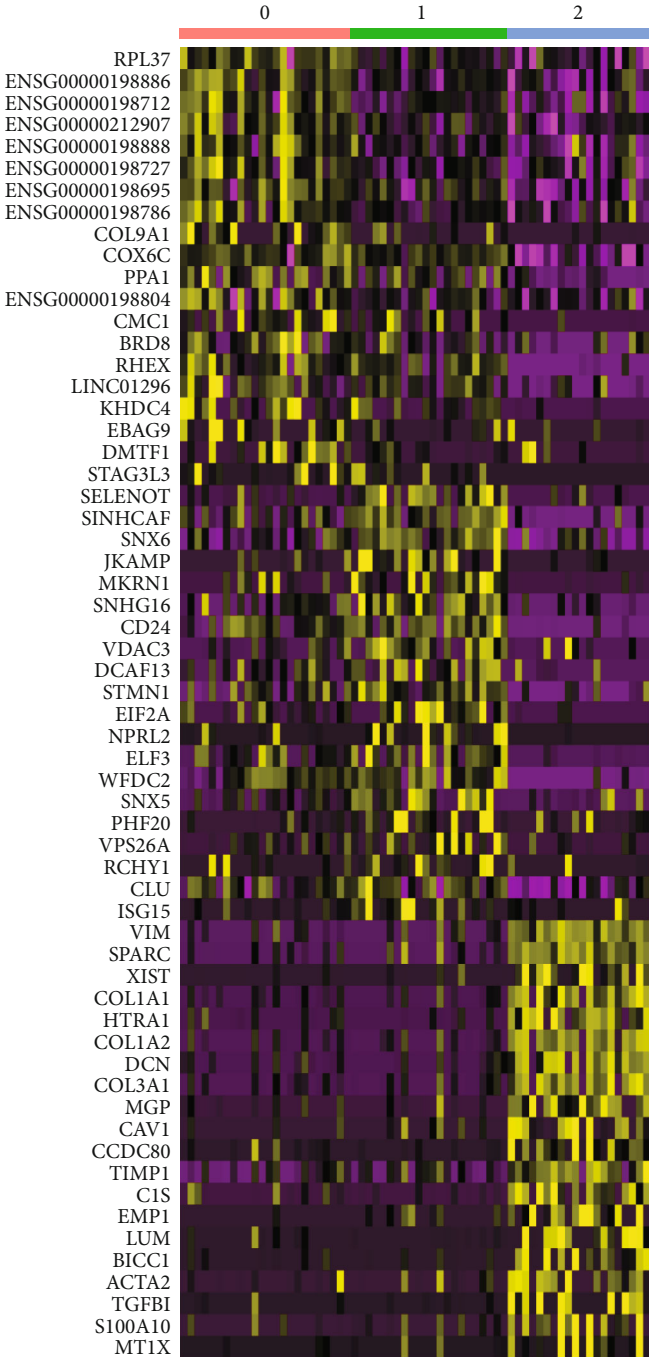


FIGURE 3: Individual principal component analysis. (a) The top 30 genes in PC1 and PC2. (b) The correlation between PC1 and PC2. (c) The p value of PCs. (d) The PCs were sorted based on the standard deviation. (e) UMAP plots showing the three cell clusters.

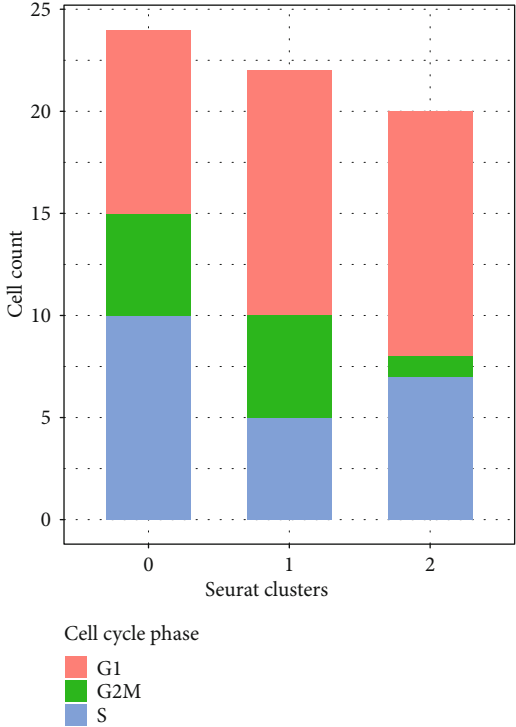
the Monocle package. A pseudotime plot was generated that can account for both branched and linear differentiation processes based on the top 2000 highly variable marker genes.

2.8. Differential Expression Analysis and Function Enrichment Analysis Using Ovarian Cancer Datasets. A total of 593 ovarian cancer samples were obtained from The Cancer Genome Atlas

(TCGA) using the UCSC Xena (<https://tcga.xenahubs.net>), including gene expression profiles and clinical information. Supplementary table 1 listed the IDs of all samples. After removing 17 relapse ovarian cancer, 568 ovarian cancer samples and 8 normal ovarian tissue samples were employed for this study. Differential expression analysis was then performed according to $|\log_2 \text{FC}| > 1$ and adjusted p value

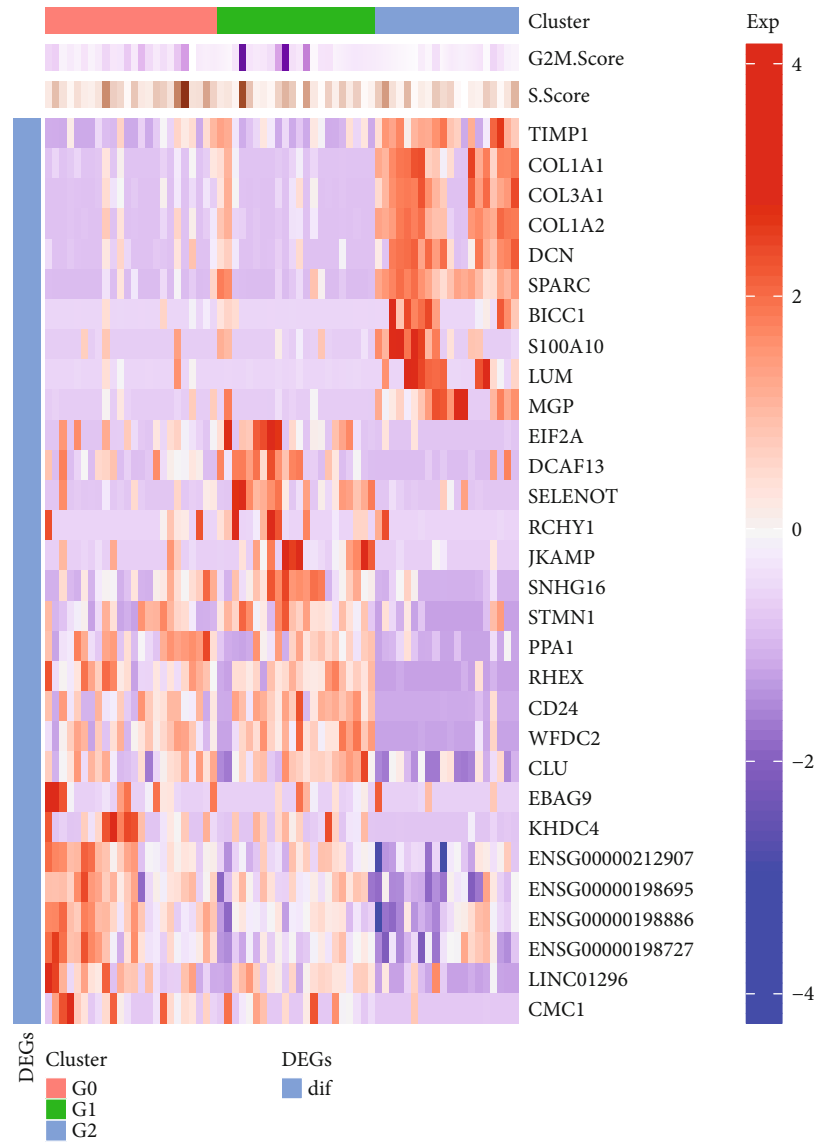


(a)



(b)

FIGURE 4: Continued.



(c)

FIGURE 4: Cell cycle analyses. (a) The top 20 marker genes in the three cell subpopulations. (b) Cell cycle phase. (c) Cell cycle heatmap.

<0.05 using the limma package in R. To explore potential biological processes and pathways, functional enrichment analyses of upregulated and downregulated genes were performed using the gProfiler package in R, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). The GO terms include biological process (BP), cellular component (CC), and molecular function (MF). Terms with p value <0.05 were significantly enriched.

2.9. Overall Survival Analysis. Marker genes and differentially expressed genes were overlapped. Overall survival and recurrence-free survival analyses of differentially expressed marker genes were performed. Kaplan-Meier survival curves and log-rank tests were performed to evaluate the associations between ovarian cancer prognosis and the expression of these prognostic genes.

3. Results

3.1. Identification of Three Cell Subpopulations across Ovarian Cells Based on Single-Cell RNA-seq. In total, 66 ovarian cancer cells were included in this study. Considering that the amount of data and the number of cells was relatively small, we used all the cells without filtering (Figures 1(a)–1(e)). Then, we detected 6000 highly variable genes across the single cells after calculating the mean and the variance to mean ratio of each gene. The top 20 highly variable genes such as LUM, COL3A1, and SPARC are shown in Figure 2.

To overcome the various technical noise in any single feature of scRNA-seq data, the Seurat package was used to cluster cells according to their PCA scores, where each PC represented a “meta-feature” (Figures 3(a) and 3(b)). Jack-Straw function was used to resample the test. We randomly replaced a subset of the data (default was 1%) and rerun

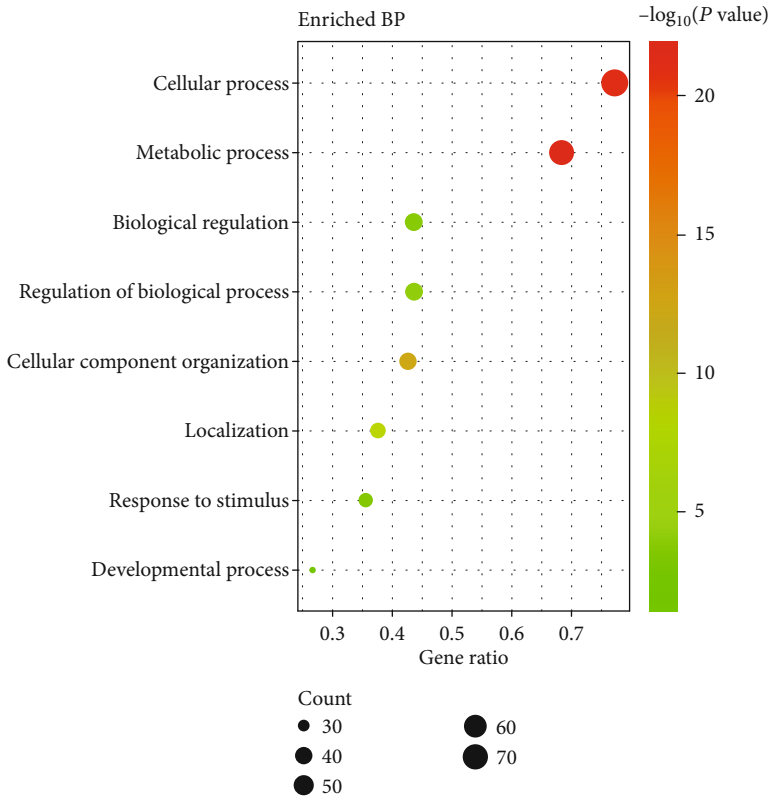
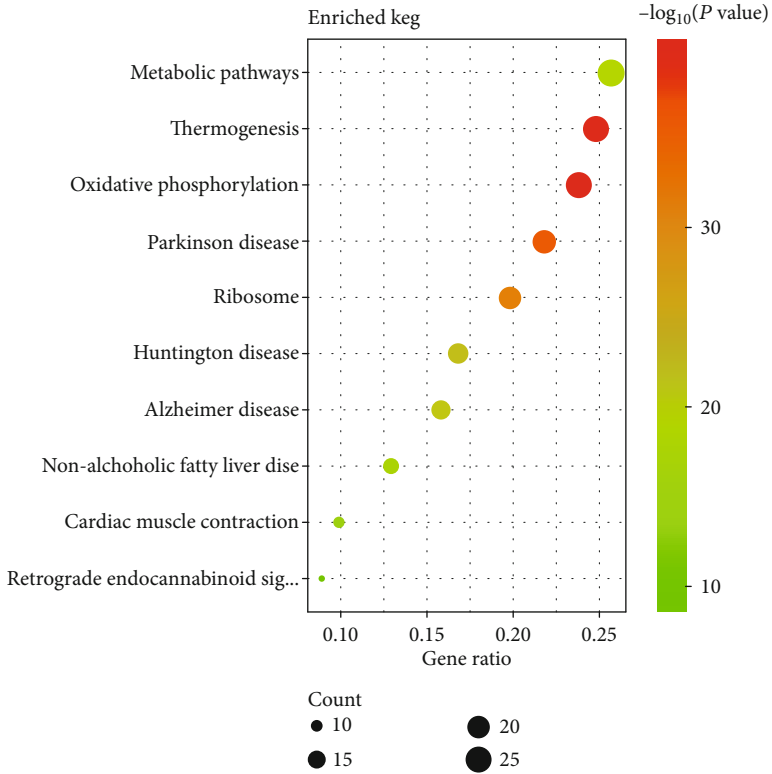
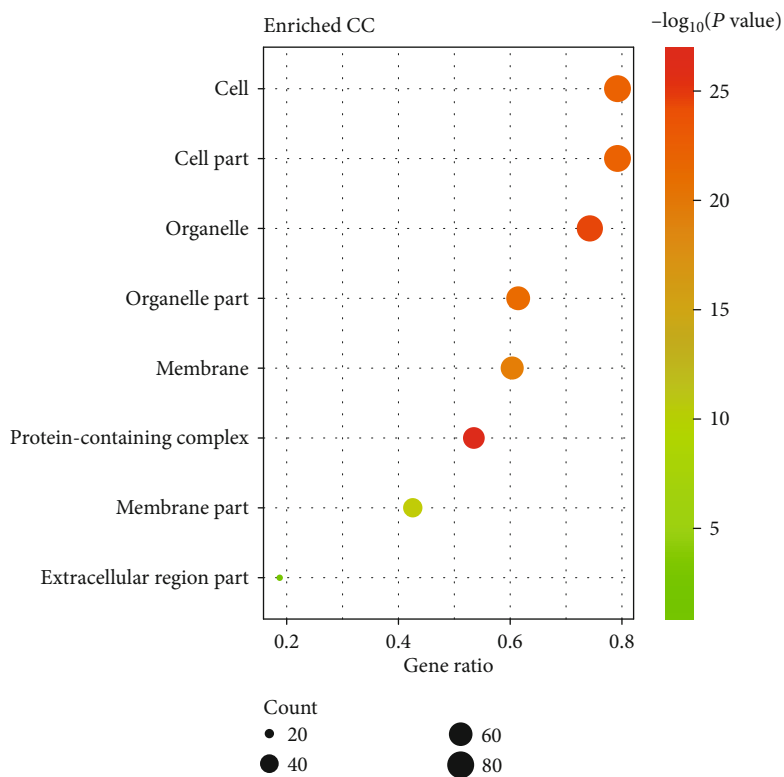
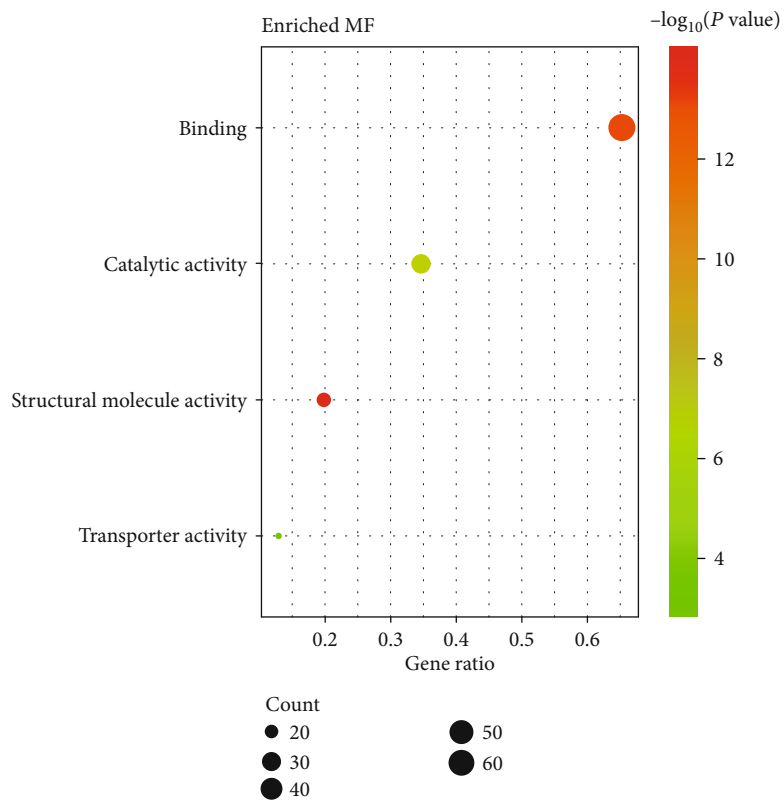


FIGURE 5: Continued.



(c)



(d)

FIGURE 5: Continued.

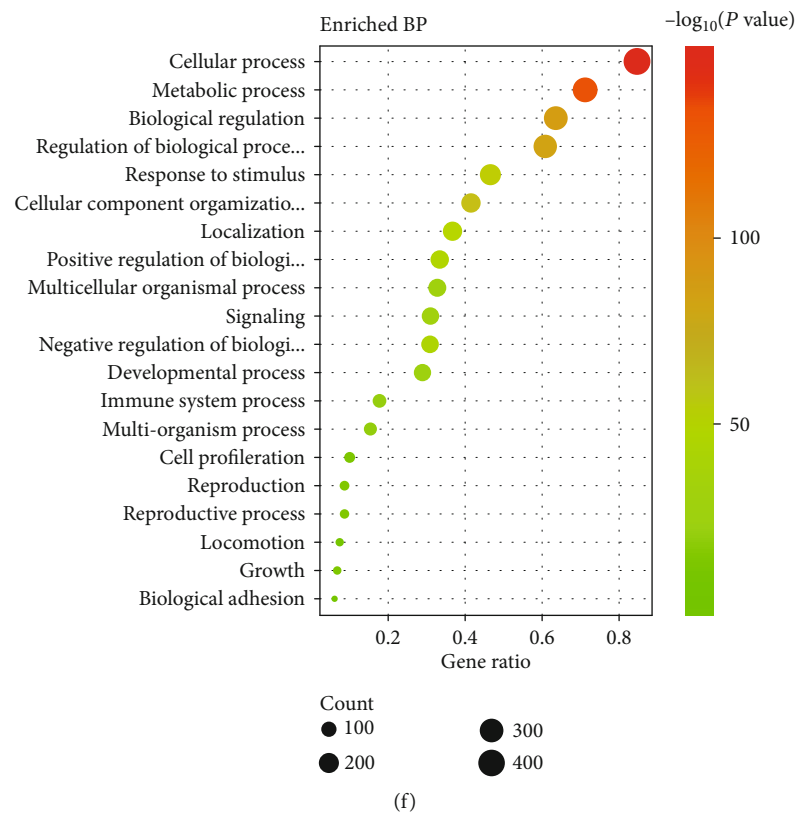
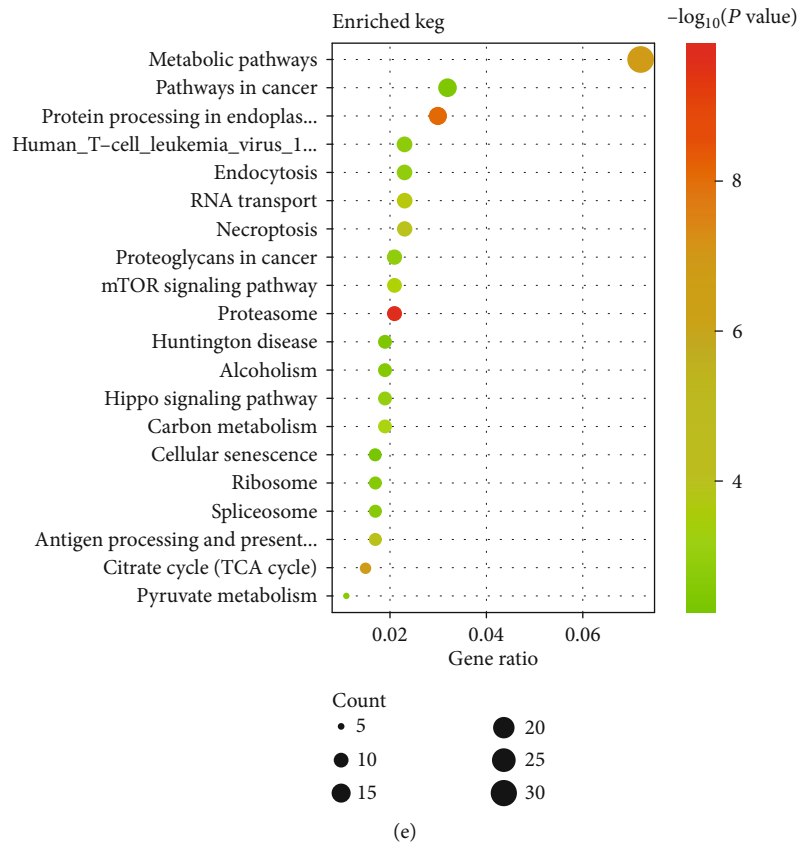
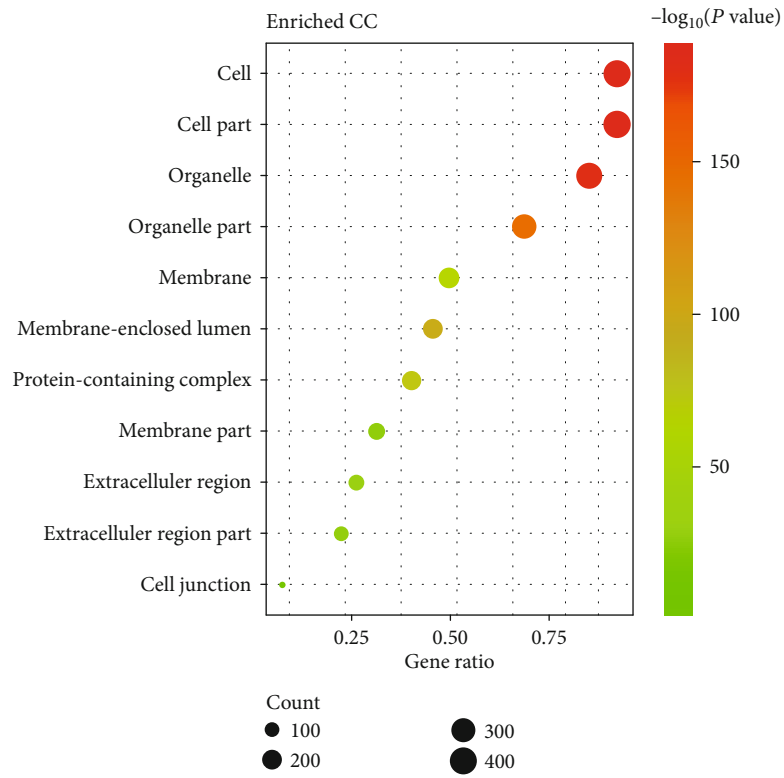
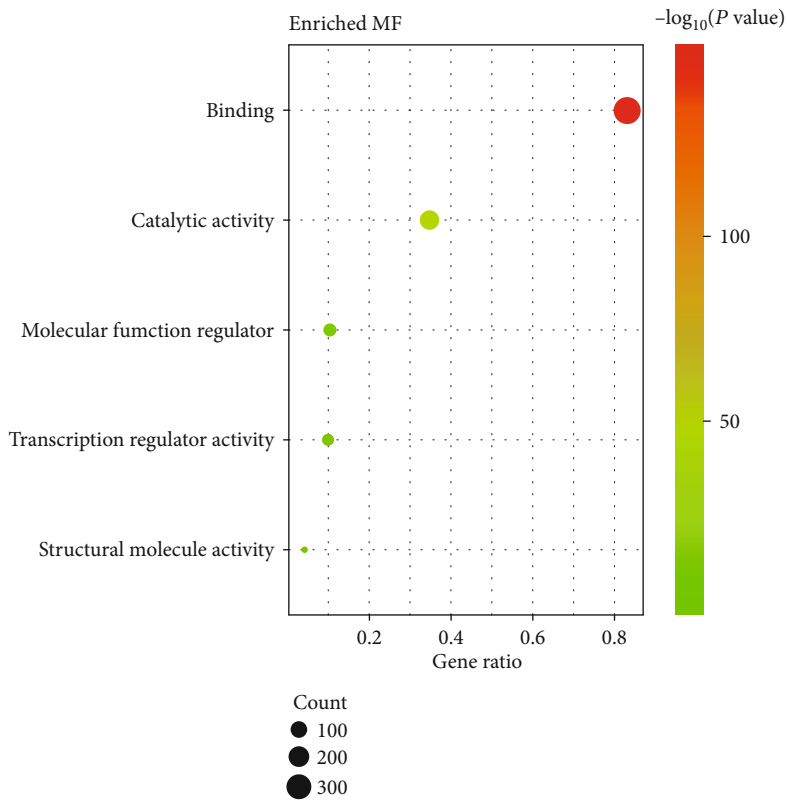


FIGURE 5: Continued.

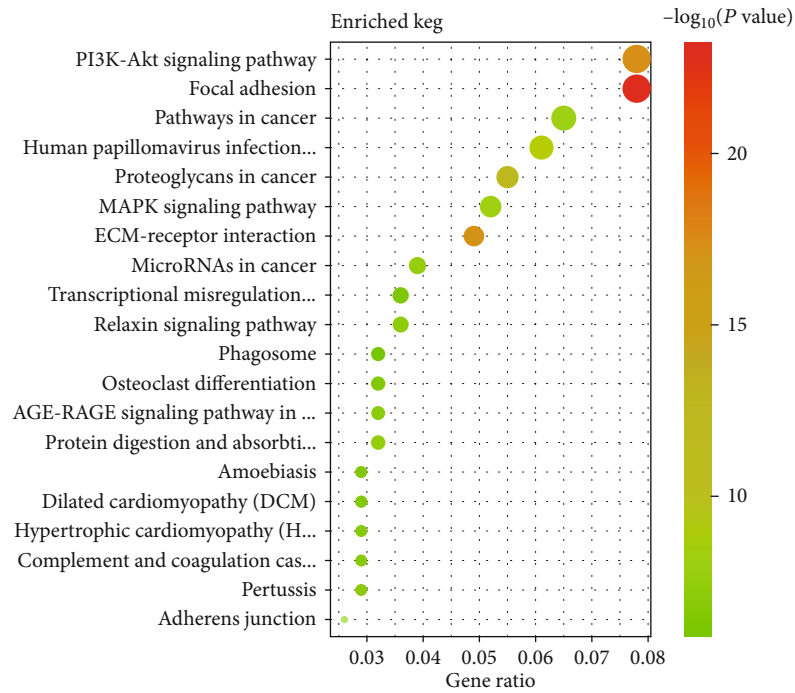


(g)

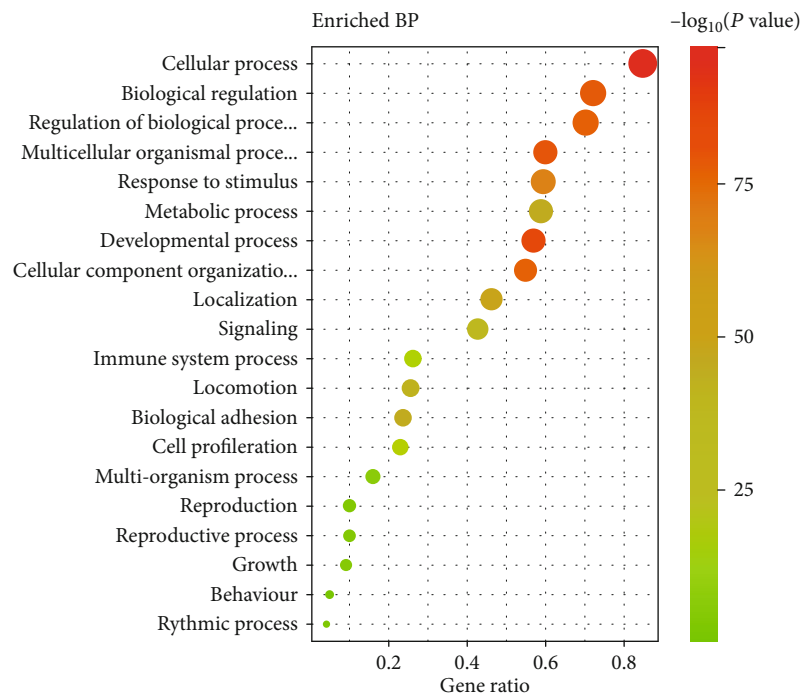


(h)

FIGURE 5: Continued.

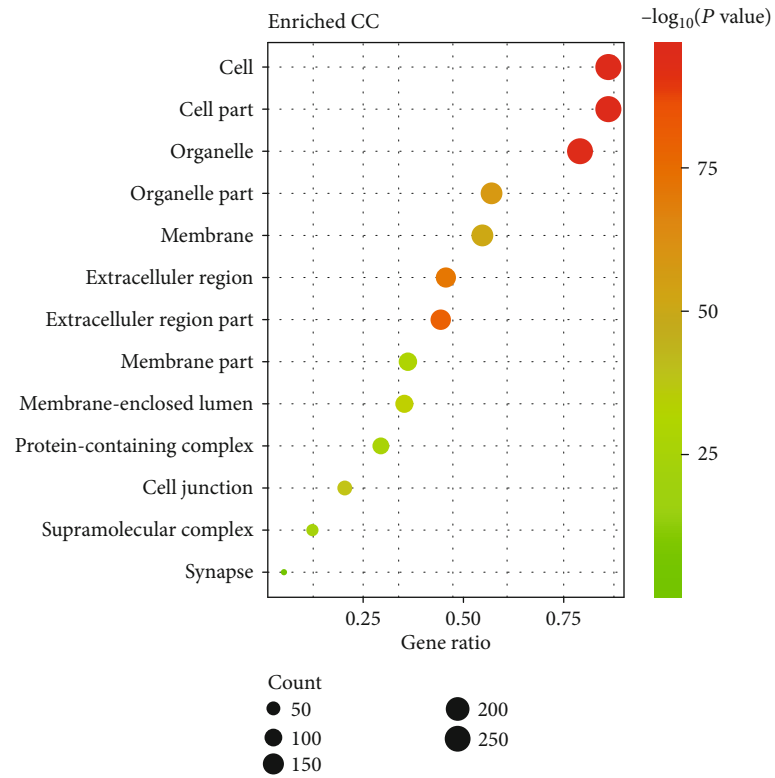


(i)

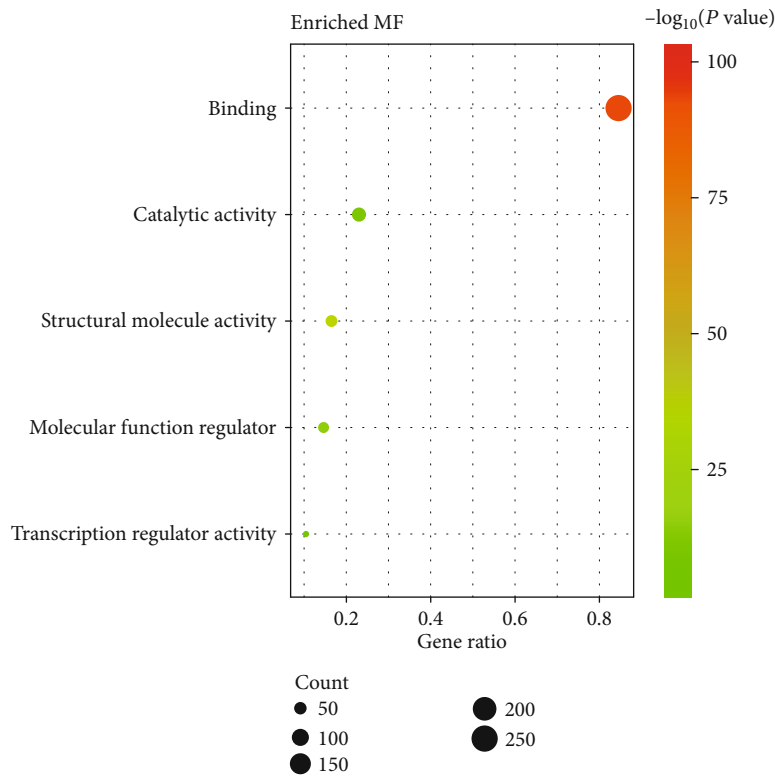


(j)

FIGURE 5: Continued.



(k)



(l)

FIGURE 5: Function enrichment analyses. (a–d) The enriched KEGG pathways and GO terms including BP, CC, and MF in cluster 0. (e–h) The enriched KEGG pathways and GO terms including BP, CC, and MF in cluster 1. (i–l) The enriched KEGG pathways and GO terms including BP, CC, and MF in cluster 2.

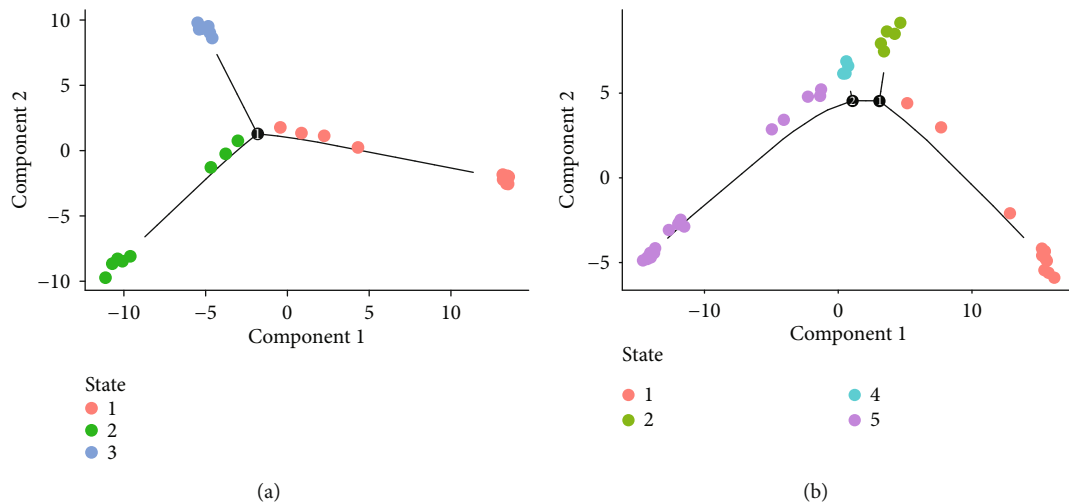


FIGURE 6: Reconstruction of differentiation trajectories to ovarian cancer. (a, b) The trajectory plot in pseudotime of epithelial cancer cells and stromal cells using Monocle analysis. Different colors represent different cell states.

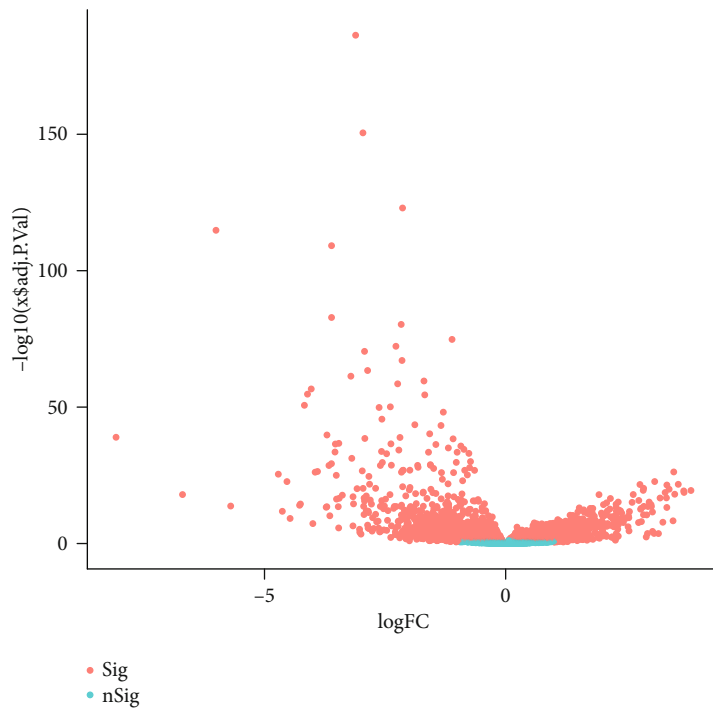
PCA to construct an “empty distribution” of feature scores and repeated the process (Figure 3(c)). We identified “important” PCs with low p values. Furthermore, the PCs were sorted based on the standard deviation using ElbowPlot function (Figure 3(d)). Because there was no obvious elbow point, we selected 19 PCs for downstream analysis. After cluster analysis, we divided the cells into 3 cell populations across ovarian cancer cells (Figure 3(e)). The number of cells in clusters 0, 1, and 2 was 24, 22, and 20. Supplementary table 2 listed which cells were in which cluster.

3.2. Analysis of Marker Genes in the Three Cell Subpopulations. The top 20 marker genes in the three cell subpopulations are listed in heatmap (Figure 4(a)). We used the Seurat tool to score the marker genes in the G1, G2M, and S cell cycles. Figure 4(b) shows the cell counts in the G1, G2M, and S cell cycles. By Fisher’s test, there was no significant difference between the three cell subpopulations and cells in each cell cycle (p value = 0.2834). Cell cycle heatmap shows the top ten differentially expressed genes and cell cycle scores in each cell subpopulation (Figure 4(c)). To explore potential biological processes and pathways, GO and KEGG enrichment analyses were performed (Figure 5). Genes in cluster 1 (Figures 5(a)–5(d)) and cluster 2 (Figures 5(e)–5(h)) were mainly enriched in metabolic processes and pathways. Meanwhile, genes in cluster 2 were primarily involved in cancer-related pathways such as PI3K-Akt pathway and pathways in cancer (Figures 5(i)–5(l)). We found that these marker genes were enriched in different biological processes and pathways in different cell subpopulations such as metabolic pathways, pathways in cancer, and mTOR signaling pathway.

3.3. Reconstruction of Differentiation Trajectories Using Monocle Package. Cell fate decisions and differentiation trajectories were reconstructed with the Monocle package. The pseudotime estimation analysis of epithelial cancer cells and stromal cells was performed based on the top 2000 highly variable marker genes (Figures 6(a) and 6(b)).

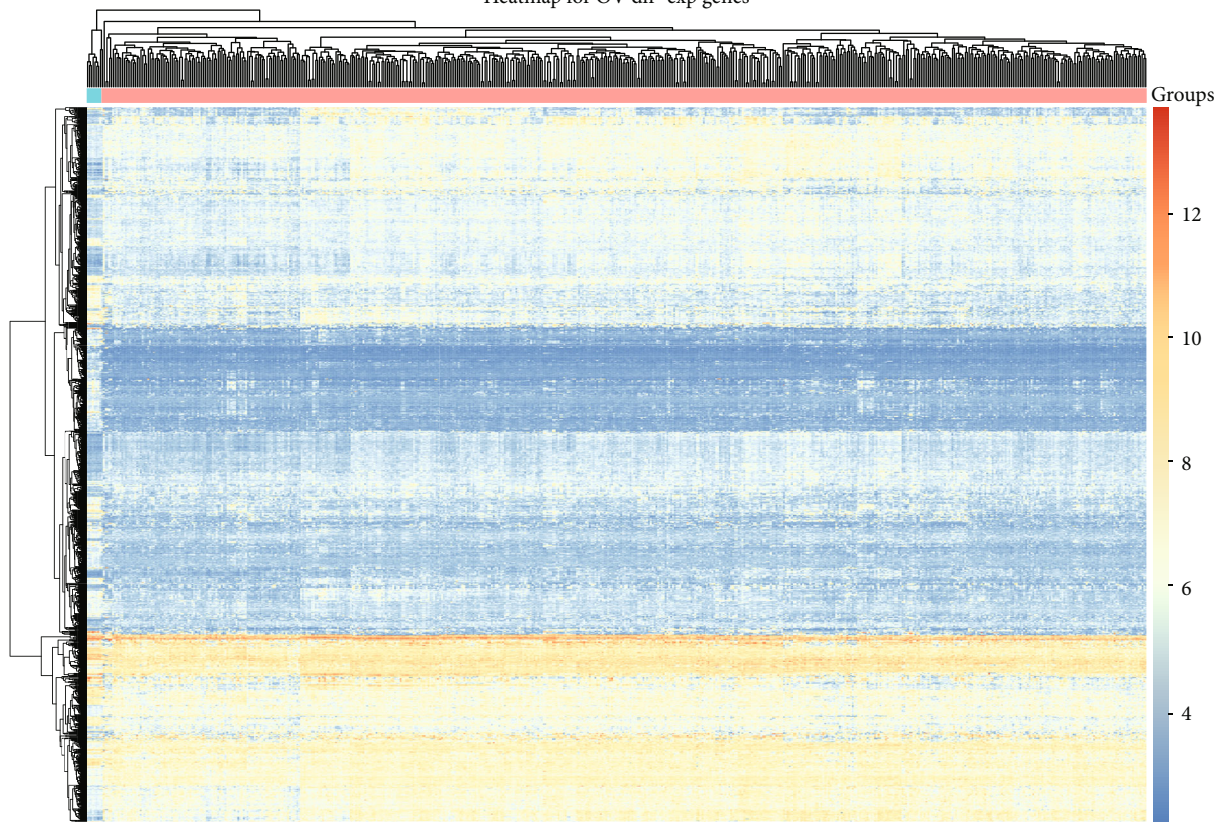
3.4. Identification of Differentially Expressed Genes Using TCGA Ovarian Cancer Datasets. A total of 1,124 differentially expressed genes with $|\log 2FC| > 1$ and adjusted p value < 0.05 were identified between 568 ovarian cancer samples and 8 normal samples (Figures 7(a) and 7(b)). GO enrichment analysis results showed that upregulated genes were primarily enriched in intracellular membrane-bounded organelle, nucleus, nuclear lumen, cytosol, nucleoplasm, cellular nitrogen compound metabolic process, heterocycle metabolic process, cellular aromatic compound metabolic process, and protein metabolic process (Figure 7(c)). Meanwhile, upregulated genes were involved in cell cycle, Herpes simplex virus 1 infection, human papillomavirus infection, human T cell leukemia virus 1 infection, and PI3K-Akt signaling pathway (Figure 7(d)). Downregulated genes primarily participated in multicellular organism development, plasma membrane, cytosol, vesicle, animal organ development, extracellular exosome, extracellular vesicle, positive regulation of cellular metabolic process, cellular response to organic substance, and positive regulation of nitrogen compound metabolic process (Figure 7(e)). In Figure 7(f), downregulated genes were mainly enriched in MAPK, metabolic, pathways in cancer, PI3K-Akt, and Ras signaling pathways.

3.5. Identification of Differentially Expressed Marker Genes Associated with Prognosis of Ovarian cancer. All marker genes were overlapped with 1,124 differentially expressed genes in TCGA samples. Survival analysis was used for identifying prognosis-related differentially expressed marker genes. The results showed that marker genes STAT1, ANP32E, GPRC5A, and EGFL6 were highly expressed in ovarian cancer (Figures 8(a)–8(d)). Furthermore, marker genes PMP22, FBXO21, and CYB5R3 were lowly expressed in ovarian cancer (Figures 8(e)–8(g)). The high expression of ANP32E ($p = 0.031$, HR: 0.79 (0.64-0.98)), STAT1 ($p = 0.005$, HR: 0.74 (0.59-0.91)), GPRC5A ($p = 0.03$, HR: 1.27 (1.02-1.57)), EGFL6 ($p = 0.018$, HR: 0.77 (0.62-0.96)), and PMP22 ($p = 0.043$, HR: 1.25 (1.01-1.54)) was significantly associated with better overall survival time than their low expression (Figures 9(a)–9(e)). The high expression of



(a)

Heatmap for OV dif-exp genes



(b)

FIGURE 7: Continued.

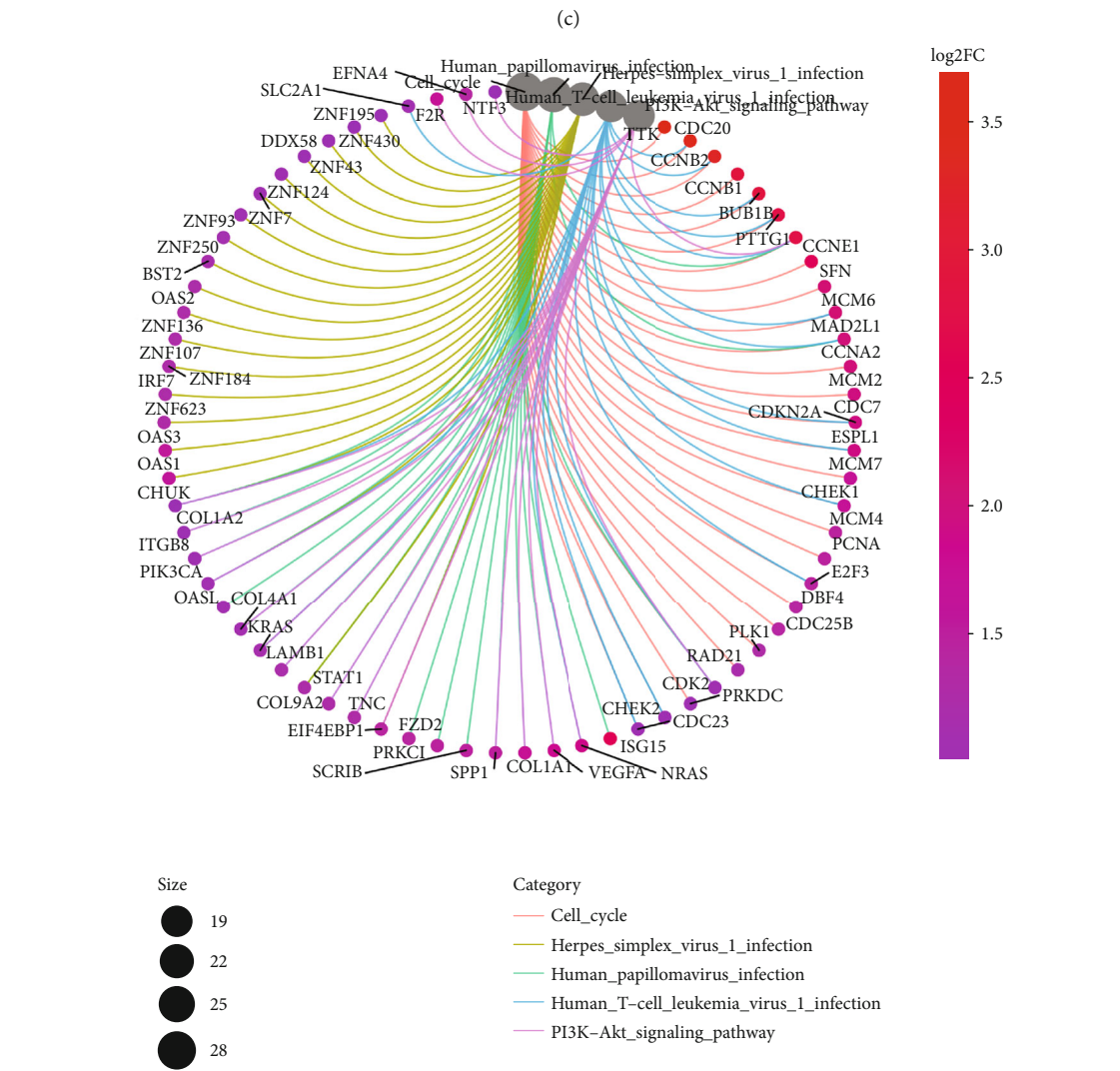
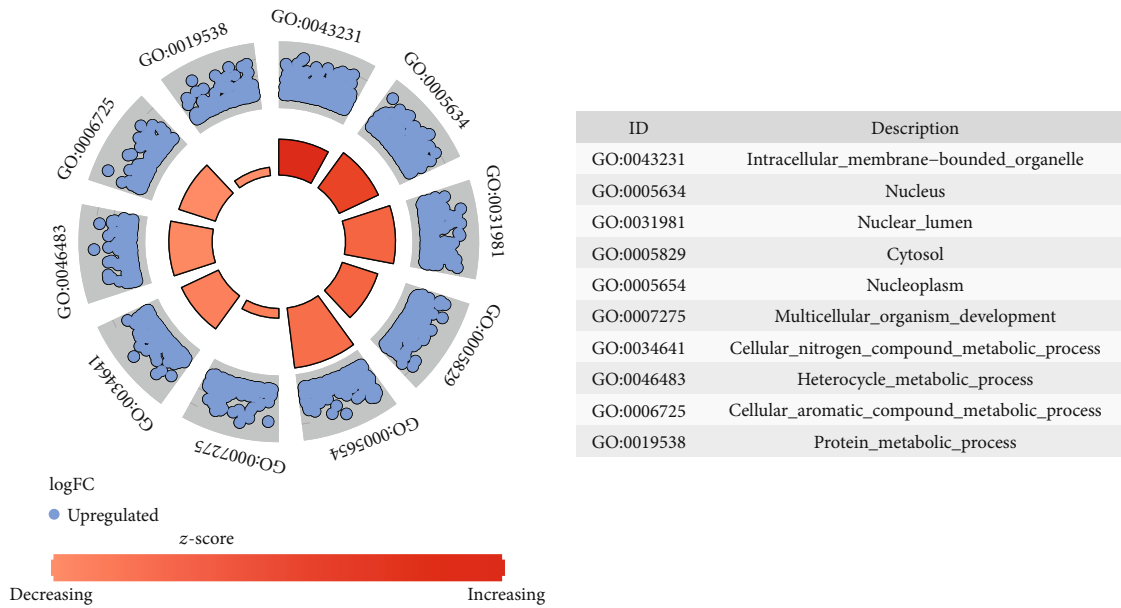
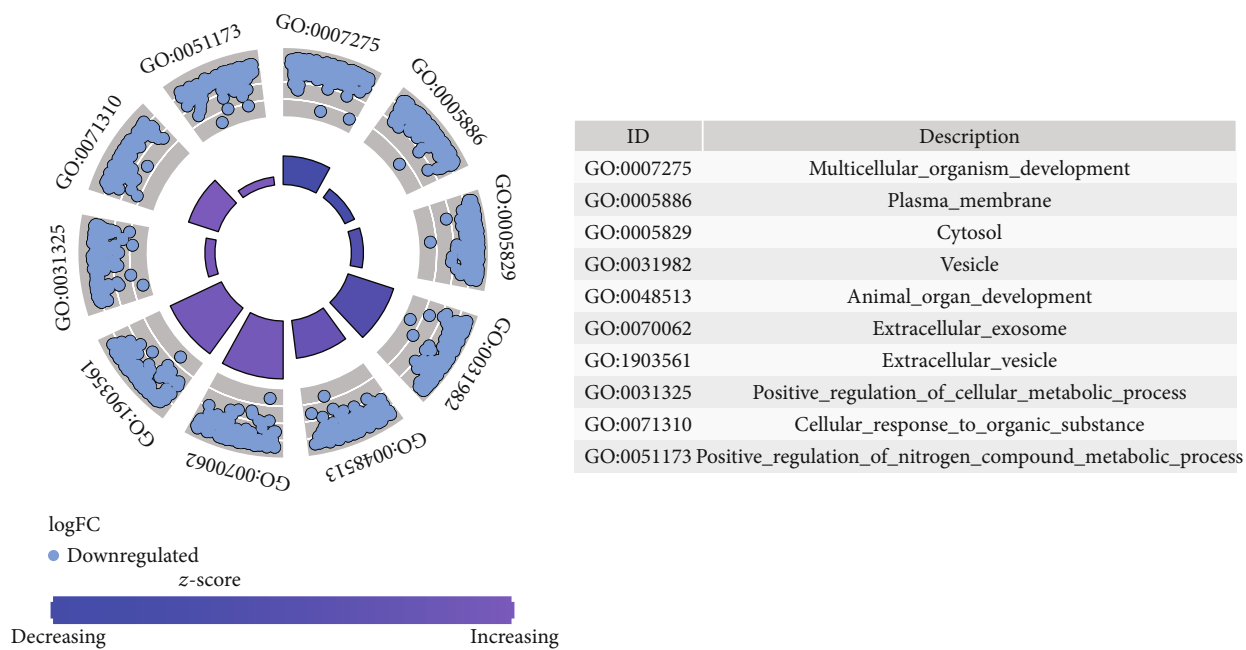
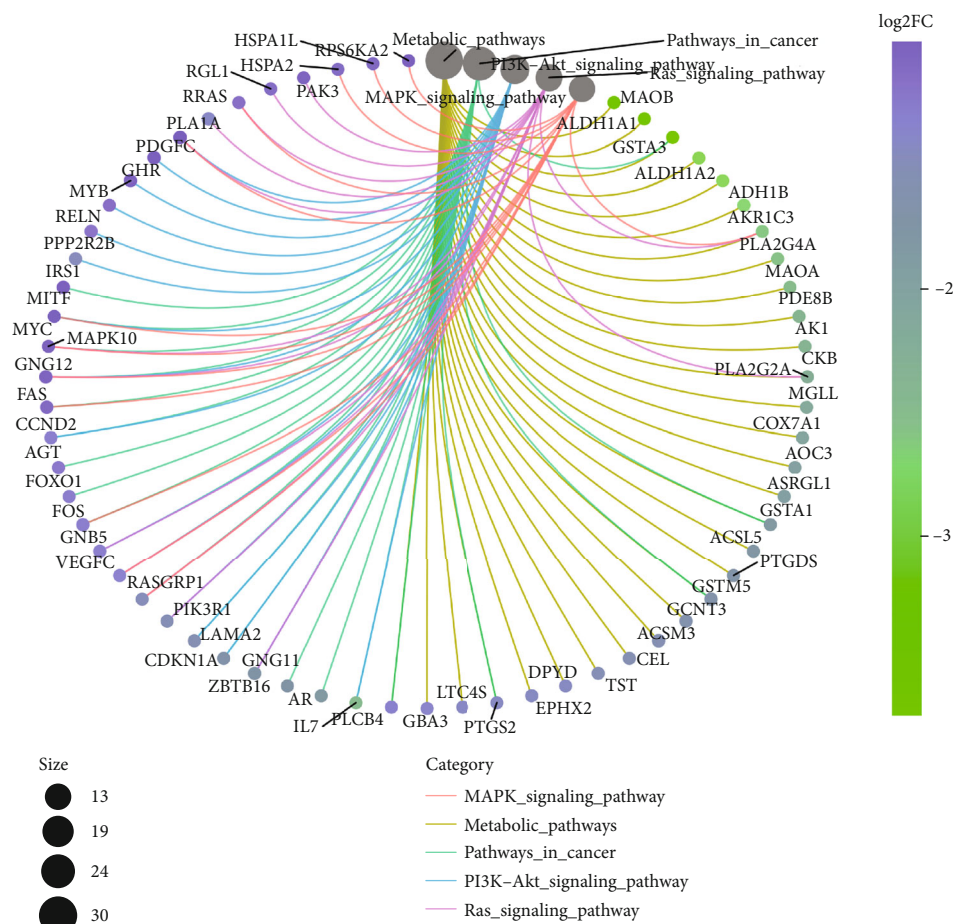


FIGURE 7: Continued.



(e)



(f)

FIGURE 7: Differentially expressed genes of ovarian cancer. (a, b) Volcano plots and heatmap showing the differentially expressed genes with $|\log_2FC| > 1$ and adjusted p value < 0.05 between ovarian cancer and normal tissues, respectively. (c, d) GO and KEGG enrichment analysis results of upregulated genes. (e, f) GO and KEGG enrichment analysis results of downregulated genes.

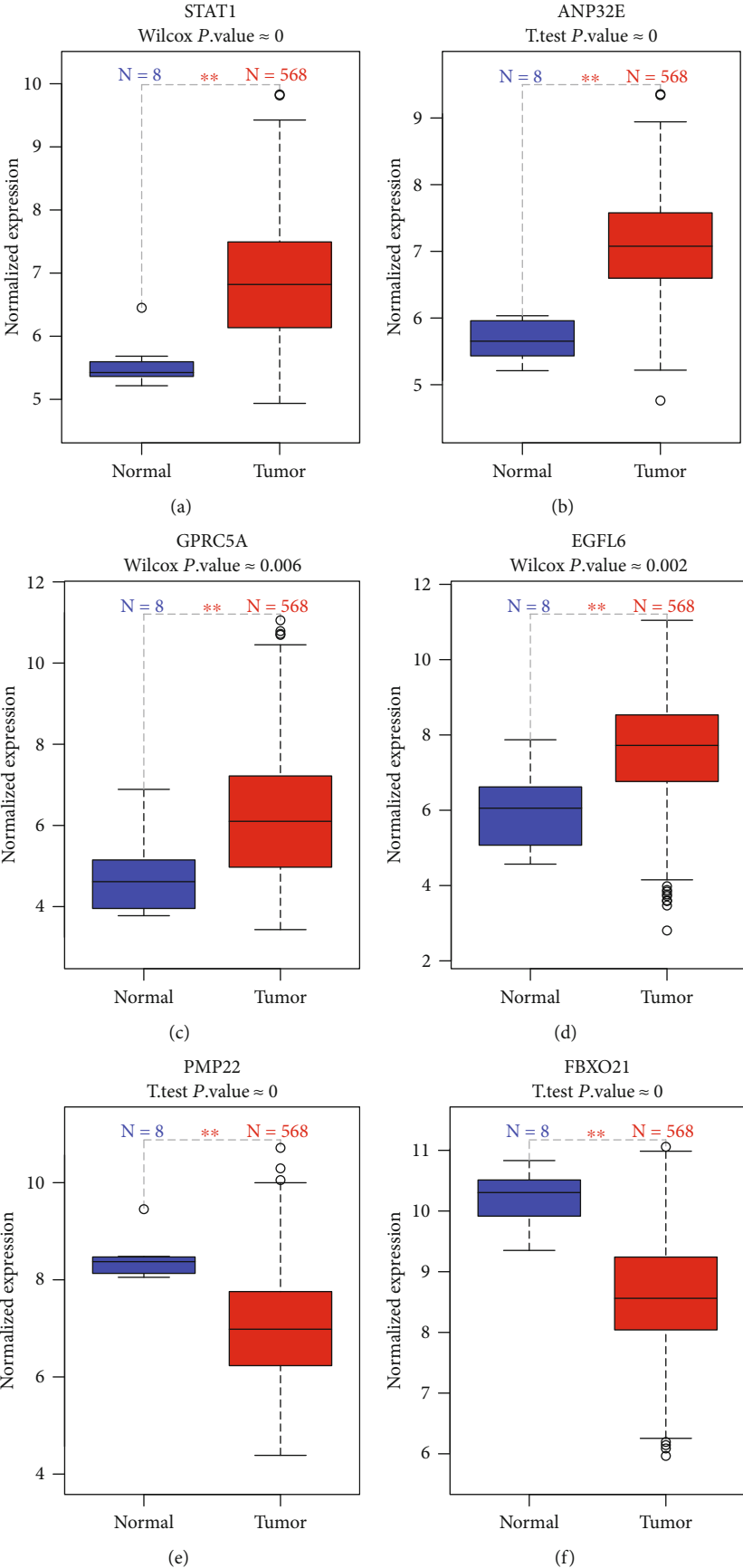


FIGURE 8: Continued.

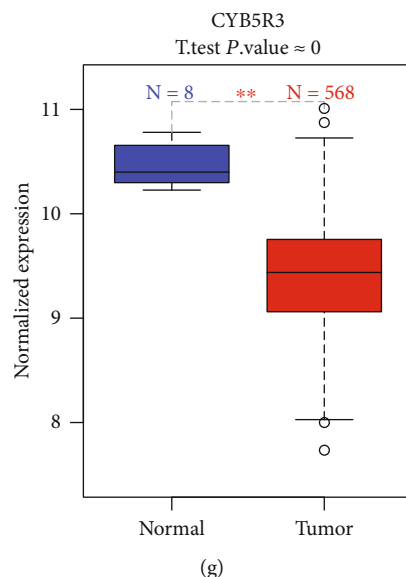


FIGURE 8: The differential expression of marker genes associated with prognosis of ovarian cancer. (a) STAT1; (b) ANP32E; (c) GPRC5A; (d) EGFL6; (e) PMP22; (f) FBXO21; (g) CYB5R3.

FBXO21 ($p = 0.027$, HR: 0.57 (0.35-0.94)), ANP32E ($p = 0.007$, HR: 0.51 (0.31-0.84)), and CYB5R3 ($p = 0.015$, HR: 1.86 (1.12-3.08)) indicated better recurrence-free survival time compared with their low expression (Figures 9(f)–9(h)). Furthermore, we found that STAT1 had the highest expression in stage II among all stages (Figure 10(a)). PMP22 had the highest expression in stage III among all stages (Figure 10(b)).

4. Discussion

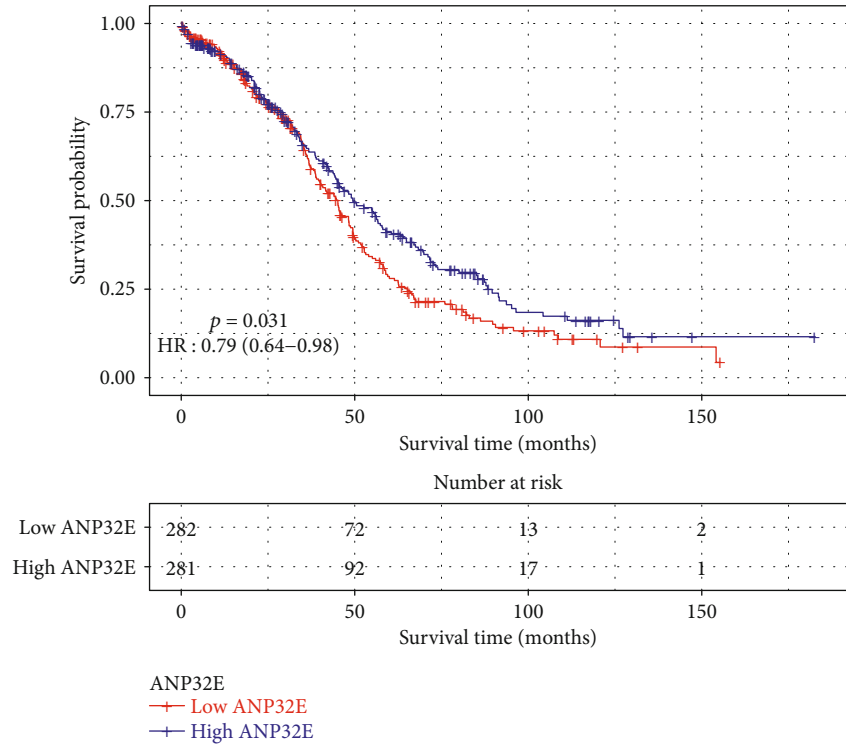
The treatment of ovarian cancer is complicated by the heterogeneity of the tumor. Different histological types of epithelial ovarian cancer have different cell origins, different mutation profiles, and different prognosis [16, 17]. Even in a histological type, different molecular subtypes with different prognosis can be found. To solve these problems, it is necessary to better characterize the heterogeneity of these ovarian cancer cells, to find reliable biomarkers, and develop appropriate targeted therapies. Single-cell RNA sequencing technology can explore the intercellular heterogeneity at the single-cell level and reconstruct lineage hierarchies. This method allows an unbiased analysis of the heterogeneity profile within a population of cells as it utilizes transcriptome reconstitution from a single cell. Our reanalysis of the ovarian cancer single-cell transcriptome may provide a deeper insight into the heterogeneity spectrum of ovarian cancer cells.

Totally, 66 ovarian cancer cells were included in our study. To remove cells with low quality, quality control was performed using the Seurat package. Proliferation induced by abnormal regulation of the cell cycle is thought to be critical for ovarian cancer progression. The G1/S phase is the most critical rate-limiting step in cell cycle promotion. Some studies have shown that the expression of cell cycle-related genes is significantly associated with poor prognosis in patients with ovarian cancer. Therefore, we studied molecules involved in

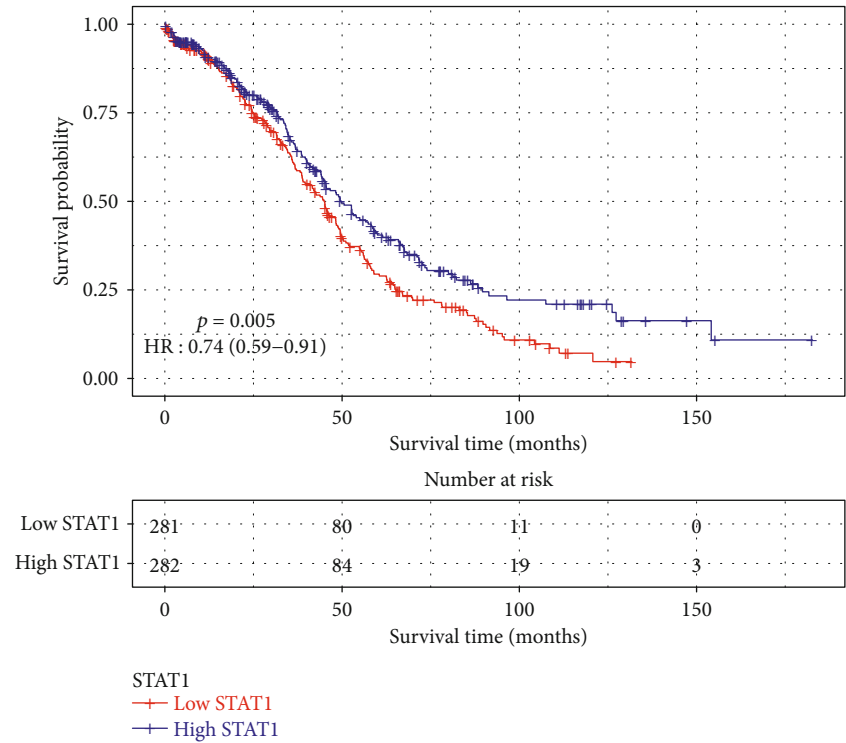
cell cycle progression to discover new prognostic factors and therapeutic targets. In this study, 66 ovarian cancer cells were clustered into three groups (G1, G2M, and S). The marker genes in each cluster were identified. To explore potential biological processes and pathways, KEGG and GO enrichment analyses of these marker genes were performed. The results showed that the marker genes in each cluster were enriched in different biological processes and pathways.

Using ovarian cancer dataset from TCGA, a total of 1,124 differentially expressed genes with $|\log 2FC| > 1$ and adjusted p value < 0.05 were identified between 568 ovarian cancer tissues and 8 normal tissues. To explore potential biological processes and pathways, these differentially expressed genes were mainly enriched in metabolic pathways, pathways in cancer, PI3K-Akt signaling pathway, and the like. For example, most ovarian cancer cells are highly proliferative; therefore, they are highly dependent on the metabolism of glucose by the aerobic glycolysis or the Warburg effect [18, 19]. PI3K-Akt signaling pathway is deregulated in various malignant cancers including ovarian cancer, which participates in tumor cell proliferation, survival, metabolism, and angiogenesis [20, 21].

The intercellular heterogeneity is one of the major drivers of cancer progression [22]. Gene variation at the single-cell level can rapidly produce cancer heterogeneity [23]. Prognosis-related differentially expressed marker genes were identified. We found that the expression levels of STAT1, ANP32E, GPRC5A, and EGFL6 were all significantly higher in ovarian cancer tissues compared with normal tissues. Furthermore, PMP22, FBXO21, and CYB5R3 expression was significantly lower in ovarian cancer tissues compared with normal tissues. The low expression of ANP32E, STAT1, GPRC5A, EGFL6, and PMP22 was positively associated with overall survival time of ovarian cancer. The low expression of FBXO21, ANP32E, and CYB5R3 was significantly associated with longer recurrence-free survival

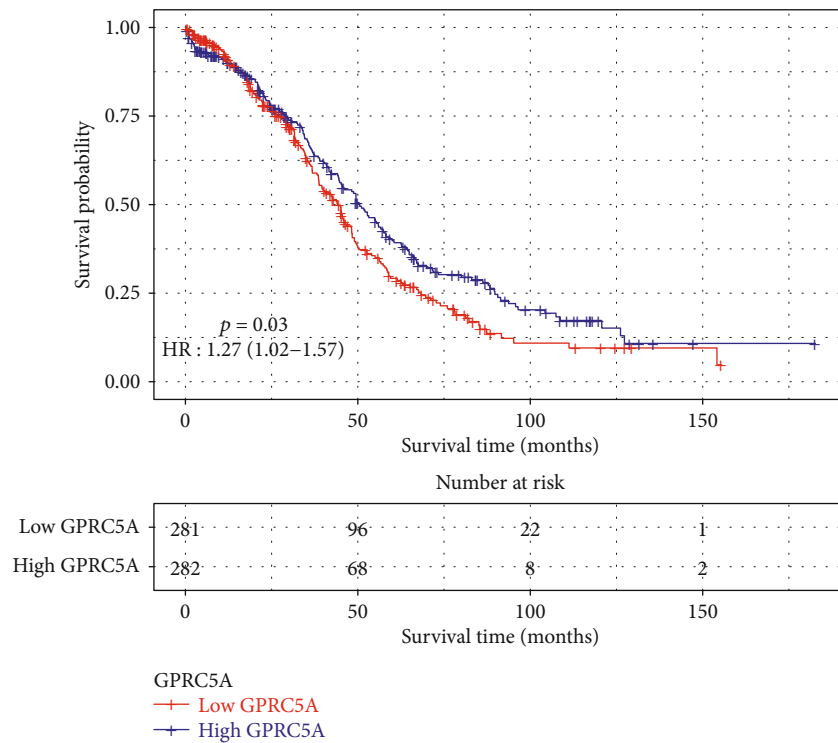


(a)

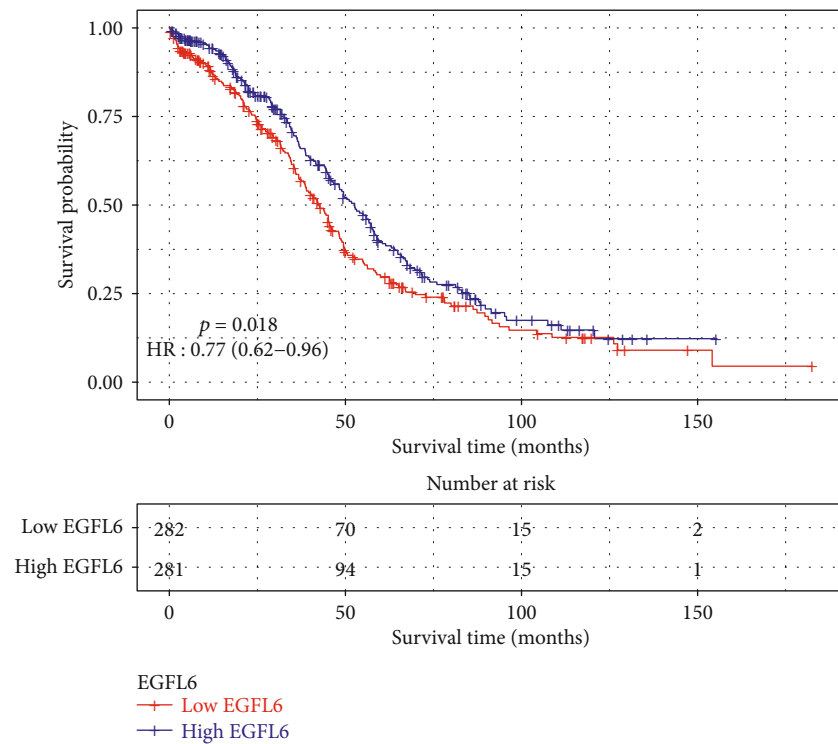


(b)

FIGURE 9: Continued.

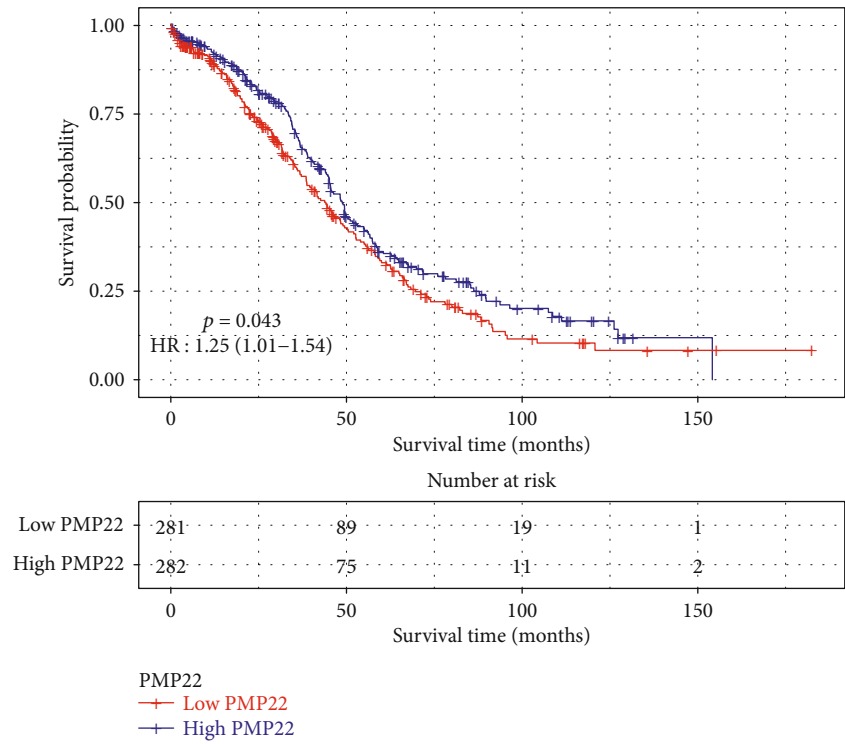


(c)

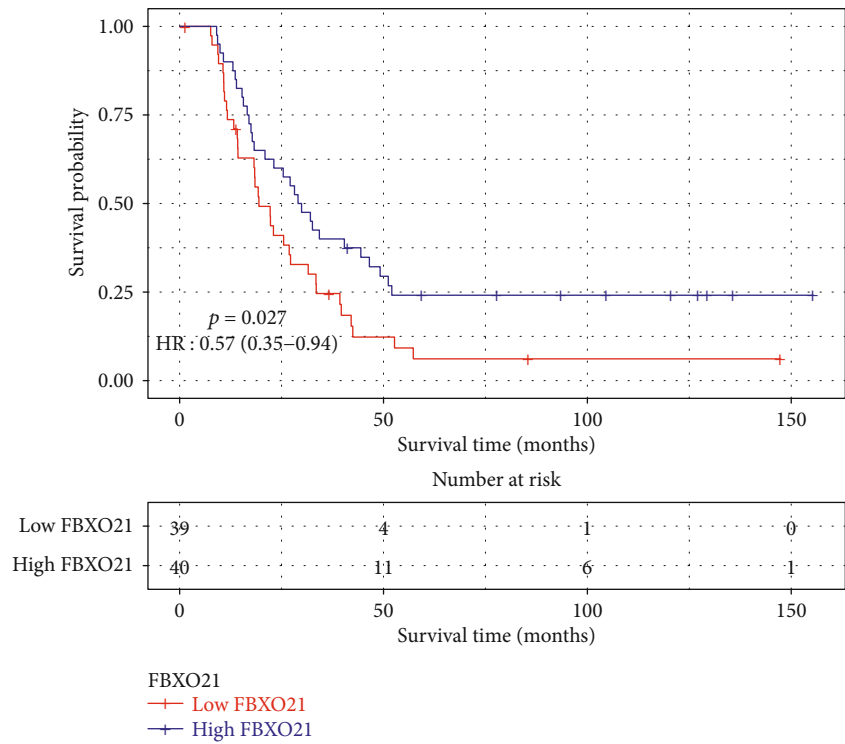


(d)

FIGURE 9: Continued.



(e)



(f)

FIGURE 9: Continued.

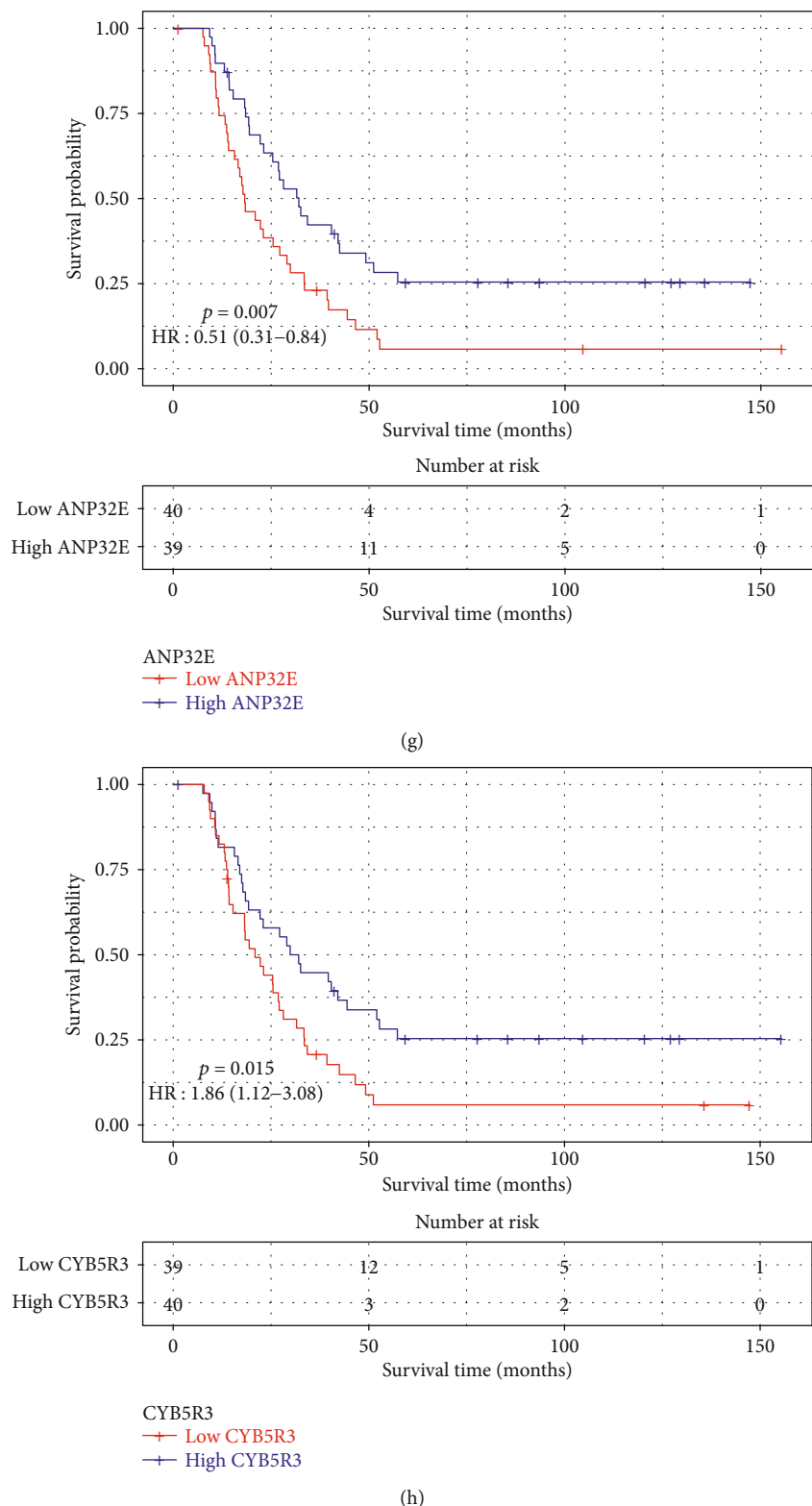


FIGURE 9: The survival analysis of differentially expressed marker genes in ovarian cancer. (a-e) The overall survival analysis results of ANP32E, STAT1, GPRC5A, EGFL6, and PMP22. (f-h) The recurrence-free survival analysis results of FBXO21, ANP32E, and CYB5R3.

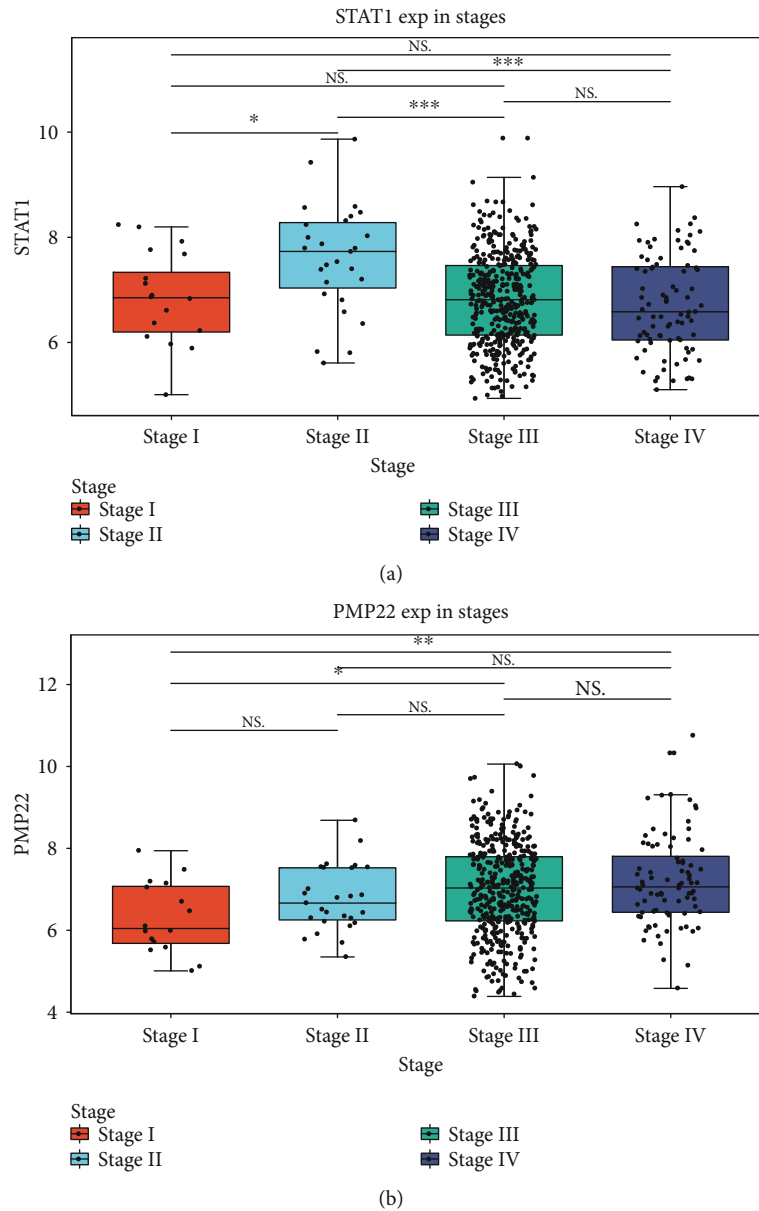


FIGURE 10: The differential expression of STAT1 and PMP22 across different stages in ovarian cancer. (a) STAT1; (b) PMP22.

time of ovarian cancer. STAT1, a member of STAT family, has been confirmed to be highly expressed in ovarian cancer [24, 25]. The high expression of ANP32E is in association with better prognosis, contributing to the proliferation and tumorigenesis of triple-negative breast cancer cells [26, 27]. GPRC5A variants may drive self-renewal of bladder cancer stem cells according to single-cell RNA-seq analysis [28]. EGFL6, a stem cell regulator expressed in ovarian tumor cells and vasculature, may induce the growth and metastasis of ovarian cancer [29, 30]. A previous study has found that EGFL6 is upregulated in drug-resistant ovarian cancer cell lines using microarray analysis [31]. The expression and function of PMP22 in tumors remain unclear. Some studies have shown that PMP22 is a potential tumor suppressor, and others have indicated that PMP22 has a potential carcinogenic function in

tumors [32–35]. Studies on the role of PMP22 in the regulation of ovarian cancer have not been reported. Furthermore, there is no report concerning the expression and role of FBXO21 and CYB5R3 in ovarian cancer. Collectively, our study identified specific cell subpopulations and marker genes in ovarian cancer.

5. Conclusion

In our study, we analyzed the intercellular heterogeneity in ovarian cancer using single-cell RNA sequencing and identified marker genes in each cluster. Combining TCGA ovarian cancer dataset, we identified differentially expressed marker genes that were significantly associated with prognosis of ovarian cancer, including ANP32E, STAT1, GPRC5A, EGFL6, PMP22, FBXO21, and CYB5R3.

Abbreviations

RNA-seq:	RNA sequencing
GEO:	Gene Expression Omnibus
PCA:	Principal component analysis
TCGA:	The Cancer Genome Atlas
FC:	Fold change
UMAP:	Uniform Manifold Approximation and Projection
GO:	Gene Ontology
KEGG:	Kyoto Encyclopedia of Genes and Genomes
BP:	Biological process
CC:	Cellular component
MF:	Molecular function.

Data Availability

The data used to support the findings of this study are included within the supplementary information files.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Yan Li and Juan Wang contributed equally to this work.

Supplementary Materials

Supplementary 1. Supplementary table 1: the IDs of TCGA ovarian cancer samples.

Supplementary 2. Supplementary table 2: the list of genes in each cluster.

References

- [1] Y. Jiang, T. Lyu, X. Che, N. Jia, Q. Li, and W. Feng, "Overexpression of SMYD3 in ovarian cancer is associated with ovarian cancer proliferation and apoptosis via methylating H3K4 and H4K20," *Journal of Cancer*, vol. 10, no. 17, pp. 4072–4084, 2019.
- [2] A. McPherson, A. Roth, E. Laks et al., "Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer," *Nature Genetics*, vol. 48, no. 7, pp. 758–767, 2016.
- [3] S. E. Bulun, Y. Wan, and D. Matei, "Epithelial mutations in endometriosis: link to ovarian cancer," *Endocrinology*, vol. 160, no. 3, pp. 626–638, 2019.
- [4] S. McGuire, B. Kara, P. C. Hart et al., "Inhibition of fascin in cancer and stromal cells blocks ovarian cancer metastasis," *Gynecologic Oncology*, vol. 153, no. 2, pp. 405–415, 2019.
- [5] F. C. Nielsen, T. van Overeem Hansen, and C. S. Sørensen, "Hereditary breast and ovarian cancer: new genes in confined pathways," *Nature Reviews. Cancer*, vol. 16, no. 9, pp. 599–612, 2016.
- [6] B. Davidson, "Recently identified drug resistance biomarkers in ovarian cancer," *Expert Review of Molecular Diagnostics*, vol. 16, no. 5, pp. 569–578, 2016.
- [7] D. Yuan, Y. Tao, G. Chen, and T. Shi, "Systematic expression analysis of ligand-receptor pairs reveals important cell-to-cell interactions inside glioma," *Cell communication and signaling*, vol. 17, no. 1, p. 48, 2019.
- [8] C. Blassl, J. D. Kuhlmann, A. Webers, P. Wimberger, T. Fehm, and H. Neubauer, "Gene expression profiling of single circulating tumor cells in ovarian cancer - establishment of a multi-marker gene panel," *Molecular Oncology*, vol. 10, no. 7, pp. 1030–1042, 2016.
- [9] A. Jiménez-Sánchez, D. Memon, S. Pourpe et al., "Heterogeneous tumor-immune microenvironments among differentially growing metastases in an ovarian cancer patient," *Cell*, vol. 170, no. 5, pp. 927–938.e20, 2017.
- [10] T. Baslan and J. Hicks, "Unravelling biology and shifting paradigms in cancer with single-cell sequencing," *Nature Reviews. Cancer*, vol. 17, no. 9, pp. 557–569, 2017.
- [11] K. R. Campbell, A. Steif, E. Laks et al., "clonealign: statistical integration of independent single-cell RNA and DNA sequencing data from human cancers," *Genome biology*, vol. 20, no. 1, 2019.
- [12] D. Abrams, P. Kumar, R. K. M. Karuturi, and J. George, "A computational method to aid the design and analysis of single cell RNA-seq experiments for cell type identification," *BMC Bioinformatics*, vol. 20, no. S11, 2019.
- [13] E. Papalexi and R. Satija, "Single-cell RNA sequencing to explore immune cell heterogeneity," *Nature Reviews. Immunology*, vol. 18, no. 1, pp. 35–45, 2018.
- [14] B. J. Winterhoff, M. Maile, A. K. Mitra et al., "Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells," *Gynecologic Oncology*, vol. 144, no. 3, pp. 598–606, 2017.
- [15] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, and R. Satija, "Integrating single-cell transcriptomic data across different conditions, technologies, and species," *Nature Biotechnology*, vol. 36, no. 5, pp. 411–420, 2018.
- [16] V. Krishnan, J. S. Berek, and O. Dorigo, "Immunotherapy in ovarian cancer," *Current Problems in Cancer*, vol. 41, no. 1, pp. 48–63, 2017.
- [17] V. Rojas, K. M. Hirshfield, S. Ganesan, and L. Rodriguez-Rodriguez, "Molecular characterization of epithelial ovarian cancer: implications for diagnosis and treatment," *International Journal of Molecular Sciences*, vol. 17, no. 12, 2016.
- [18] Z. Ai, Y. Lu, S. Qiu, and Z. Fan, "Overcoming cisplatin resistance of ovarian cancer cells by targeting HIF-1-regulated cancer metabolism," *Cancer Letters*, vol. 373, no. 1, pp. 36–44, 2016.
- [19] P. K. Chakraborty, S. B. Mustafi, X. Xiong et al., "MICU1 drives glycolysis and chemoresistance in ovarian cancer," *Nature Communications*, vol. 8, no. 1, 2017.
- [20] H. Li, J. Zeng, and K. Shen, "PI3K/AKT/mTOR signaling pathway as a therapeutic target for ovarian cancer," *Archives of Gynecology and Obstetrics*, vol. 290, no. 6, pp. 1067–1078, 2014.
- [21] S. Mabuchi, H. Kuroda, R. Takahashi, and T. Sasano, "The PI3K/AKT/mTOR pathway as a therapeutic target in ovarian cancer," *Gynecologic Oncology*, vol. 137, no. 1, pp. 173–179, 2015.
- [22] D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul et al., "Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types," *Science*, vol. 343, no. 6172, pp. 776–779, 2014.

- [23] U. M. Litzénburger, J. D. Buenrostro, B. Wu et al., "Single-cell epigenomic variability reveals functional cancer heterogeneity," *Genome Biology*, vol. 18, no. 1, 2017.
- [24] J. A. Josahkian, F. P. Saggiaro, T. Vidotto et al., "Increased STAT1 expression in high grade serous ovarian cancer is associated with a better outcome," *International journal of gynecological cancer*, vol. 28, no. 3, pp. 459–465, 2018.
- [25] X. Tian, W. Guan, L. Zhang et al., "Physical interaction of STAT1 isoforms with TGF- β receptors leads to functional crosstalk between two signaling pathways in epithelial ovarian cancer," *Journal of experimental & clinical cancer research*, vol. 37, no. 1, 2018.
- [26] A. Obri, K. Ouararhni, C. Papin et al., "ANP32E is a histone chaperone that removes H2A.Z from chromatin," *Nature*, vol. 505, no. 7485, pp. 648–653, 2014.
- [27] Z. Xiong, L. Ye, H. Zhenyu et al., "ANP32E induces tumorigenesis of triple-negative breast cancer cells by upregulating E2F1," *Molecular Oncology*, vol. 12, no. 6, pp. 896–912, 2018.
- [28] Z. Yang, C. Li, Z. Fan et al., "Single-cell Sequencing Reveals Variants in *_ARID1A_*, *_GPRC5A_* and *_MLL2_* Driving Self-renewal of Human Bladder Cancer Stem Cells," *European Urology*, vol. 71, no. 1, pp. 8–12, 2017.
- [29] S. Bai, P. Ingram, Y. C. Chen et al., "EGFL6 regulates the asymmetric division, maintenance, and metastasis of ALDH+ ovarian cancer cells," *Cancer Research*, vol. 76, no. 21, pp. 6396–6409, 2016.
- [30] K. Noh, L. S. Mangala, H. D. Han et al., "Differential effects of EGFL6 on tumor versus wound angiogenesis," *Cell Reports*, vol. 21, no. 10, pp. 2785–2795, 2017.
- [31] R. Januchowski, P. Zawierucha, M. Rucinski, and M. Zabel, "Microarray-based detection and expression analysis of extracellular matrix proteins in drugresistant ovarian cancer cell lines," *Oncology Reports*, vol. 32, no. 5, pp. 1981–1990, 2014.
- [32] W. Cai, G. Chen, Q. Luo et al., "PMP22 regulates self-renewal and chemoresistance of gastric cancer cells," *Molecular Cancer Therapeutics*, vol. 16, no. 6, pp. 1187–1198, 2017.
- [33] H. Liu, H. Q. Cao, J. B. Ta, W. Zhang, and Y. H. Liu, "Knock-down of peripheral myelin protein 22 inhibits the progression of chronic myeloid leukemia," *Oncology Research*, vol. 22, no. 5, pp. 259–265, 2015.
- [34] H. Qu, M. Zhu, Y. Tao, and Y. Zhao, "Suppression of peripheral myelin protein 22 (PMP22) expression by miR29 inhibits the progression of lung cancer," *Neoplasma*, vol. 62, no. 6, pp. 881–886, 2015.
- [35] S. Winslow, K. Leandersson, and C. Larsson, "Regulation of PMP22 mRNA by G3BP1 affects cell proliferation in breast cancer cells," *Molecular Cancer*, vol. 12, no. 1, 2013.

Research Article

A Two-Phase Mitosis Detection Approach Based on U-Shaped Network

Wenjing Lu 

School of Information Engineering, Harbin University, China

Correspondence should be addressed to Wenjing Lu; luwj116@outlook.com

Received 29 July 2021; Accepted 11 September 2021; Published 5 October 2021

Academic Editor: Qiushi Zhao

Copyright © 2021 Wenjing Lu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a deep learning-based method for mitosis detection in breast histopathology images. A main problem in mitosis detection is that most of the datasets only have weak labels, i.e., only the coordinates indicating the center of the mitosis region. This makes most of the existing powerful object detection methods hardly be used in mitosis detection. Aiming at solving this problem, this paper firstly applies a CNN-based algorithm to pixelwisely segment the mitosis regions, based on which bounding boxes of mitosis are generated as strong labels. Based on the generated bounding boxes, an object detection network is trained to accomplish mitosis detection. Experimental results show that the proposed method is effective in detecting mitosis, and the accuracies outperform state-of-the-art literatures.

1. Introduction

Breast cancer is one of the main threats to woman health and becomes one of the most leading causes of cancer-related death all over the world. Early diagnosis is believed to be an effective way for promoting the prognosis of breast cancer.

Generally, breast cancer can be classified into three levels in histopathology based on the morphological microstructure of cancerous and the normal cells, i.e., well differentiated, poorly differentiated, and intermediate. Classification is important to the diagnosis and prognosis of breast cancer. The most commonly used classification standard is the BRE system proposed by WHO, in which three indications are used to evaluate the differentiation level. The indications are vasculogenesis degree, nuclear atypia, and mitotic counting.

Among the indications, mitotic counting is the most important, which can be described as the number of cells under mitosis in tumor and around areas. In traditional methods, mitotic counting is done by pathologists. Since the nuclei of cells experiencing mitosis are extremely small, therefore, attention should be highly concentrated. Moreover, the morphology of cells under various stages of mitosis is different, and there may exist enormous normal cells

which are similar to mitotic cells. Due to these reasons, mitotic counting is a tedious and error-prone task.

In order to reduce the workload of pathologists, many computer algorithms and systems are proposed to automatically detect mitosis. Traditional automatic mitosis detection methods usually extract handcrafted features from breast histopathology images and train a machine learning algorithm and then perform predictions on testing images based on the trained model. The key issue of such methods is the feature definition. Effective features can greatly increase the accuracy of detection, while badly defined features may dramatically influence the accuracy.

Recently, deep learning has attracted the attention of researchers and becomes a new focus of computer vision. Convolutional neural networks (CNNs) are applied to images, and discriminative features are extracted under appropriate loss functions. Compared with traditional learning-based methods, the most notable advantage of CNN-based ones is that no human interventions are needed throughout the whole procedure. In consideration of the superiority of CNNs, many researchers begin to employ CNN-based methods to detect mitosis in breast histopathology images and achieve competent performances.

In this paper, a CNN-based mitosis detection method is proposed. The whole procedure can be separated into two phases. For the first phase, in consideration of the lack of strong labels in current datasets, a U-shaped network is trained over pixellevel-labeled datasets and used to predict bounding boxes on datasets which only have weak labels. In the second phase, the predicted bounding boxes are then taken as strong labels to train a network for mitosis detection. Contributions of this paper can be summarized as follows.

- (1) A mitosis detection method is proposed aiming at solving the problem of insufficiency of strongly labeled breast histopathology image dataset. We propose to use pixel-wise-labeled datasets to train a segmentation network, and then, strong labels (bounding boxes) can be generated based on the prediction of the segmentation network. Thus, most of the current weakly labeled datasets can be used for mitosis detection network training
- (2) For mitosis segmentation, a U-shaped network is trained using a pixelwisely labeled dataset. Benefiting from the multiple frequency downsampling and upsampling layers in the network, the ability of segmenting small targets is promoted, which is suitable for the mitosis segmentation task

The rest of this paper will be organized as follows. Section 2 gives a brief literature review of mitosis detection in breast histopathology images. Section 3 proposes the mitosis detection method with detailed network structure. Section 4 provides sufficient experiments and comparisons to show the effectiveness of the proposed method, and finally in Section 5, conclusions are drawn.

2. Related Works

Most of the traditional mitosis detection methods are based on image features which are manually designed by computer scientists and pathologists. Huang et al. [1] propose an exclusive independent component analysis (XICA) algorithm to detect mitosis. It is based on the fact that the mitotic nucleus is darker; then, it surrenders in color, and a sparse representation-based classifier is used to extract mitosis from candidates. Considering there may be distinct color variations in breast histopathological images, Tashk et al. [2] propose to use texture features for mitosis detection. An object-oriented complete local binary pattern is designed, and the support vector machine (SVM) is used to separate mitosis from background. Khan et al. [3] propose to employ a Gaussian mixture model (GMM) to model the distributions of mitosis pixels and background pixels, and a context-aware postprocessing is used to reduce false negatives. Tek et al. [4] investigate a set of generic features, i.e., color, binary shape-based, Laplacian, and morphological features to represent mitosis regions. The AdaBoost algorithm is then used to detect mitosis. All these methods employ handcrafted image features and a machine learning algo-

rithm to predict whether image pixels belong to the mitosis region or not. However, since mitotic morphology may vary greatly and the collection of tissue sections obtained by different instruments also increases the diversity of the appearances of histopathological images, image features such as color, texture, and intensity may be incompetent to fully represent mitosis. Besides, designing of such features requires rich experiences of computer scientists and pathologists, and therefore, such methods are not so satisfactory for automatic mitosis detection.

Recently, researchers begin to put their focus on CNN-based automatic mitosis detection methods. CNNs construct high-level semantic features from low-level features and obtain competent performances in many areas of computer vision such as classification, segmentation, and object detection [5–9]. Ciresan et al.'s team [10] is one of the earliest researchers who employ CNNs for mitosis detection. In their work, a CNN with sliding window and max pooling is proposed, and they achieved the first place in ICPR2012 mitosis detection competition. Wang et al. [11] propose a mitosis detection method which combines both handcrafted and CNN features. Although the computational load is reduced by incorporating manually designed features, the overall performance is limited due to the disability of handcrafted features in representing mitosis morphology. Veta et al. [12] propose a similar method and obtain the first place in ICPR2013 mitosis detection contest. Chen et al. [13] use a cascaded CNN to detect mitosis by constructing a two-stage deep network. A rough search network is used to search mitosis candidates and a discriminant network to further select mitosis from candidates. Inspired by the residual conception proposed by He et al. [14], Zerhouni et al. [15] propose a wide residual network (WRN) for mitosis detection. Recently, Li et al. [16] employ faster R-CNN [17] as the detector for mitosis detection. However, since the faster R-CNN is designed as a general purpose object detection network, it is hard to get satisfactory performance on mitosis detection task without sufficient training data.

A main difficulty of mitosis detection is that most existing datasets only have weak labels, i.e., only center points of mitotic nucleus are labeled. It is difficult to construct a valid training set to train a powerful detection network based on such labels. At most occasions, it is unavoidably for pathologists to perform a pixel-level or bounding box-level labeling on such datasets, which is a labor-consuming task. Some researchers intend to solve this problem. For example, Li et al. [18], Zerhouni et al. [15], Cai et al. [19], and Yancey [20] crop a square area around the label point as the region of interest. However, since mitosis is often heteromorphic in shape, a large amount of background pixels are included into the cropped area, which influences the detection performance.

3. Method

Mitosis detection can be seen as a special case of object detection. In the past few years, along with the successful employment of CNNs, object detection methods, such as R-CNN [21], fast R-CNN [22], faster R-CNN [17], SPP-Net [23], and YOLO [24], achieve reasonable accuracies

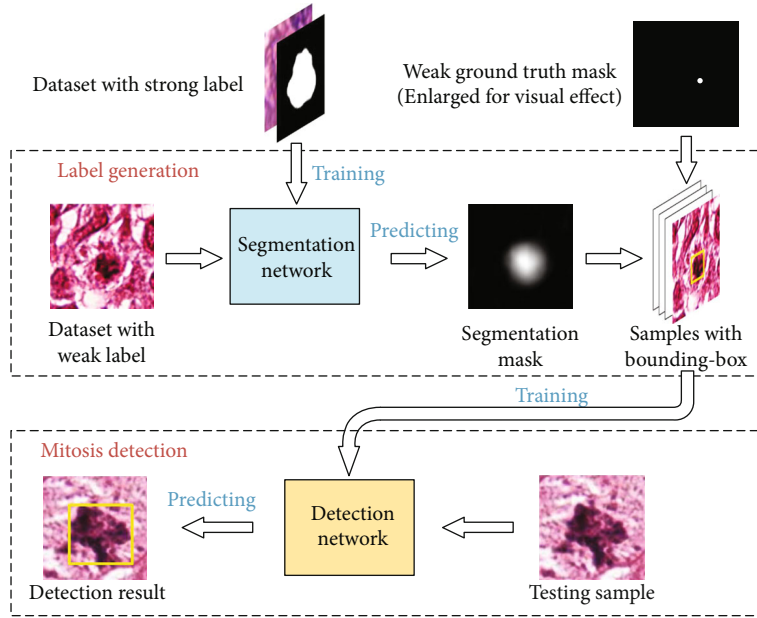


FIGURE 1: Flowchart of the proposed method.

and efficiencies. However, such general purpose object detection methods are hard to be employed into the mitosis detection task directly. The main reason is that such methods usually need to have a training dataset with labeled bounding boxes indicating regions of interest. In most existing breast histopathology image datasets, mitosis are weakly labeled, i.e., only the coordinates indicating the centers of mitosis regions are labeled. With such weakly labeled samples, a manual labeling is needed to generate bounding boxes in order to employ a detection algorithm.

To solve this problem, this paper proposes to generate bounding boxes for weakly labeled breast histopathology images and construct a mitosis detection method based on object detection networks. Firstly, a U-shaped network is trained using pixel level-labeled dataset, and this network is used for segmenting mitosis. Based on the segmentation and weak labels, bounding boxes of mitosis are generated for training a detection network. Then, the detection network is used to detect mitosis. The whole process is depicted in Figure 1. In this section, the two stages of the proposed method will be described in detail.

3.1. Label Generation. CNNs have been successfully used in biomedical image segmentation. Inspired by the U-Net [25] and the multilevel wavelet CNN (MWCNN) [26], a segmentation network is used in this paper to promote the ability of segmenting small targets. The structure of the network is depicted in Figure 2.

As shown in Figure 2, each CNN block is a 4-layered fully convolutional network (FCN) without pooling and takes the discrete wavelet transform (DWT) subband image as the input except the first layer. Low-frequency and high-frequency bands of DWT in CNN help to fully explore all frequency information of the input image, and the inverse wavelet transform (IWT) plays the role of reconstructing subband images into whole. Each layer of the CNN block

is composed of 3×3 convolution (Conv), batch normalization (BN), and rectified linear unit (ReLU) operations. In fact, this CNN block structure is the one that has been proven to be effective in network training by He et al. [14].

Similar to the U-Net, the pixelwise cross entropy is used as the loss function of the network, as defined in

$$E = \sum_x \log \left(p_{l(x)}(x) \right), \quad (1)$$

where l is the true label of each pixel and $p_k(x)$ is the softmax of pixel x in the output feature map, which is defined as

$$p_k(x) = \frac{\exp(a_k(x))}{\sum_{k'}^K \exp(a_{k'}(x))}, \quad (2)$$

where $a_k(x)$ denotes the activation in feature channel k of pixel x and K is the number of classes.

The network is trained using the MITOS2012 dataset [27] which has pixel-level strong labels. Input images and their corresponding segmentation maps are fed into the network for training. The trained network performs an end-to-end prediction, and the output feature map shares the same width and height as those of the input images. The output feature map indicates a probability of a pixel that it belongs to mitosis.

After getting the segmented mitosis, the minimum circumscribed rectangle is labeled as the bounding box of mitosis, which will be taken as the ground truth. In this paper, if the weak label (a point labeled at the center of the mitotic nucleus) is within the marked bounding box, the sample will be taken as positive. Figure 3 shows some examples of the labeled bounding boxes.

3.2. Mitosis Detection. After generating strong labels, the R-CNN [21] algorithm is employed for mitosis detection.

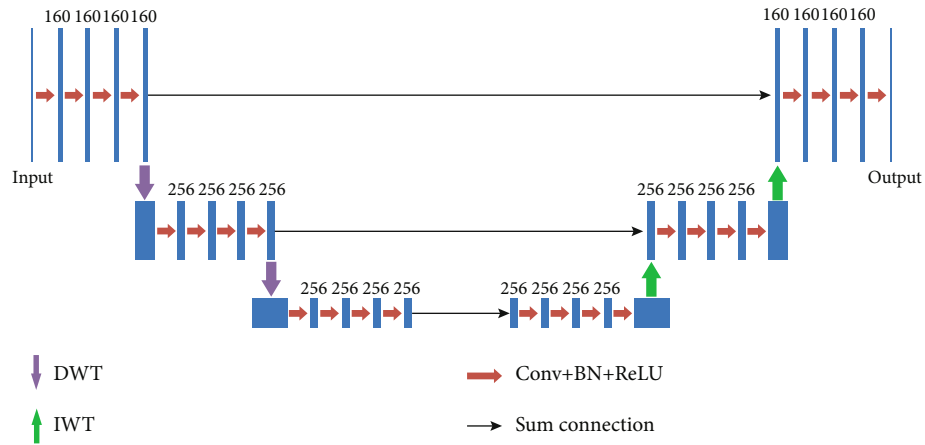


FIGURE 2: Architecture of the segmentation network.

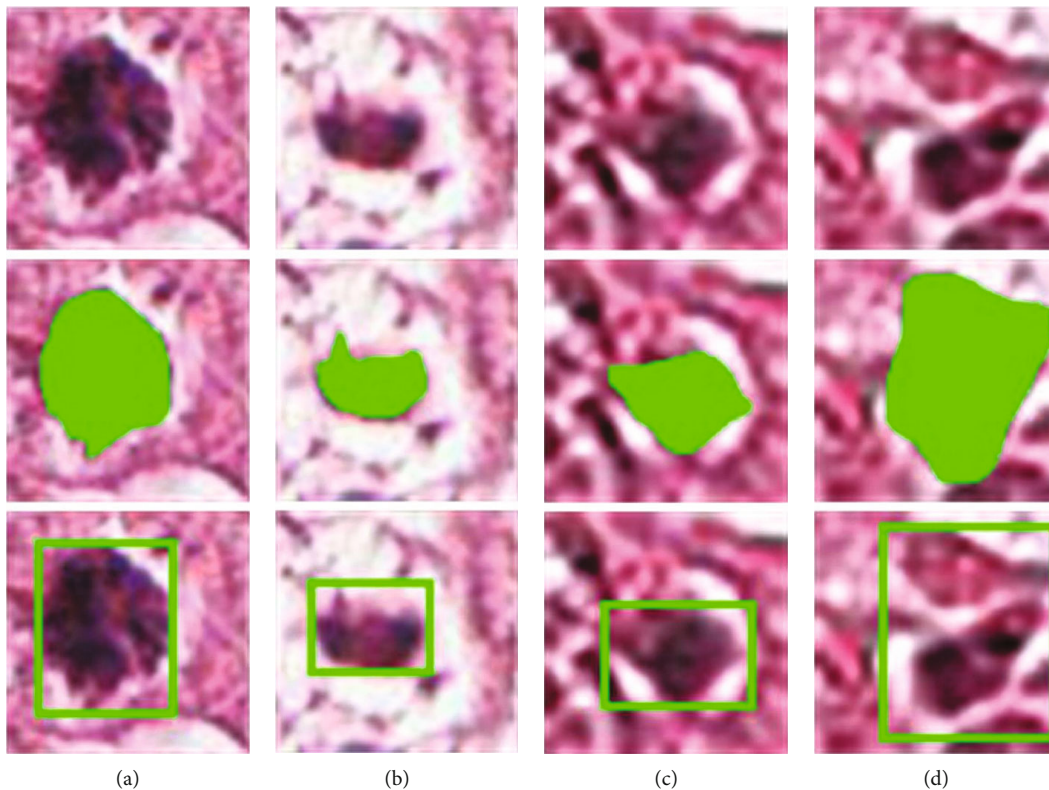


FIGURE 3: Examples of labeled bounding boxes. (A) Original. (B) Original with mask. (C) Original with bounding box. (a, b) True positives. (c, d) False positives.

R-CNN has been a representative and powerful CNN-based object detection method since recent years. R-CNN performs a four-step detection routine, i.e., region proposal generation, CNN-based feature extraction, SVM-based region classification, and bounding box regression. A selective search algorithm is employed to generate region proposals, which will be fed into a CNN to extract features after region wrapping. During the training phase, the generated bounding boxes in Section 3.1 are used as the ground truth. 32 positive samples and 96 negative samples are composed together as a mini-

batch, which is consistent to the original R-CNN, and are proven to be an optimized combination. It should be noticed that theoretically any object detection method can be employed here, and more powerful object detection method may obtain more accurate results.

4. Experimental Results

4.1. *Datasets.* In the experiments, three mostly used datasets for mitosis detection are employed, i.e., AMIDA2013 [12],

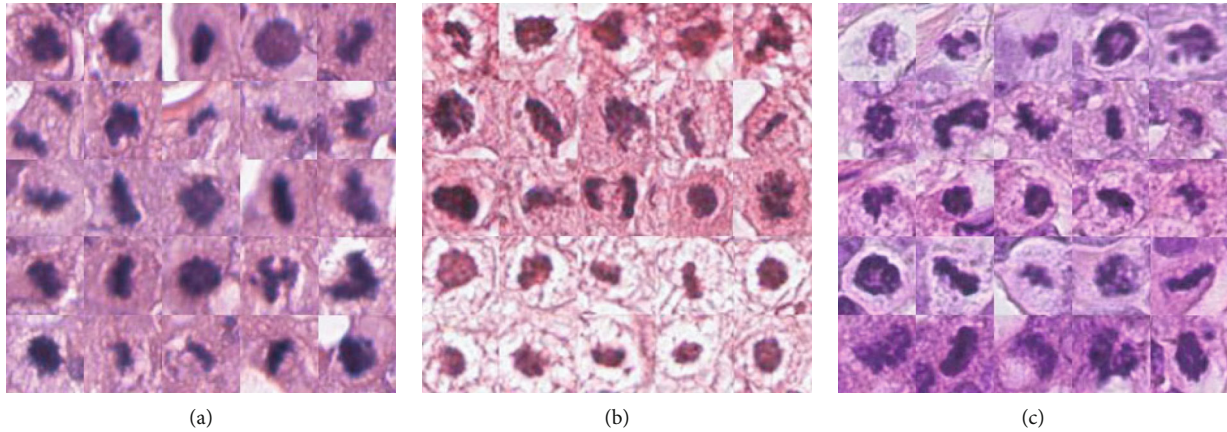


FIGURE 4: Exemplars from the datasets: (a) AMIDA2013, (b) ICPR2014, and (c) TUPAC2016.

ICPR2014 [28], and TUPAC2016 [29]. All these datasets contains H&E-stained breast cancer histopathological section view. Figure 4 shows some exemplars from the datasets.

The AMIDA2013 dataset consists of 1083 samples from 23 volunteers, and each volunteer has at least 10 views with dimension of 2000×2000 and resolution of $0.25 \text{ pixels}/\mu\text{m}$. Images from the AMIDA2013 dataset are obtained by Aperio Scanscope XT slice scanner. Images are labeled by experts from University Medical Center Utrecht, and each view is labeled by two experts independently. For the AMIDA2013 dataset, 550 samples of 12 volunteers are taken as the training set and others as the validation set.

The ICPR2014 dataset was firstly proposed for the ICPR2014 mitosis detection competition. It is composed of two subsets of images scanned by Aperio Scanscope XT and Hamamatsu Nanozoomer 2.0-HT digital slice scanner, respectively. The dimension of images from this dataset is 1539×1376 and the resolution is $0.2455 \text{ pixels}/\mu\text{m}$. The dataset is labeled by two pathologists independently, and 749 mitosis are labeled out of 1200 views from 11 volunteers. In this paper, 816 views are selected as the training set, and 96 are taken as the validation data, which is consistent with Ref. [18] and Ref. [20].

The TUPAC2016 dataset consists of two parts. The first part contains samples from 23 volunteers, and this part is the same as the AMIDA2013 dataset. The second part contains samples from 50 volunteers with dimension of 5657×5657 , which are scanned using Leica SCN 400 digital slice scanner. Each image is labeled by two pathologists independently. In the experiment, validation samples are selected by every 7 volunteers from volunteer no.30, and the others are taken as training samples. This strategy is consistent with Ref. [18].

4.2. Evaluation Metrics. Evaluation is performed according to the ICPR2014 contest criteria. A detected mitosis is counted as correct if its center point is localized within a range of $8 \mu\text{m}$ from its ground truth. Here, the center point of detected mitosis is defined as the diagonal intersection of its bounding box. Three metrics, namely, precision, recall,

and F_1 -score are employed as the quantitative indicators, which are defined in

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F_1\text{-score} &= \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}, \end{aligned} \quad (3)$$

where TP is the number of true mitosis which are detected, FP is the number of falsely detected mitosis, and FN is the number of true mitosis which are not detected. Precision indicates how many true mitosis are detected out of all the detected instances. Recall indicates how many true mitosis are detected out of all mitosis. F_1 -score gives a comprehensive combination of Precision and Recall.

4.3. Implementation Details. The experiments are carried out on a computer with Intel Core i7 CPU, Nvidia GTX 2080Ti GPU, and 16 GB RAM. All the codes are implemented using Python 3.6 as the programming language and PyTorch 1.9.0 as the deep learning framework. Both networks are trained 100 epochs with Adam optimizer and batch size of 8. During training, the learning rate is initialized as 10^{-3} and decreased to 10^{-6} after 100 epochs.

4.4. Results and Comparisons. Detection results and comparisons with some state-of-the-art methods considering the three criteria are listed in Table 1. It should be noticed that all the results of the referred methods are reported by the literatures. Some visual results are shown in Figure 5. In Figure 5, rectangles indicate the bounding boxes of detected mitosis, and dot marks indicate the ground truth label (enlarged for visual effect).

From the results, we can see that the proposed method achieves the best results compared with the referred methods in most of the cases. Li et al.'s method [18] is a CNN-based mitosis detection method which uses two concentric circles to label a mitosis area. The proposed method

TABLE 1: Detection results.

Datasets	Methods	Recall	Precision	F_1 -score
AMIDA2013	IDSIA [12]	0.612	0.610	0.611
	Li et al. [18]	0.677	0.669	0.673
	Proposed	0.689	0.690	0.689
ICPR2014	Yancey et al. [20]	—	—	0.507
	Li et al. [18]	0.682	0.541	0.603
TUPAC2016	Proposed	0.733	0.539	0.621
	Li et al. [18]	—	—	0.717
	Proposed	0.766	0.843	0.803

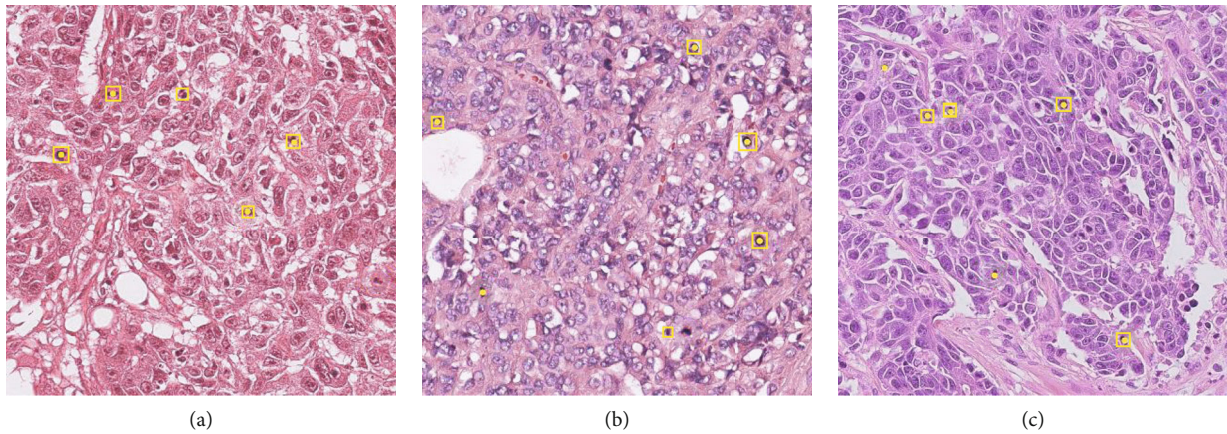


FIGURE 5: Some detection results. Rectangles indicate the bounding boxes of detected mitosis, and dot marks indicate the ground truth label (enlarged for visual effect): (a) AMIDA2013, (b) ICPR2014, and (c) TUPAC2016.

TABLE 2: Detection results.

Label generation method	Recall	Precision	F_1 -score
Proposed	0.766	0.843	0.803
U-Net	0.751	0.840	0.793
Manual	0.779	0.850	0.813

outperforms Li et al.’s method on all the three considered datasets with respect to most of the criteria. Since the proposed method employs a segmentation network to generate strong labels, the detection network can be better trained with this more adequate training data, and thus, more competent prediction performance is obtained. The IDSIA [12] method employs a multicolored max pooling convolutional neural network (MCMPCNN) for supervised pixel classification. However, only weak labels are used in its network training, which influences the prediction ability of IDSIA.

4.5. Impact of Label Generation. To further corroborate the effectiveness of the label generation strategy used in this paper, an additional experiment is performed on the TUPAC2016 dataset, and the results are listed in Table 2, in which Manual means the bounding boxes of mitosis are marked manually, i.e., the minimum rectangle that can surround the mitosis and include the ground truth point, U-Net

means the mitosis is segmented using the original U-Net network [25], and Proposed means the mitosis is segmented using the proposed segmentation network.

From the results, we can see that labels generated by the proposed segmentation network are slightly better than the original U-Net, which can be attributed to the fact that the proposed network can provide a more accurate pixel class prediction and thus provide a more accurate bounding box. However, different segmentation methods have similar final detection results, and they are also similar with that of manual labeling. We can conclude that segmentation-based label generation is beneficial to mitosis detection with nearly no accuracy loss.

5. Conclusions

In this paper, a deep learning-based method for mitosis detection in breast histopathology images is proposed. The method is aimed at solving the problem of the insufficiency of strongly labeled samples by incorporating a bounding box label generation process before mitosis detection. Experimental results show that the proposed label generation strategy can promote the mitosis detection performance in a large extent. The main limitation of the proposed method is that the employed object detection method R-CNN needs

a long time of training, although it can be solved by integrating a more powerful object detection method into the proposed detection routine.

Data Availability

The data used to support the findings of this study are all publicly available datasets deposited in the following websites: (1) AMIDA2013 dataset, <https://tupac.grand-challenge.org/Dataset/>. The AMIDA2013 dataset is contained as part of the TUPAC2016 dataset (also declared in the manuscript). These two datasets are published by the same organization, so the AMIDA2013 dataset cannot be accessed as its own. Anyone can get the AMIDA2013 dataset through downloading the TUPAC2016 dataset. Please refer to [29]. (2) ICPR2014 dataset, <https://mitos-atypia-14.grand-challenge.org/Dataset/>. (3) TUPAC2016 dataset, <https://tupac.grand-challenge.org/Dataset/>. (4) MITOS2012 dataset, <http://www.icpr2012.org/contests.html>

Conflicts of Interest

The author declares that they have no conflicts of interest.

References

- [1] C.-H. Huang and H.-K. Lee, "Automated mitosis detection based on exclusive independent component analysis," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR 2012)*, pp. 1856–1859, Tsukuba, Japan, 2012.
- [2] A. Tashk, M. S. Helfroush, H. Danyali, and M. Akbarzadeh, "An automatic mitosis detection method for breast cancer histopathology slide images based on objective and pixel-wise textural features classification," in *The 5th Conference on Information and Knowledge Technology*, pp. 406–410, Shiraz, Iran, 2013.
- [3] N. M. Rajpoot, A. M. Khan, and H. Eldaly, "A gamma-gaussian mixture model for detection of mitotic cells in breast cancer histopathology images," *Journal of Pathology Informatics*, vol. 4, no. 1, pp. 149–152, 2013.
- [4] F. B. Tek, "Mitosis detection using generic features and an ensemble of cascade adaboosts," *Journal of Pathology Informatics*, vol. 4, no. 1, p. 12, 2013.
- [5] Z. Cai and N. Vasconcelos, "Cascade R-CNN: delving into high quality object detection," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6154–6162, Salt Lake City, UT, USA, 2018.
- [6] H. Lei, H. Xie, W. Zou, X. Sun, K. Kpalma, and N. Komodakis, "Hierarchical saliency detection via probabilistic object boundaries," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no. 6, pp. 1755010.1–1755010.23, 2017.
- [7] D. Lin, G. Chen, D. Cohen-Or, P.-A. Heng, and H. Huang, "Cascaded feature network for semantic segmentation of rgb-d images," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 1320–1328, Venice, Italy, 2017.
- [8] X. Qian, F. Yanwei, Y.-G. Jiang, T. Xiang, and X. Xue, "Multi-scale deep learning architectures for person re-identification," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5409–5418, Venice, Italy, 2017.
- [9] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5689–5697, Venice, Italy, 2017.
- [10] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 411–418, Nagoya, Japan, 2013.
- [11] H. Wang, A. Cruz-Roa, A. Basavanthally et al., "Mitosis detection in breast cancer pathology images by combining hand-crafted and convolutional neural network features," *Journal of Medical Imaging*, vol. 1, no. 3, article 034003, 2014.
- [12] M. Veta, P. J. van Diest, S. M. Willems et al., "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Medical Image Analysis*, vol. 20, no. 1, pp. 237–248, 2015.
- [13] H. Chen, Q. Dou, X. Wang, J. Qin, and P.-A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 1160–1166, Phoenix Arizona, 2016.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [15] E. Zerhouni, D. Lányi, M. Viana, and M. Gabrani, "Wide residual networks for mitosis detection," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 924–928, Melbourne, VIC, Australia, 2017.
- [16] C. Li, X. Wang, W. Liu, and L. J. Latecki, "Deepmitosis: mitosis detection via deep detection, verification and segmentation networks," *Medical Image Analysis*, vol. 45, pp. 121–133, 2018.
- [17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [18] C. Li, X. Wang, W. Liu, L. J. Latecki, B. Wang, and J. Huang, "Weakly supervised mitosis detection in breast histopathology images using concentric loss," *Medical Image Analysis*, vol. 53, pp. 165–178, 2019.
- [19] D. Cai, X. Sun, N. Zhou, X. Han, and J. Yao, "Efficient mitosis detection in breast cancer histology images by RCNN," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 919–922, Venice, Italy, 2019.
- [20] R. Yancey, "Multi-stream faster RCNN for mitosis counting in breast cancer images," 2020, <https://arxiv.org/abs/2002.03781>.
- [21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages, pp. 580–587, Columbus, OH, USA, 2014.
- [22] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago, Chile, 2015.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788, Las Vegas, NV, USA, 2016.

- [25] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Munich, Germany, 2015.
- [26] P. Liu, H. Zhang, K. Zhang, L. Liang, and W. Zuo, “Multi-level wavelet-cnn for image restoration,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 886–88609, Salt Lake City, UT, USA, 2018.
- [27] L. Roux, D. Racoceanu, N. Loménie et al., “Mitosis detection in breast cancer histological images an icpr 2012 contest,” *Journal of Pathology Informatics*, vol. 4, no. 1, p. 8, 2013.
- [28] L. Roux, “Mitosis atypia 14 grand challenge,” <https://mitos-atypia-14.grand-challenge.org/Home/>.
- [29] M. Veta, Y. J. Heng, N. Stathonikos et al., “Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge,” *Medical Image Analysis*, vol. 54, pp. 111–121, 2019.

Research Article

A Pyramid Architecture-Based Deep Learning Framework for Breast Cancer Detection

Dong Sui ¹, Weifeng Liu,¹ Jing Chen,² Chunxiao Zhao,¹ Xiaoxuan Ma,¹ Maozu Guo,¹ and Zhaofeng Tian ²

¹School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

²Department of Laboratory and Diagnosis, Changshai Hospital, Navy Medical University, Shanghai 200433, China

Correspondence should be addressed to Dong Sui; suidong@bucea.edu.cn and Zhaofeng Tian; tian_zhao_feng@163.com

Received 30 June 2021; Accepted 20 August 2021; Published 1 October 2021

Academic Editor: Qiushi Zhao

Copyright © 2021 Dong Sui 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.

Breast cancer diagnosis is a critical step in clinical decision making, and this is achieved by making a pathological slide and gives a decision by the doctors, which is the method of final decision making for cancer diagnosis. Traditionally, the doctors usually check the pathological images by visual inspection under the microscope. Whole-slide images (WSIs) have supported the state-of-the-art diagnosis results and have been admitted as the gold standard clinically. However, this task is time-consuming and labour-intensive, and all of these limitations make low efficiency in decision making. Medical image processing protocols have been used for this task during the last decades and have obtained satisfactory results under some conditions; especially in the deep learning era, it has exhibited the advantages than those in the shallow learning period. In this paper, we proposed a novel breast cancer region mining framework based on deep pyramid architecture from multilevel and multiscale breast pathological WSIs. We incorporate the tissue- and cell-level information together and integrate these into a LSTM model for the final sequence modelling, which successfully keeps the WSIs' integration and is not mentioned by the prevalence frameworks. The experiment results demonstrated that our proposed framework greatly improved the detection accuracy than that only using tissue-level information.

1. Introduction

Breast cancer is the leading death cause among women all over the world [1]. Great progresses of microscopic imaging make digital pathology come into the whole-slide image (WSI) stage. These techniques allow a WSI image (a whole-slide image at 40x magnification is about 2 GB) to be stored, served, and viewed in multiscale, multiview, and multilevel than the light microscopy. In this context, modern precision medicine approaches require careful diagnostic with personality and precision survival prediction for each case so as to tailor suitable therapy protocol [2]. A straight diagnostic protocol of breast cancer is the interpretation of digital pathology slides, which used to be a time-consuming and labour-intensity pathway for manual interpretation with significant inter- and introobserver variability [3]. The computer-aided digital pathology analysis combines

the image processing technique that opens the door to automatically depicting the pathology slides with a more objective and quantitative way [4]. Over these decades, due to the breakthroughs in Artificial Intelligence (AI), it allows computers to reach the state of the art in many vision-based tasks, better than human counterparts for specific tasks, especially in the field of medical image processing [5].

Clinically, pathological analyses on axilla lymph nodes can indicate the original, spread, and metastasis of breast cancer; furthermore, the pathological changes in axilla lymph nodes are a critical factor for prognostic evaluation [6]. However, pathological image analysis in lymph nodes in tissue level is usually time-consuming and prone to subjective variances. In addition, small metastases such as isolated tumor cell clusters (ITCs) are a small region with few cells and difficult to detect or missed [7]. Based on these merits, there are great demands of automated breast cancer

detection frameworks for improving the robustness and precision of the decision making [4]. WSIs excited the development of quantitative histopathology analysis, which can now capture nuclear and tissue architecture from different levels and scales [8, 9]. Furthermore, cell-level analysis based on cell shape, nuclear morphology, and cytoplasm distribution is crucial for the tumor grading and reorganization [10]. It has been reported that many cell analysis frameworks give the possibility for incorporating with cell-level information to analyze the multiscale pathological image [5, 11].

In this paradigm, many deep learning-based metastasis detection methods have been proposed and achieved excellent results at the challenges in MICCAI and ISBI from 2016~2020 [12, 13]. A consistent evaluation of digital pathology analysis protocols for breast cancer diagnosis with accuracy and efficiency was performed; all of these successes are dependent on the proper designing and integrating exiting pipeline by using transfer learning. In this way, researchers can reduce the risk of cancer patient and slide misidentification; furthermore, tissue loss and damage can be better fixed to facilitate covering the gaps among pathology laboratories and clinical primary diagnosis [14]. Thus, the most important advantage of WSIs is that the researchers can apply deep learning-based methods in the diagnostic workflow. However, most of these frameworks mainly focus on tissue level in the WSIs and cannot depict the details of cell-level information.

2. Related Work

Detection of the suspected tumor region, characterization of tumor subtype, and quantification of tumor invasive extent are the critical procedure in breast cancer diagnostic. In deep learning context, CNN is employed for breast cancer WSI patch classification, in which the CNN model uses manually annotated labels for training and gets ideal results [15]. Cruz-Roa et al. proposed a deep learning pipeline that is pre-trained using image net for distinguishing the benign and malignant breast cancer; at the same time, some data augmentation methods have been adopted to prevent overfitting [16]. Lymph node metastasis is the most suitable background for AI algorithm application; Liu et al. and Steiner et al. proposed a challenge competition and establish a testbed for breast cancer diagnosis; the comparison results exhibited great superiority for the deep learning-based method than the pathologist achievement [17, 18]. On the other hand, tumor histologic grading and invasive tumor characterization can give a deep inspection for breast cancer prognostic evaluation. But cell-level and tissue-level feature identification used to be a laborious task, such as the tubular formation and mitotic cell analysis, which are important prognostic factors that were mined by using manual operation, and the labour-intensive nature of mitotic counting can lead to discordance.

Some shallow learning methods have achieved remarkable results on this paradigm [19, 20]; recently, deep learning methods have shown their excellent performance on this task [21–23]. For the early achievement, Rexhepaj et al. proposed a nuclear detection algorithm to quantify IHC staining cell for protein expression and get a correlation of 0.9 with manual counting [24]. Nonetheless, this method has

not achieved the state of the art at that time; the subsequent work shows that the deep learning-based method can help improve the level of concordance among human pathologists [24]. Romo et al.'s team employed a CNN model to detect tubule nuclei and use this information for Oncotype DX risk category [25]. Veta et al. propose a framework based on non-CNN model to perform cell nucleus detection and segmentation jointly for the cell morphological analysis [22, 25]. Biomarker finding is another element that related with diagnosis of breast cancer. More recently, WSI-based biomarker detection is becoming a prominent pathway for tumor evaluation directly using image information. Couture et al. introduced a deep learning-based multimodule framework for ER status prediction and get the accuracy up to 84% [26]. Shamaï et al. implemented a deep learning pipeline with less data of 19 biomarkers, and within the subgroup, they get the 92% accuracy of confidence score [27]. Image of aspiration biopsy was also employed to mining malignant and benign tumors by fitting cellular features with machine learning paradigm [28]. The conditional GAN model is a promising pathway of image data augmentation for the deep model training. Sahiner et al.'s works supported the application of GANs to boost the training phase to optimize tumor classification [29].

The morphological features extracted from the breast cancer WSIs are known to be valuable for the prognosis evaluation. Veta et al. use breast cancer microarray data and nuclear handcraft features to construct a general model for patients' prognosis evaluation, and their research open a window for the possible direction of tumor prognosis analysis by multimodule feature fusion [19, 20]. As a standard deep learning model, CNN can extract multiple-level features to represent the tumor region; Yuan proposed a CNN-based model to analyze the lymphocyte spatial distribution for classifying different tumors in WSIs [30]. In addition, the spatial relationship is also used for cell morphology analysis, tumor detection, and prognosis evaluation, but there are still few researches that focus on multilevel and multiscale information. In this paper, we proposed an open and multitask framework for tumor detection and grading; we also concern some spatial information and multilevel fusion features to depict the hidden relationship between tumor statues and information from each level.

3. Materials and Methods

3.1. Datasets. For the framework construction, we use Camelyon 2017 data set for the tissue-level network training and TIM 2015 data set for cell-level detection network training, respectively [7, 31]. Although there are many other data sets for the cancer region detection of breast cancer, we still choose the Camelyon 2017 data set for the evaluation; this is because of the restriction of data scale and the processing ability of our hardware platform. Figure 1 shows the details of the two data samples. For the framework construction, we use the Camelyon 2017 data set for the tissue-level network training and the TIM 2015 data set for cell-level detection network training, respectively. The TIM2015 data set can be found at <http://haeckel.case.edu/data/TMI2015.tgz>,

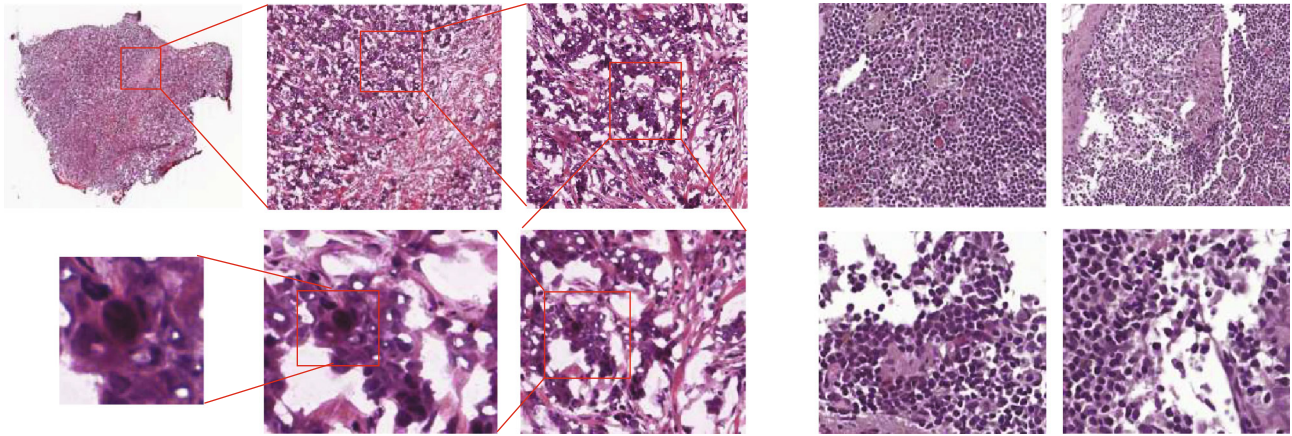


FIGURE 1: Data set in this paper. (a) Is the Camelyon 2017 whole-slide images [7]; (b) is the TMI 2015 data set from Xu et al. [31].

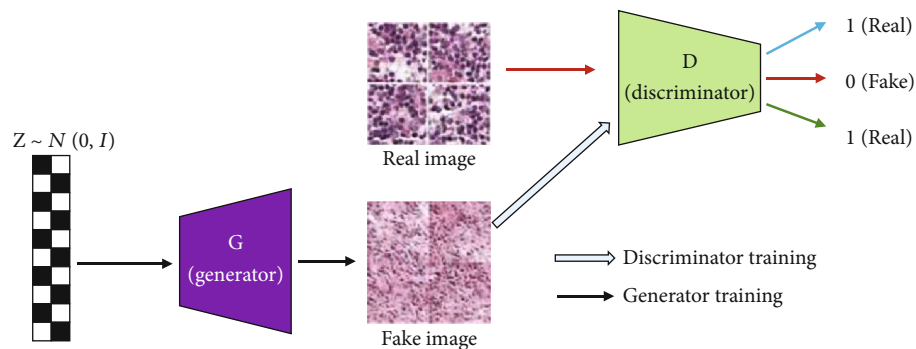


FIGURE 2: GAN model for pathological image generation.

and the Camelyon 2017 data set can be found at <https://camelyon17.grand-challenge.org/>.

The Camelyon 2017 data set is a multilevel WSI data set including patients from different medical centers and up to 500 slides; for the network training, we manually selected 100 slides for this task. The glass slides were digitized by whole-slide scanners with a pixel size of 0.23 to 0.26 μm , respectively. The WSI is a multiple-resolution and multiple-level images that is about 1×10^5 by 2×10^5 pixels at the highest resolution level. The whole WSI contains 10 resolution levels; each consecutive level doubled the pixel size in both directions and halved the pixel in each dimension; in addition, the file size of a WSI with 10 levels is about 2~4GB which varies depending on the scanner and tissue anatomy structure of the input image. The scanned images were converted into standard multiresolution TIFF image files according to open slide standard [7].

The TMI 2015 data set is also H&E-stained histopathological images, which were obtained from digitized glass slides corresponding to 49 lymph node-negative and estrogen receptor-positive breast cancer (LN-, ER+ BC) patients at Case Western Reserve University. The size of each image is about 2200×2200 pixels, and there are about 1500 nuclei in each image [31].

3.2. Pathological Region Generation. For the training of the pyramid deconvolution network, balance of the training sample is a critical problem, but in this data set, the tumor

types are unbalance in different tumor stages. For this situation, we adopted the deep convolution generative adversarial networks (DCGAN) for boosting the tumor region patches and extended the origin patch set into an expanded edition. Under this situation, the data unbalance can be addressed as a data generation problem from the existing classes, so as to promote the balance of the training data. Figure 2 shows the details of the pathological image generation process for the microregion of breast cancer.

3.3. Whole-Slide Image Preprocessing. In this paper, we employed two data sets, part of Camelyon 2017 and the TMI 2015 from breast cancer tissue section slide. For WSIs of Camelyon 2017, we extracted the slide area on each level by OTSU algorithm. Following this step, we split the original image into small patches with the same size at each scale. At the same time, we transform the counter label of Camelyon 2017 into mask and perform the same manipulation with its WSI image at each level. Concerning the image quality, we chose a commonly used color equalization to reduce the effect of unequalization and uneven illumination of original WSI staining [32].

Because the cancer regions in the Camelyon 2017 data set only cover a small region in the WSI slides, this led to an unbalance situation for the deep model training. We adopt some data augmentation methods to merge the gap, such as random cropping, colour jittering, scaling, and rotation. For some class, we use deep convolution generative

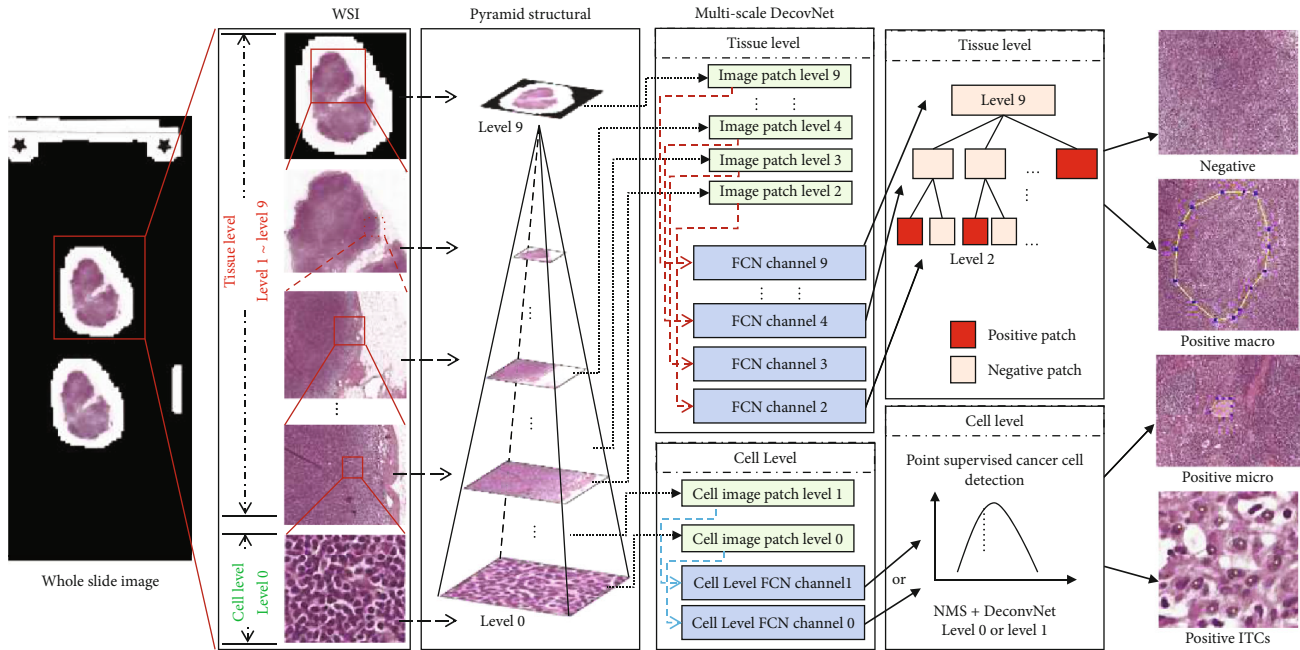


FIGURE 3: The scheme of pyramid deconvolution network framework for breast cancer detection.

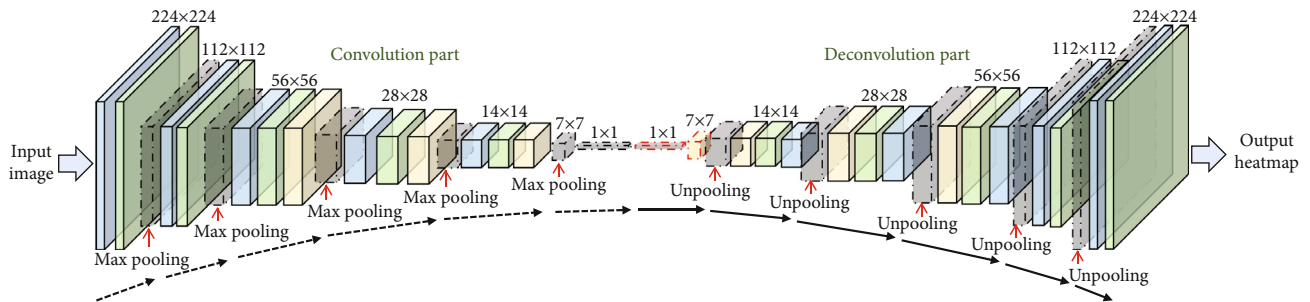


FIGURE 4: The backbone network architecture of the DeconvNet. We make a slight change on kernel size and stride for different input image scales.

adversarial networks (DCGAN) for boosting the tumor region patches; finally, we extended the origin patch set into an expanded edition.

For the TIM2015 data set [31], this data set is used for training and evaluating a cell detection model; the original data was labeled by using a binding box; for convenience and application, we transform the binding box into a point label by making an average position through the four coordinates of the binding box. Then, following our proposed method, we fill the dot label with Gaussian kernel for the cell mask generation, and at last, we constructed the final training set for cell-level analysis pipeline.

3.4. Framework Architecture and Network Design. In this paper, we construct a multilevel and multiscale tumor region detection and segmentation framework for breast cancer. As it is shown in Figure 3, the WSIs are stored into a multiple-level pyramid structure with 10 levels; the user can zoom into any level and depict the details to perform diagnosing tasks. For this situation, we divided the total framework into

two levels of tumor analysis procedure pipeline, in the tissue level and cell level, we adopt the same network backbone architecture derived from the DeconvNet (see Figure 4), and some parameter settings are also changed according to the input image. In the following part, we will introduce the details of our framework in the tissue level and cell level, respectively.

The aim of the tissue level is to get the tumor region according to the labeled mask so as to assign the TNM stage of WSI. As shown in Figure 4, the tissue level and cell level share the same basic network. They are all based on DeconvNet which consists of the stacked convolution layers, max-pooling layers, and deconvolution layers. In the front three blocks, max-pooling layers are followed with convolution layers, and ReLU activation layer and convolution layer which form deconvolution layers are followed with upsampling layers which are followed by the last three blocks. Finally, a convolution layer was used to replace fully convolutional layers in the end of the network in order to obtain a density map. Totally speaking, 3×3 kernel size, ReLU

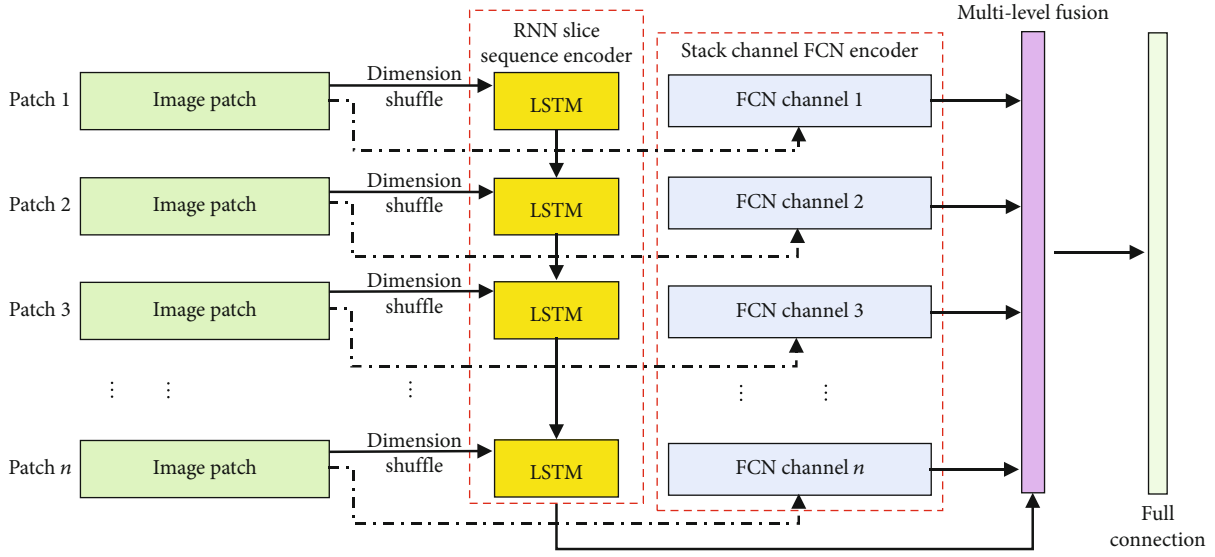


FIGURE 5: Long short-term memory enhanced channeled fully convolutional network pipeline with a dimension shuffle layer.

activation, Mean Squared Error (MSE) loss function, and Adam optimizer were used in all network structures. In order to handle different tasks, the basic network was fine-tuned. As shown in Figure 2, the fully connected layer was connected to the last convolution layer to get confidence result of patch binary classification. With characteristic of DeconvNet, images of any size can be input. However, in actual application, images with the dimensions reduction were used to avoid cutting the edge of the output density map.

For the high-resolution WSI images, the content is usually up to 2 GB for a single image, and the tumor region locates at somewhere in the whole slide. For a large image, the network has to be trained on the image patches generated from the ground scale. In order to keep the patch sequencing, we proposed a LSTM-CFCN-based model for the segmentation task. In this model, the stacked channel FCN is for the patch feature encoding and the LSTM model is used to merge these patches into a large image from these FCN encoded patches; more details are illustrated in Figure 5. In this part, the cancer region in the training data set is labelled by using a mask; the channel FCN is used to estimate the density map of a tumor region in each sequential patch; and LSTM block is used to combine the detected result into an integrated figure. From this point of view, the tumor density map is predicted by the deconvolution part, and a Euclidean distance is used for measuring the difference between the generated density map and the ground truth. And the loss function is defined as follows:

$$L_t = \frac{1}{2N} \sum_{i=1}^N \sum_{p=1}^p \|F_i(p; \Theta_{\text{CFCN}}) - F_i^0(p)\|_2^2, \quad (1)$$

where N is the batch size and $F_i(p)$ is the probability of tumor at pixel p in the i th sequential patch. For some situation in low level of WSI, the global tumor region is generated from several sequential image patches; the integration is

learned by the LSTM block. So, the total loss function includes the basic tumor segmentation part and the LSTM residual part. The final result is a sum of the two parts:

$$T_i = R(F_i; \Upsilon, \Psi) + \sum_{p=1}^p F_i(p), \quad (2)$$

where $R(F_i; \Upsilon, \Psi)$ is the residual count, F_i is the estimated heat map from patch i , Υ is the parameter of LSTM, and Ψ is the fully connected layer's parameter. The whole-slide tumor detection loss function is defined as

$$L_{\text{lstm}} = \frac{1}{2N} \sum_{i=1}^N (T_i - T_i^0)^2, \quad (3)$$

where T_i^0 is the ground truth in the i th image patch and T_i is the learned tumor region. In this way, the total loss function for the multiple-scale LSTM-CFCN is defined as

$$L = L_t + \alpha L_{\text{lstm}}, \quad (4)$$

where alpha is a weight parameter of the LSTM residue and to be tuned for the suitable accuracy. At the same time, the tumor detection is trained with fewer parameters to achieve a better training process. In the framework, the Adam optimizer and backpropagation are used to optimize the loss function L for different scenes.

3.5. Tissue-Level Pyramid-DeconvNet. In the tissue level, conventional deep learning methods usually take amount of time and space to handle all the small patches. In order to overcome this obstacle, we introduce a flexible automatic decision method based on the pyramid deconvolution networks, which can target the RoIs (Region of Interest) quickly with higher accuracy. We first cast the problem as a supervised learning problem that tries to learn a mapping between a patch $I_l(x)$ and a density map $D_l(x)$, denoted

```

Input: WSI image  $I$ , patch level DeconvNet for scale  $l$ ,  $C_p$ , layer number  $L$ , and confidence threshold  $t$ .
Output: Selected patches  $Ps$ .
1: Generate patches  $Ps_{L-1}$  with step  $w$  and  $h$  in  $I_{L-1}$ , and location code  $LCIS_{L-1}$ .
2: patches initialization with  $Ps = Ps_{L-1}$ .
3: for  $i = L-2$  to 2 do
4:   if  $Ps$  is empty then
5:     Break
6:   for patch  $p$  in  $Ps$  do
7:     Calculate cancer confidence of  $p$  named  $c$  with  $C_i$ 
8:     if  $c > t$  then
9:       Add  $LCI_p$  to  $LCIS_i$ 
10:    for  $LCI$  in  $LCIS_i$  do
11:      Calculate  $LCI$  in  $i-1$  layer
12:    Generate patches with all  $LCI_s$  named  $Ps_{i-1}$ 
13:  Set current patch set  $Ps = Ps_{i-1}$ 

```

ALGORITHM 1: Tissue-level pathological RoI extraction.

as $F_l : I_l(x) \rightarrow D_l(x)$ which, $(I_l \in R^{m \times n}, D_l \in R^{m \times n})$ in layer l and obtained different weights w_l and biases b_l . Then, we fixed w_l and b_l to train the last fully connected layer for classification $C_l(w_l, b_l, w_l^c, b_l^c)$. Utilizing a tree-like searching protocol, networks in layer l will test their layer separately with C_l and obtained the classification confidence of cancer $C(x)$. RoI prediction probability was compared with threshold t . If we find RoI, we introduce location code information, short for LCI, which is a series of continuous coding to represent the position of each RoI and find the coordinates and LCIs of next layer $l-1$. Loop until we reach the top layer of cell level. Details are shown in Algorithm 1.

3.6. Cell-Level DeconvNet. In the small tumor region of the WSI images, it even contains few illness cells. The cell-encoding information is obtained for better judgment of small tumor areas, such as micro and ITC. For the tumor region affirmative, we introduce a cell-level DeconvNet. The basic network is a DeconvNet with a NMS layer embedded to the last convolution layer. After training, the proposed network with patches and their mask images were filled with Gaussian kernels with Equation ((1)):

$$P(x) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}. \quad (5)$$

Density maps which consist of 0-1 float values of each layer were obtained. After that, NMS was employed to get the accuracy of each cancer cell. And then, we view the value of detection count results as cancer cell counting results and cluster the detected points to form the region of cancer area. As shown in Figure 3, a Nonmaximum Suppression (NMS) algorithm was connected to the final layer to get the position of each cell and obtain the result of cancer cell counting and region area. In this way, the small region such as ITC and micro are detected by using cell-level screening.

3.7. Cancer Region Generation. Through tissue-level detection method, if Algorithm 1 breaks with layer number larger than two, the WSI is negative. And macro areas and large

size of microareas were detected if the total area of connected RoI patches in layer two were larger than the threshold of the area in the TNM staging system. And then, cell level determined the areas of micro or ITC, which were passed to a cancer cell detector to obtain the location and number of cancer cells. The size of cancer areas and the number of cancer cells were breast cancer detection standard to classify the images according to the TNM system.

3.8. Evaluation Methods. For the tumor segmentation, we evaluated the segmentation performance by using the tradition method of image segmentation. The results of our pipeline are reported in terms of recall, precision, and F_1 -measure value, as follows:

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN}, \\ \text{Precision} &= \frac{TP}{TP + FP}, \\ F_1 &= \frac{2PR}{P + R}. \end{aligned} \quad (6)$$

For the evaluation, we choose the intersection over union (IoU) metric and Jaccard index to quantify the percent overlap between the target mask and the framework prediction output results. The Jaccard similarity coefficient of the segmentation result and the original labelled mask is expressed as

$$\text{Jaccard}(S, M) = \frac{|\text{intersection}(S, M)|}{|\text{union}(S, M)|}, \quad (7)$$

where $|*|$ represents the cardinal of set $*$. The Jaccard index can also be expressed in terms of True Positives (TP), False

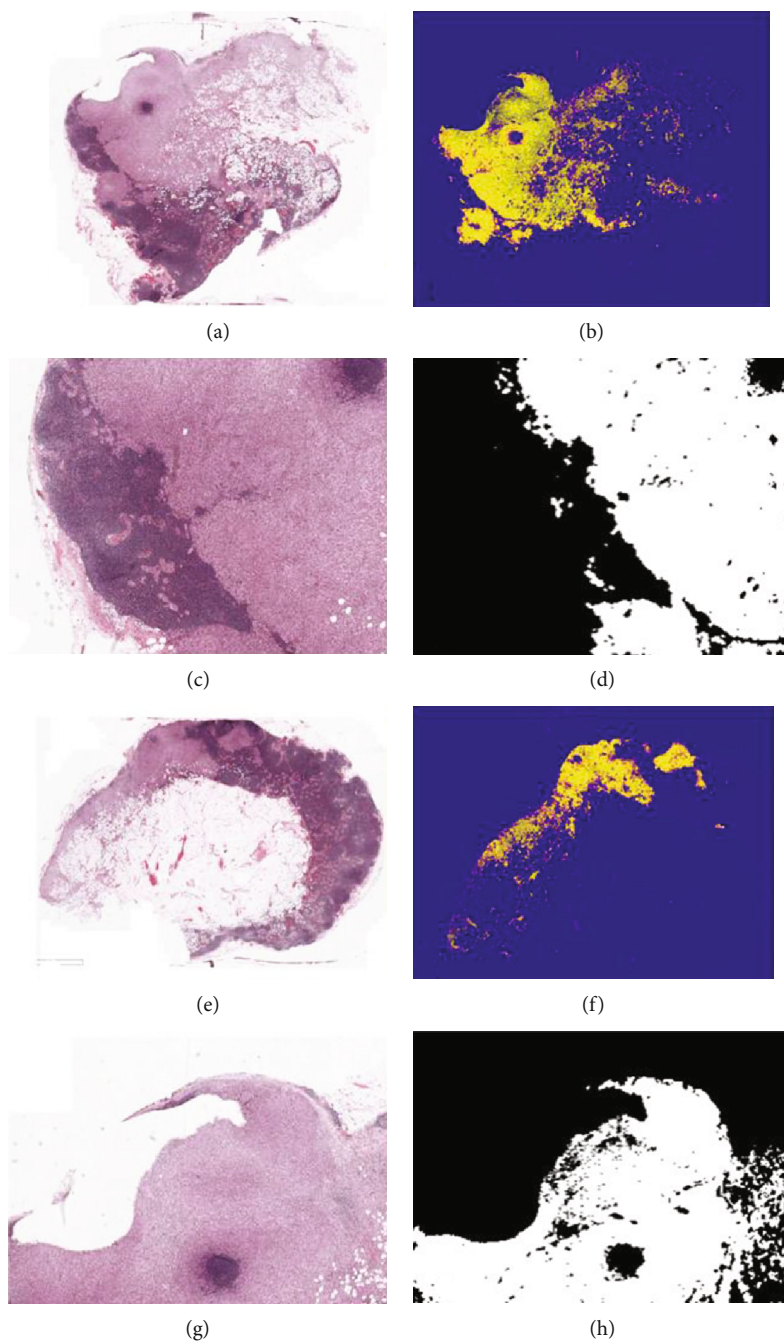


FIGURE 6: Pathological images and their detection results. The subimages (a), (c), (e), and (g) are the original pathological images; (b) and (f) are the heat maps generated from our framework.

TABLE 1: Comparison results of different methods.

Methods/accuracy	Normal (%)	Macro (%)	Micro (%)	ITCs (%)	Average (tumor, %)
FCN	64.3	61.1	53.6	17.3	44
U-Net	77.8	67.1	55.2	21.2	47.83
Ours	80.1	73.6	57.7	20.6	50.63
FCN + cell detection	68.7	64.7	54.2	14.1	44.33
U – Net + cell detection	79.6	70.1	58.4	23.1	50.53
Ours + cell detection	85.3	74.3	60.1	24.6	53

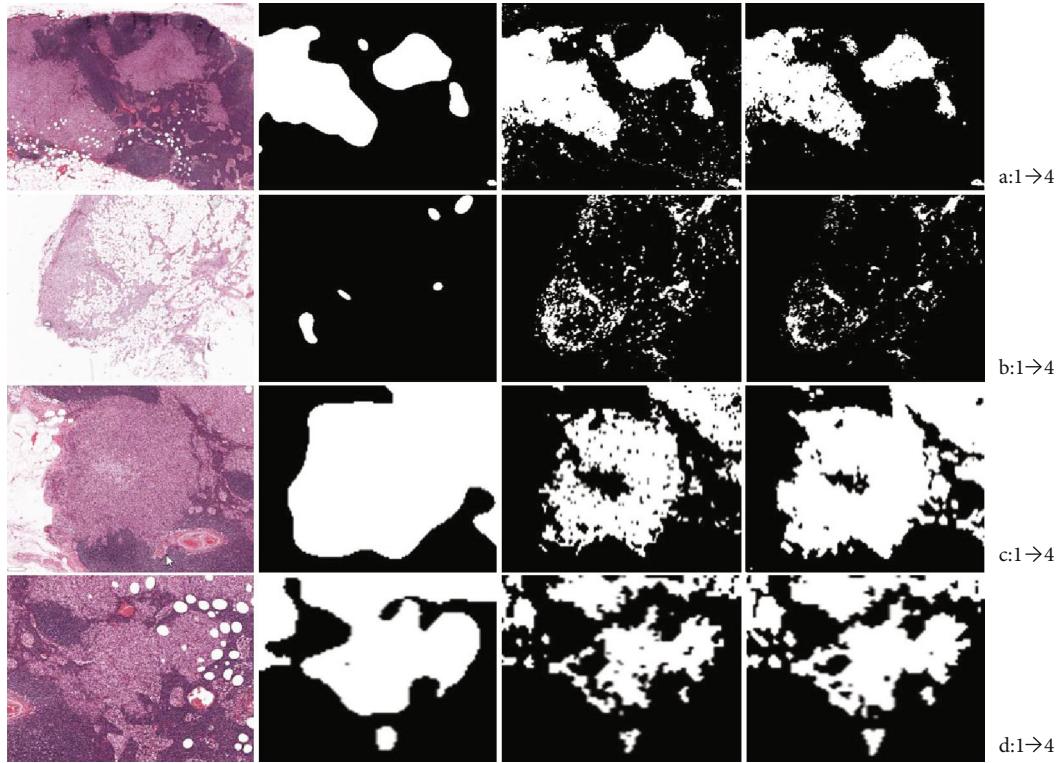


FIGURE 7: Comparisons of the segmentation results among different methods. (a–d) Are different views of the WSIs; from 1 to 4 are the segmentation results of original image, FCN, U-net, and our proposed method.

Positives (FP), and False Negatives (FN) as

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \quad (8)$$

4. Results and Discussion

4.1. Network Training. All experiments and test bed are carried out on Intel Core (TM) i9, 12 cores, 2.90 GHz processor with 32 GB of RAM and a NVIDIA GeForce GTX 2080Ti Graphics Processor Unit. The software implementation is performed using CUDA10.0, CuDNN8.5, and Pytorch1.5.

The proposed models have been tested on the Camelyon 2017 data sets; in this part, we divided each data set into two parts, 70% and 30% for training and testing. We kept the same parameters in all FCN blocks to make sure the network stability. For making sure the cells in LSTM model, we performed several tests about the cell number; finally, we choose the N with 10 cells. The training epoch generally kept at 3000 for the pathological images. In the training processes, random sampling and truncated back propagation are employed for handling the huge image data and LSTM model tuning. Adam optimizer is used in these models with $1e-3$ and final learning rate with $1e-5$, and the learning rate is reduced by an equal step in each epoch until the final learning rate is reached.

4.2. Results of Tumor Region Detection. In this paper, we proposed a multiple-level CFCN-based framework for the tumor region detection in breast cancer WSI image data

set. The final aim is to combine the pyramid information from different image scales to ensure the accuracy of final segmentation results. In this part, we choose several tissue-level and cell-level information for the segmentation task and get a relative satisfactory decision on the test set.

The segmentation results of our proposed framework are shown in Figure 6. For the tissue-level detection, we applied the framework to carry out the tumor region detection task on the Camelyon 2017 data set. Here, we selected some segmentation result as the final tissue-level detection; in Figure 6, the heat maps (b) and (f) exhibit the tumor region in the WSI slides. For the segmentation details, (d) and (h) are the partial details of the segmentation result on the WSIs; it can depict that our proposed framework can distinguish the tumor region and normal region clearly in the testing set.

For the comparisons, we choose some state-of-the-art methods such as FCN and U-Net as the test bed for the final method evaluation. In tissue-level segmentation, we compared our proposed method with the abovementioned methods; results are shown in Table 1; for the normal region in WSIs, our proposed framework gets the highest accuracy compared with the traditional methods. In addition, U-Net usually takes the priority position in medical image processing especially on segmentation task, but the U-net model cannot depict the level-wise information. In our framework, we use both cell-level and tissue-level information for the final decision and get a higher accurate result. Micro and ITCs are small regions that contain few cancer cells with unstable variability in morphology, usually existing at an unstable statute in the whole slide. All of these induced the

low accuracy of detection and segmentation. In this situation, our framework still catches the better results for some fixed situation.

To refine the segmentation results, we combined the cell-level information into the framework and make sure the segmented tumor region contains multilevel information. Especially for the micro and ITC regions, the cell-level information can indicate the existence of small tumor regions even without screening by tissue-level scans. For cell-level information incorporation, we can see that it can improve the detection result, Table 1 shows the detection results by employing cell-level and tissue-level information, and there is a certain improvement on detection accuracy. Figure 7 shows the segmentation result by using different methods; it can be addressed; our proposed method can greatly improve the segmentation result compared with FCN and U-Net.

5. Conclusions

In this paper, we propose an automatic cancer lesion detection approach using pyramid deconvolution network (PDN) for multilevel and multiscale H&E-stained breast pathological WSIs. In this framework, we integrate tissue- and cell-level information for the cancerous region detection and segmentation, which is neglected by state-of-the-art methods. The results demonstrated that our workflow greatly improved the performance compared with those only using tissue-level information. The comparison results showed our framework can get better accuracy on the same testing data set. In the future, our aim will focus on multiscale feature extraction and fine-tuning the new representation network for improving the detection and segmentation performance.

Data Availability

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

Disclosure

This manuscript is an extension edition of the conference paper published on 2020 the 4th International Conference on Big Data Research (ICBDR'20), titled "Pyramid Deconvolution Net: Breast Cancer Detection Using Tissue and Cell Encoding Information," but the main framework is changed greatly different from the early version.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Dong Sui and Jing Chen have equal contribution to this manuscript.

Acknowledgments

This research is supported by National Natural Science Foundation of China (Grant No.61702026, 62031003), the Pyramid Talent Training Project of Beijing University of Civil Engineering and Architecture (Grant No. JDYC20200318), the National Key Research and Development Program of China (Grant No. 2020YFF0305504), and the Doctoral Research Initiation Fund of Beijing University of Civil Engineering and Architecture (No. X20040).

References

- [1] C. Li, X. Wang, W. Liu, and L. J. Latecki, "DeepMitosis: mitosis detection via deep detection, verification and segmentation networks," *Medical Image Analysis*, vol. 45, pp. 121–133, 2018.
- [2] D. S. Gomes, S. S. Porto, D. Balabram, and H. Gobbi, "Inter-observer variability between general pathologists and a specialist in breast pathology in the diagnosis of lobular neoplasia, columnar cell lesions, atypical ductal hyperplasia and ductal carcinoma in situ of the breast," *Diagnostic Pathology*, vol. 9, no. 1, pp. 121–121, 2014.
- [3] K. H. Allison, L. M. Reisch, P. A. Carney et al., "Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel," *Histopathology*, vol. 65, no. 2, pp. 240–251, 2014.
- [4] A. Ibrahim, P. Gamble, R. Jaroensri et al., "Artificial intelligence in digital breast pathology: techniques and applications," *Breast*, vol. 49, pp. 267–273, 2020.
- [5] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [6] E. A. Rakha, M. Aleskandarani, M. S. Toss et al., "Breast cancer histologic grading using digital microscopy: concordance and outcome association," *Journal of Clinical Pathology*, vol. 71, no. 8, pp. 680–686, 2018.
- [7] P. Bandi, O. Geessink, Q. Manson et al., "From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 550–560, 2019.
- [8] X. Hou, J. Liu, B. Xu et al., "Dual adaptive pyramid network for cross-stain histopathology image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019*, D. Shen, Ed., pp. 101–109, Springer International Publishing, 2019.
- [9] H. Jia, Y. Song, H. Huang, W. Cai, and Y. Xia, "HD-Net: hybrid discriminative network for prostate segmentation in MR images," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019*, D. Shen, Ed., pp. 110–118, Springer International Publishing, Cham, 2019.
- [10] X. Shi, F. Xing, Y. Xie, H. Su, and L. Yang, "Cell encoding for histopathology image classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 30–38, Springer, 2017.
- [11] S. Zhang and D. Metaxas, "Large-scale medical image analytics: recent methodologies, applications and future directions," *Medical Image Analysis*, vol. 33, pp. 98–101, 2016.
- [12] Y. Zhang, H. Chen, Y. Wei et al., "From whole slide imaging to microscopy: deep microscopy adaptation network for histopathology cancer image classification," in *Medical Image*

- Computing and Computer Assisted Intervention–MICCAI 2019*, D. Shen, Ed., pp. 360–368, Springer International Publishing, Cham, 2019.
- [13] H. Chen, X. Han, X. Fan et al., “Rectified cross-entropy and upper transition loss for weakly supervised whole slide image classifier,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019*, D. Shen, Ed., pp. 351–359, Springer International Publishing, Cham, 2019.
- [14] B. J. Williams, D. Bottoms, and D. Treanor, “Future-proofing pathology: the case for clinical adoption of digital pathology,” *Journal of Clinical Pathology*, vol. 70, no. 12, pp. 1010–1018, 2017.
- [15] A. Cruz-Roa, H. Gilmore, A. Basavanthally et al., “Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent,” *Scientific Reports*, vol. 7, no. 1, article 46450, 2017.
- [16] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A dataset for breast cancer histopathological image classification,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2016.
- [17] Y. Liu, T. Kohlberger, M. Norouzi et al., “Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists,” *Archives of Pathology & Laboratory Medicine*, vol. 143, no. 7, pp. 859–868, 2019.
- [18] D. F. Steiner, R. MacDonald, Y. Liu et al., “Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer,” *The American Journal of Surgical Pathology*, vol. 42, no. 12, pp. 1636–1646, 2018.
- [19] M. Veta, P. J. van Diest, S. M. Willems et al., “Assessment of algorithms for mitosis detection in breast cancer histopathology images,” *Medical Image Analysis*, vol. 20, no. 1, pp. 237–248, 2015.
- [20] M. Veta, Y. J. Heng, N. Stathonikos et al., “Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge,” *Medical Image Analysis*, vol. 54, pp. 111–121, 2019.
- [21] M. P. Humphries, S. Hynes, V. Bingham et al., “Automated tumour recognition and digital pathology scoring unravels new role for PD-L1 in predicting good outcome in ER-/HER2+ breast cancer,” *Journal of Oncology*, vol. 2018, Article ID 2937012, 14 pages, 2018.
- [22] M. Veta, R. Kornegoor, A. Huisman et al., “Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer,” *Modern Pathology*, vol. 25, no. 12, pp. 1559–1565, 2012.
- [23] D. Tellez, M. Balkenhol, I. Otte-Holler et al., “Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2126–2136, 2018.
- [24] E. Rexhepaj, D. J. Brennan, P. Holloway et al., “Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry: application to measurement of oestrogen and progesterone receptor levels in breast cancer,” *Breast Cancer Research*, vol. 10, no. 5, p. R89, 2008.
- [25] D. Romo-Bucheli, A. Janowczyk, H. Gilmore, E. Romero, and A. Madabhushi, “Automated tubule nuclei quantification and correlation with Oncotype DX risk categories in ER+ breast cancer whole slide images,” *Scientific Reports*, vol. 6, no. 1, article 32706, 2016.
- [26] H. D. Couture, L. A. Williams, J. Geradts et al., “Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype,” *NPJ Breast Cancer*, vol. 4, no. 1, p. 30, 2018.
- [27] G. Shamaï, Y. Binenbaum, R. Slossberg, I. Duek, Z. Gil, and R. Kimmel, “Artificial intelligence algorithms to assess hormonal status from tissue microarrays in patients with breast cancer,” *JAMA Network Open*, vol. 2, no. 7, article e197700, 2019.
- [28] V. Madekivi, P. Bostrom, A. Karlsson, R. Aaltonen, and E. Salminen, “Can a machine-learning model improve the prediction of nodal stage after a positive sentinel lymph node biopsy in breast cancer?,” *Acta Oncologica*, vol. 59, no. 6, pp. 689–695, 2020.
- [29] C. Senaras, B. Sahiner, G. Tozbikian, G. Lozanski, and M. N. Gurcan, “Creating synthetic digital slides using conditional generative adversarial networks: application to Ki67 staining,” in *Medical Imaging 2018: Digital Pathology*, Houston, Texas, United States, 2018.
- [30] Y. Yuan, “Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer,” *Journal of The Royal Society Interface*, vol. 12, no. 103, article 20141153, 2015.
- [31] J. Xu, L. Xiang, Q. Liu et al., “Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 119–130, 2016.
- [32] K. S. Song, H. Kang, and M. G. Kang, “Hue-preserving and saturation-improved color histogram equalization algorithm,” *Journal of the Optical Society of America A*, vol. 33, no. 6, pp. 1076–1088, 2016.

Review Article

Review of Breast Cancer Pathological Image Processing

Ya-nan Zhang ^{1,2}, Ke-ruì XIA ², Chang-yi LI,¹ Ben-li WEI,¹ and Bing Zhang¹

¹*School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China*

²*HRG International Institute (Hefei) of Research and Innovation, Hefei 230000, China*

Correspondence should be addressed to Ke-ruì XIA; xiakr@hitrobotgroup.com

Received 12 July 2021; Accepted 24 August 2021; Published 21 September 2021

Academic Editor: Lei Zhang

Copyright © 2021 Ya-nan 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.

Breast cancer is one of the most common malignancies. Pathological image processing of breast has become an important means for early diagnosis of breast cancer. Using medical image processing to assist doctors to detect potential breast cancer as early as possible has always been a hot topic in the field of medical image diagnosis. In this paper, a breast cancer recognition method based on image processing is systematically expounded from four aspects: breast cancer detection, image segmentation, image registration, and image fusion. The achievements and application scope of supervised learning, unsupervised learning, deep learning, CNN, and so on in breast cancer examination are expounded. The prospect of unsupervised learning and transfer learning for breast cancer diagnosis is prospected. Finally, the privacy protection of breast cancer patients is put forward.

1. Introduction

Breast cancer is one of the most common malignant tumors. According to Chinese Women's survey, breast cancer is the most common malignant tumor in Chinese women, and the incidence rate is increasing year by year. The key to reduce the mortality of breast cancer is early diagnosis and treatment. At present, mammography is the most commonly used method to detect breast cancer. However, because of the huge amount of data and the poor imaging features of early breast cancer, early diagnosis is very difficult. With the development of image processing technology and early diagnosis technology, image processing of breast pathology has become an important way of early diagnosis of breast cancer, which mainly includes the study of masses, calcifications, and breast density. One of the main manifestations of breast cancer in breast mammography is mass. The basic steps of pathological image processing are as follows: first is image preprocessing, which includes removing background, marker, pectoral muscle and noise, and breast segmentation and image enhancement. Secondly, the region of interest is found by a basic image processing method. Then, features that can represent the quality, such as texture features and morphological features, are extracted. Finally, the tumor and normal tissue were separated according to

the extracted features. Another manifestation of breast cancer on X-ray images is a large breast density [1].

The object of breast pathology image processing is a variety of medical images with different imaging mechanisms. The types of medical imaging widely used in clinic include X-ray imaging (X-CT), magnetic resonance imaging (MRI), nuclear medicine imaging (NMI), and ultrasonic imaging (UI). X-ray imaging (X-CT), mainly through X-ray tomography, such as head tomography, is used to diagnose cerebral vascular diseases and intracranial hemorrhage. X-ray tomography has a good effect in the diagnosis of traumatic skull and facial fractures. Magnetic resonance imaging (MRI) is a noninvasive imaging technique, which can produce three-dimensional anatomical images [2]. Nuclear medicine imaging (NMI) is based on the difference of radioactivity concentration inside and outside organs or between different parts of organs [3, 4]. UI was used to observe the shape, location, size, number and scope of the mass, and the activity of abdominal organs. The edge echo, capsule, smoothness, wall thickness, and halo were observed. According to the clinical data, the symptoms of breast cancer can be adjusted by means of X-ray examination, which is identified by X-ray mammography. MRI results of tumor detection are higher than the actual value, while CT results of tumor detection are lower than the actual value. Dynamic

enhanced CT examination of breast lesions can better overcome the situation of missed detection.

Medical image acquisition and interpretation are the basis of breast cancer diagnosis based on medical imaging. In recent years, the speed and resolution of image acquisition have been greatly improved. However, image diagnosis is limited by the doctor's experience, ability, and other subjective factors, and its ability of replication and promotion is limited. In order to minimize the dependence on doctors, image processing technology is applied to medical imaging processing. Medical image processing includes lesion detection, image segmentation, image registration, and image fusion. In addition to clinical diagnosis, medical image processing also plays an important auxiliary role in medical teaching, operation planning, operation simulation, and various medical researches [5, 6].

2. Detection of Breast Cancer

Detection of breast cancer is mainly based on detection methods, image processing of lesion detection, lesion location matching, and extraction of lesion feature values. Breast cancer detection can detect candidate lesion location by supervised learning or classical image processing. The combination of image processing and Convolutional Neural Networks (CNN) is one of the successful examples of deep learning in recent years. The CNN is applied in breast cancer image analysis, mapping the input layer, pool layer, modified linear unit, all link layer, and output layer, respectively, and predicting the information represented by medical images. For example, Setio et al. extracted the features of pulmonary nodules in nine different directions of 3D chest CT scanning, selected the appropriate candidate as the center, and classified the candidates through CNN [3, 7–11]. Ross et al. decomposed the 3D image into 2D patches and then rotated the 2D patches randomly to get the "2.5D" view. The CNN was used to detect the early features of cancer from the 2.5D view. The combination of deep learning and image processing greatly improves the accuracy of lesion detection [12], while it is difficult to achieve high accuracy by using nondeep learning classifiers such as support vector machine. The accuracy of CNN algorithm depends on the training of initial markers by experts and needs a wide range of case coverage. Therefore, the promotion of CNN in the field of medical image processing is constrained by resources "transfer learning" that can reduce the dependence of CNN on initial marker training to a certain extent [13, 14], but the application of transfer learning itself is limited, so it is difficult to find the application conditions of transfer learning between medical images of human organs.

Early diagnosis of breast cancer can be detected by tumor markers. Tumor markers are substances produced and secreted by tumor cells during growth and reproduction. When these substances reach a certain amount, they can be extracted from breast images. The early feature values of breast cancer can be identified by using SIFT (scale invariant feature transform) or HOG (Histogram of Oriented Gradient) and so on to provide support for the early diagnosis of

breast cancer. The combination of image processing technology and reinforcement learning technology can reduce the dependence on a human doctor's experience to the greatest extent. Image processing technology is used to process two-dimensional slices, and then reinforcement learning is combined to set the enhancement target. Through the judgment of each discrete two-dimensional slice, the optimal decision strategy is found, so as to maximize the benefit of judging the pathological correctness of the whole group of two-dimensional slices. Through the analysis and processing of two-dimensional slice image, the segmentation, extraction, three-dimensional reconstruction, and three-dimensional display of human breast, surrounding soft tissue and lesion are realized. After the calibration of features, the reinforcement learning is used to quantitatively analyze the lesion and the region around the breast. Combined with the revenue target, the learning is carried out through continuous attempts. The goal is to obtain the maximum revenue value. Based on the reinforcement learning method, it is found that breast cancer does not need to know how to produce correct breast cancer recognition action. Reinforcement learning relies on its own learning, constantly trying and making mistakes and constantly recording the maximum value of income in the process of trial and error until the method of finding the maximum value of income is found.

3. Breast Cancer Image Segmentation

Based on the given feature factors, the medical image segmentation compares the similarity of feature factors between images and divides the image into several regions. The objects of medical image segmentation mainly include cells, tissues, and organs. The region-based segmentation method is mainly based on the spatial local features of the image, such as gray, texture, and other pixel statistical characteristics. The boundary-based segmentation method mainly uses gradient information to determine the boundary of the target. For example, the fast marching algorithm and the medical image segmentation method of watered transform can segment the image quickly and accurately [15].

In recent years, with the development of other emerging disciplines, image segmentation technology develops rapidly, and new methods generated by interdisciplinary emerge in endlessly. Some new image segmentation techniques have been developed for breast cancer detection, such as the method based on statistics, the method based on fuzzy theory, the method based on neural network, the method based on wavelet analysis, the model-based snake model (dynamic contour model), and the combination optimization model. Although new segmentation methods have been proposed, the results are not ideal. At present, the research focus is a knowledge-based segmentation method; that is, some prior knowledge is introduced into the segmentation process by some means, so as to constrain the computer segmentation process, so that the segmentation results can be controlled within the range we can understand without going too far [16]. For example, when the gray value of the tumor in the liver is very different from that of the normal

liver, the tumor and the normal liver will not be regarded as two independent tissues. All of the above methods have their own limitations. The results are good in specific scenarios, but not ideal beyond specific scenarios. Because the boundaries of internal organs, muscles, blood vessels, and other organs are usually very complex, the diseased areas of organs are mixed with normal tissues, and the differences between the gray levels and boundaries of diseased areas and normal tissues are relatively small, it is very difficult to identify these organs and diseased areas in medical images, and the existing image segmentation algorithms can not complete the task of image segmentation independently. Human intervention is also needed [17]. Medical image segmentation is quite different from image segmentation in other fields. The effect of existing classical algorithms in medical image segmentation is not good. It is still necessary to continue to study in improving the accuracy, speed, adaptability, and robustness of image segmentation [18]. The image segmentation method based on prior knowledge can well control the segmentation boundary of the image. For example, in the segmentation of intrahepatic mass, the image segmentation method based on prior knowledge can recognize intrahepatic mass and normal liver by gray value. However, image segmentation based on prior knowledge requires a large number of prior data. The more prior data, the more accurate the results. For example, Ghesu et al., based on 2891 times of cardiac ultrasound data, used deep learning and edge space learning to detect and segment cardiac ultrasound images [17, 18]. Parameter space exploration and data sparsity are important factors to improve the efficiency of medical image segmentation. Brosch et al. proposed a 3D deep convolution coder network through convolution and deconvolution to segment multiple sclerosis brain lesions and normal brain regions [19]. Data normalization and data enhancement techniques are applied to image enhancement and core regions of suspected tumors in brain tumor segmentation research and achieved good results [20]. The research of medical image segmentation methods has the following remarkable characteristics: it is difficult for any single existing image segmentation algorithm to achieve satisfactory results for general images, so more attention should be paid to the effective combination of multiple segmentation algorithms. Due to the complexity of human anatomical structure and the systematicness of function, although there have been studies on the methods of automatic segmentation of medical images to distinguish the required organs and tissues or find the lesion area, the existing software packages generally can not complete the automatic segmentation, and the manual intervention of anatomy is still needed [21]. At present, it is impossible for computer to complete the task of image segmentation, so the human-computer interactive segmentation method has gradually become the focus of research. The research of new segmentation methods mainly focuses on automatic, accurate, fast, adaptive, and robust features. The comprehensive utilization of classical segmentation technology and modern segmentation technology is the development direction of medical image segmentation technology in the future [22, 23].

4. Breast Cancer Image Registration

Image registration is the first mock exam of image fusion. In breast cancer clinical diagnosis, multiple modes or modes of image registration and fusion are needed. More information can help doctors give more accurate diagnosis [24]. In the clinical diagnosis process of breast cancer, medical image registration mainly locates reference points in two or more images; through spatial location transformation, such as rotation, the reference point is located in a coordinate system. Registration requires that the mapping of points between images is one-to-one correspondence; that is, each point in an image space has corresponding points in another image space, or in the sense of medical diagnosis, the points in the image can be accurately or approximately accurately corresponded [25–27]. Registration can be divided into two types: based on external features and based on internal features. Registration based on image internal features is noninvasive and traceable, which is the focus of research on registration algorithm [28].

There are two major categories of medical registration research for breast cancer: (1) a deep learning network is used to estimate the similarity of two images and drive iterative optimization, and (2) the depth regression network is directly used to predict the conversion parameters. The former only uses deep learning for similarity measurement and still needs the traditional registration method for iterative optimization. It does not give full play to the advantages of deep learning, takes a long time, and is difficult to achieve real-time registration. Therefore, only the latter is studied and discussed, and the conclusion is limited to this kind of nonrigid registration method. Based on supervised learning, there are two ways to obtain tags: (1) the traditional classical registration method is used for registration, and the deformation field is used as tags. (2) The original image is simulated as a fixed image, the deformed image as a moving image, and the simulated deformation field as a label. Based on unsupervised learning, the registration pair is input into the network to obtain the deformation field, and the moving image is interpolated to obtain the registration image. The 3D image is similar to it. The 3D image is input into the network to obtain the deformation field (dx , dy , dz), and then, the registration image is obtained by interpolation. However, the medical image registration of breast cancer is still an unsolved classic problem. There are no universally recognized gold standard and no corresponding large database in this field. Deep learning methods have some successful cases in breast cancer image registration. There are usually several reasons: (1) the expert knowledge of the field is well utilized, (2) the data are properly preprocessed and processed by data enhancement, (3) a special network structure is designed for a single task, and (4) the appropriate super parameter optimization method is used: such as parameter adjustment based on intuition or the Bayesian method. There are still some difficulties and challenges in the field of breast cancer medical image registration: (1) there is a lack of large databases with precise annotation. (2) Specific tasks need guidance from experts in the field. (3) It is difficult to agree with the opinions of different experts in some

ambiguous images. (4) The two classification models are too simple to be competent for more complex cases. (5) the difficulties of breast cancer medical image analysis still exist in images. Besides the analysis, we need to make full use of the information about other dimensions of the patient, like age, medical history, etc. (6) The slice-based neural network is having difficulty in using the location information of the corresponding anatomical structure in the original image, but the method of transferring the whole image into the neural network has corresponding disadvantages. Although Esteva et al. made amazing progress in dermatology in 2017 and Gulshan et al. in ophthalmology in 2016, they have achieved image classification models with higher accuracy than human experts in both fields. However, the essential reason for its success is that the above two problems are simple enough, and the data volume of ImageNet dataset is very large, so the existing model is applied to the above two problems and has achieved very good results. Usually, there is no such simple structure, and there is no effective network structure that can be used for 3D gray or multichannel image preprocessing. The advantage of unsupervised learning over supervised learning is that unsupervised learning does not need a large number of precisely labeled data. In addition, unsupervised learning imitates the way of human learning and can automatically perform the required tasks without special labeling or can deal with a large number of classification problems with little supervision. At present, the main methods of unsupervised learning are self-coding VAE and counter neural network Gan. Compared with accurately labeled breast cancer medical images, unlabeled breast cancer medical image data are easier to obtain. Unsupervised learning can directly use standard breast cancer medical images without any supervision, so it has more advantages. Finally, because deep learning is similar to black box model, the interpretation of breast cancer medical image domain requires a higher model, and further work is needed. At present, some work includes introducing Bayesian statistics into deep learning, which will help to measure the uncertainty of prediction.

Multimodality medical image registration is a new direction of breast cancer registration, such as nonrigid multimodal medical image registration based on structure representation of PCANet [29]. PCANet can automatically learn intrinsic features from a large number of medical images through multilevel linear and nonlinear transformation, which has better information entropy than the artificial feature extraction method. Multilevel image features extracted from each layer of PCANet can effectively represent multimodal images. The fusion of medical image registration technology and informatics theory opens a new idea for medical image registration of breast cancer. For example, the principle of maximum information entropy is applied to image registration, which can maximize the diversity of information and retain the main information without neglecting the secondary information [30–35]. Three-dimensional multimode image registration is a new direction of medical image registration. It has more information than two-dimensional image and can support a doctor's diagnosis more effectively. In addition, some new image reg-

istration algorithms, such as image recognition of breast cancer based on topology, feature points are extracted from existing breast cancer images. They are combined into a matching area with a certain topological structure as the matching template. In the breast images to be matched, regions with similar topological structures are found; these regions may be breast cancer. The main steps of image recognition of breast cancer based on topology are as follows: (1) the first one is extracting feature points or feature regions of a specific scale and combining them into topological templates. (2) The topology of the image to be matched is extracted. (3) By comparing the topology in the image to be matched with the topology template, regions with similar topology are found. (4) The similarity between similar topology and feature points in topology template is compared, and the product of topology similarity and feature point similarity is regarded as the final similarity. The schematic diagram of image recognition based on topology structure is shown in Figure 1. In Figure 1, the left dotted line is the topology template, the middle dotted line is the topology extracted from the image to be matched, and the right dotted line is the region with similar topology [36].

Other methods such as algorithms based on wavelet transform, statistical parametric mapping algorithm, and genetic algorithm are also continuously integrated into breast cancer image registration. The combination of multi-objective optimization, reinforcement learning, and other methods with medical image registration is the future development direction of medical image registration.

5. Breast Cancer Image Fusion

Breast cancer image fusion extracts useful information from multiple images, filters redundant information, and improves the medical value of images. In general, image fusion from low to high is signal level fusion, data level fusion, feature level fusion, and decision level fusion.

- (1) Signal level: at the lowest level, the unprocessed sensor output is mixed in the signal domain to produce a fused signal. The fused signal has the same form as the source signal, but its quality is better. The signal from the sensor can be modeled as random variables mixed with different correlated noises. In this case, fusion can be considered an estimation process, and signal level image fusion is the optimal concentration or distribution detection problem of signal to a large extent, which requires the highest registration in time and space.
- (2) Pixel level: pixel level image fusion is the most basic fusion of the three levels. After pixel level image fusion, the obtained image has more detailed information, such as edge and texture extraction, which is conducive to the further analysis, processing, and understanding of the image. It can also expose the potential target, which is conducive to the operation of judging and identifying the potential target pixels. This method can save as much information as

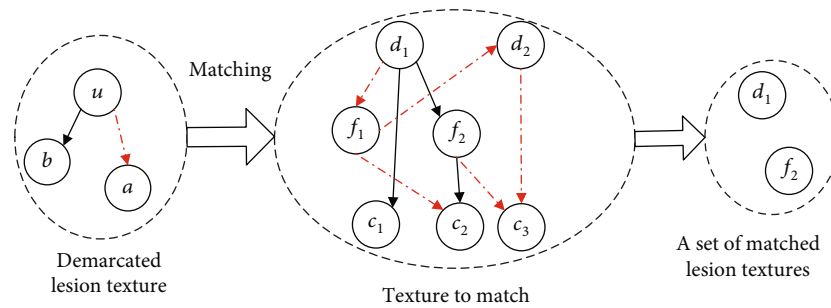


FIGURE 1: Image recognition sketch based on topological structure.

possible in the source image and increase the content and details of the fused image [37]. This advantage is unique and only exists in pixel level fusion. However, the limitations of pixel level image fusion can not be ignored, because it is to operate on pixels, so the computer has to process a large number of data, and the processing time will be relatively long, so the fused image can not be displayed in time and real-time processing can not be realized. In addition, in data communication, the amount of information is large, and it is easy to be affected by noise. In addition, if you directly participate in image fusion without strict image registration, the fused image will be blurred, and the target and details are not clear and accurate.

- (3) Feature level: feature level image fusion is to extract the feature information from the source image. The feature information is the information of the target or the region of interest in the source image, such as edge, person, building, or vehicle. Then, the feature information is analyzed, processed, and integrated to get the fused image features. The accuracy of target recognition based on fused features is obviously higher than that of the original image. The image information is compressed by feature level fusion and then analyzed and processed by computer. Compared with pixel level, the memory and time consumed will be reduced, and the real-time performance of the required image will be improved. Feature level image fusion requires less accuracy of image matching than the first layer, and its computing speed is faster than the first layer. However, it extracts image features as fusion information, so it will lose a lot of detail features.
- (4) Decision level: decision level image fusion is a cognitive-based method, which is not only the highest level of image fusion method but also the highest level of abstraction. Decision level image fusion is targeted [38]. According to the specific requirements of the problem, the feature information obtained from the feature level image is used, and then, the optimal decision is made directly according to certain criteria and the credibility of each decision, that is, the probability of the existence of the target. Among the three fusion levels, the calculation of

decision level image fusion is the smallest, but this method has a strong dependence on the previous level, and the image is not very clear compared with the former two fusion methods. It is difficult to realize the decision level image fusion, but the noise has the least influence on the image transmission.

To sum up, data level fusion is the process of directly processing the collected data to obtain the fused image, which is the basis of high-level image fusion. Feature level fusion preserves the information contained in different images. Decision level fusion is the highest level of image fusion based on subjective needs. In breast medical image fusion, data level fusion is the main method. For example, multimodality medical image fusion is a technology that integrates multiple dimensions information. It can provide more comprehensive and accurate information for clinical detection of breast cancer [39]. The steps of image fusion are mainly divided into image data fusion and fusion image display. At present, the data fusion of breast image is mainly based on pixels, which process the image point by point and sum the gray values of the corresponding pixels of the two images. However, the image will be blurred to a certain extent after using this method. The fusion method based on breast image features needs to extract image features and do target segmentation and other processing on the image. The display of fusion image includes pseudocolor display, tomographic display, and three-dimensional display. Pseudocolor display takes an image as a benchmark and superimposes the gray and contrast features of the image to be fused with the benchmark image. The tomographic display method can display the fused three-dimensional data synchronously in cross-sectional, coronal, and sagittal images, which are convenient for the observer to diagnose. The three-dimensional display method, namely, three-dimensional reconstruction, is to display the fused breast data in the form of three-dimensional images, which can more intuitively observe the spatial anatomical position of the lesions. The earliest method of 3D reconstruction is back projection. At present, there are two common reconstruction methods: filtered back projection and convolution back projection. The information content of three-dimensional image is large, and the future three-dimensional image fusion technology will be a focus of image fusion research. With the development of interdisciplinary research, new image fusion methods are emerging. The application of wavelet transform, nonlinear registration based on finite

element analysis, and artificial intelligence technology in breast image fusion will be the focus of image fusion research.

6. Forecast and Challenge

- (1) The application of unsupervised and supervised learning in the field of breast cancer image processing: breast cancer image classification based on CNN is the mainstream classification method nowadays. Fine tuning of CNN parameters directly affects the final image processing results. Unsupervised learning will have more promising results in breast cancer image processing
- (2) Transfer learning and fine tuning in the application of breast cancer image processing: transfer learning can partially alleviate the plight of not enough annotation data. In the process of transfer learning, a better plan is to use pretrained CNN as initialization of network and then carry out further supervision training. The annotated dataset is still a challenge in the field of breast cancer image processing
- (3) Patients pay more and more attention to privacy protection. The privacy of breast cancer image data has attracted much attention. In the process of breast cancer image processing, it is necessary to improve the accuracy of recognition, help doctors to give diagnosis, and pay attention to the protection of patients' privacy, protect the original information of the image from unauthorized access, and diagnose the whole process of image information authorized access and make the image access trace checkable

7. Conclusion

Deep learning and reinforcement learning are the relatively close combination of machine learning algorithm and breast cancer image processing and have made considerable progress. There are obvious differences between breast cancer image and other fields such as noise reduction, grayscale transformation, target segmentation, and feature extraction. The traditional image processing method is not directly applied to breast cancer image processing. Accumulation of deep learning in image processing can not be directly transferred to breast cancer image processing. Reinforcement learning belongs to unsupervised learning, which is different from deep learning. It uses an incentive mechanism that does not need a large number of sample space. Compared with deep learning, reinforcement learning has a wider applicability and lower promotion cost. Moreover, reinforcement learning has achieved very good results in chess, man-machine game, and other fields, which is suitable for complex logic processing. The combination of reinforcement learning and medical image processing will play a greater role in the clinical detection and prediction of breast cancer.

Conflicts of Interest

The authors declare that they have no conflicts of interest.




References

- [1] C. W. Elston, I. O. Ellis, and S. E. Pinder, "Pathological prognostic factors in breast cancer," *Critical Reviews in Oncology/Hematology*, vol. 31, no. 3, pp. 209–223, 1999.
- [2] J. K. Udupa and G. T. Herman, "Medical image reconstruction, processing, visualization, and analysis: the MIPG perspective," *IEEE Transactions on Medical Imaging*, vol. 21, no. 4, pp. 281–295, 2002.
- [3] H. R. Roth, L. Lu, J. Liu et al., "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170–1181, 2016.
- [4] G. Cserni, I. Amendoeira, N. Apostolikas et al., "Pathological work-up of sentinel lymph nodes in breast cancer. Review of current data to be considered for the formulation of guidelines," *European Journal of Cancer*, vol. 39, no. 12, pp. 1654–1667, 2003.
- [5] Y. Hou, "Breast cancer pathological image classification based on deep learning," *Journal of X-Ray Science and Technology*, vol. 28, no. 4, pp. 727–738, 2020.
- [6] S. G. Wu, J. Wang, J. Lei et al., "Prognostic validation and therapeutic decision-making of the AJCC eighth pathological prognostic staging for T3N0 breast cancer after mastectomy," *Clinical and Translational Medicine*, vol. 10, no. 1, pp. 125–136, 2020.
- [7] T. Mcinerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis*, vol. 1, no. 2, pp. 91–108, 1996.
- [8] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [9] T. M. D. (ne Lehmann), H. Handels, K. H. M. H. (ne Fritzsche) et al., "Viewpoints on medical image processing: from science to application," *Current Medical Imaging Reviews*, vol. 9, no. 2, pp. 79–88, 2013.
- [10] A. Setio, F. Ciompi, G. Litjens et al., "Pulmonary nodule detection in CT Images: False Positive Reduction using Multi-View convolutional networks," *Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.
- [11] Z. Guo, H. Liu, H. Ni et al., "Publisher correction: a fast and refined cancer regions segmentation framework in whole-slide breast pathological images," *Scientific Reports*, vol. 10, no. 1, p. 8591, 2020.
- [12] H. R. Roth, L. Lu, A. Seff et al., "A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. MICCAI 2014*, P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, Eds., vol. 8673 of Lecture Notes in Computer Science, pp. 520–527, Springer, Cham, 2014.
- [13] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Išgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.

- [14] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, 2000.
- [15] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: interpolation methods in medical image processing," *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, pp. 1049–1075, 1999.
- [16] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for medical image analysis and computer vision," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 4322, no. 1, 2001.
- [17] F. C. Ghesu, B. Georgescu, T. Mansi, D. Neumann, J. Hornegger, and D. Comaniciu, "An artificial agent for anatomical landmark detection in medical images," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016. MICCAI 2016*, S. Ourselin, L. Joskowicz, M. Sabuncu, G. Unal, and W. Wells, Eds., vol. 9902 of Lecture Notes in Computer Science, Springer, Cham, 2016.
- [18] F. C. Ghesu, E. Krubasik, B. Georgescu et al., "Marginal space deep learning: efficient architecture for volumetric image parsing," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1217–1228, 2016.
- [19] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Trabousee, and R. Tam, "Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *Medical Imaging*, vol. 35, no. 5, pp. 1229–1239, 2016.
- [20] N. Wahab, A. Khan, and Y. S. Lee, "Transfer learning based deep CNN for segmentation and detection of mitoses in breast cancer histopathological images," *Microscopy*, vol. 68, no. 3, pp. 216–233, 2019.
- [21] A. Sohail, M. A. Mukhtar, A. Khan, M. M. Zafar, A. Zameer, and S. Khan, "Deep Object Detection Based Mitosis Analysis in Breast Cancer Histopathological Images," (2020), <http://arxiv.org/abs/2003.08803>.
- [22] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," *2016 Fourth International Conference on 3D Vision (3DV)*, 2016pp. 565–571, Stanford, CA, USA, 2016.
- [23] S. Ishihara, K. Ishihara, M. Nagamachi, and Y. Matsubara, "An analysis of Kansei structure on shoes using self-organizing neural networks," *International Journal of Industrial Ergonomics*, vol. 19, no. 2, pp. 93–104, 1997.
- [24] J. B. Maintz and M. A. Viergever, "A survey of medical image registration," *Computer & Digital Engineering*, vol. 33, no. 1, pp. 140–144, 2009.
- [25] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine & Biology*, vol. 31, no. 4, pp. 1–45, 2008.
- [26] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
- [27] K. Kamnitsas, C. Ledig, V. F. J. Newcombe et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, p. 61, 2016.
- [28] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or fine tuning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [29] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.
- [30] Hoo-Chang Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1930–1943, 2013.
- [31] B. Sahiner, Heang-Ping Chan, N. Petrick et al., "Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 598–610, 1996.
- [32] D. Nie, L. Wang, Y. Gao, and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1342–1345, Prague, Czech Republic, 2016.
- [33] H. R. Roth, L. Lu, A. Farag et al., "DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. MICCAI 2015*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., vol. 9349 of Lecture Notes in Computer Science, pp. 556–564, Springer, Cham, 2015.
- [34] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," (2016), <http://arxiv.org/abs/1606.05718>.
- [35] R. Li, W. Zhang, H. I. Suk et al., "Deep learning based imaging data completion for improved brain disease diagnosis," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2014*, vol. 8675 of Lecture Notes in Computer Science, pp. 305–312, Springer, Cham, 2014.
- [36] Z. YaNan, Q. MingCheng, and L. YuPeng, "Recommendation method based on social topology for cold-start users," *Journal of Zhejiang University(Engineering Science)*, vol. 50, pp. 1001–1008, 2016.
- [37] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," in *Medical Imaging 2015: Computer-Aided Diagnosis*, International Society for Optics and Photonics, 2015.
- [38] T. Kooi, G. Litjens, B. van Ginneken et al., "Large scale deep learning for computer aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp. 303–312, 2017.
- [39] C. Yu, H. Chen, Y. Li, Y. Peng, J. Li, and F. Yang, "Breast cancer classification in pathological images based on hybrid features," *Multimedia Tools and Applications*, vol. 78, no. 15, pp. 21325–21345, 2019.

Research Article

Machine Learning Models for Survival and Neurological Outcome Prediction of Out-of-Hospital Cardiac Arrest Patients

Chi-Yung Cheng ^{1,2}, I-Min Chiu ^{1,2}, Wun-Huei Zeng,² Chih-Min Tsai,³
and Chun-Hung Richard Lin ²

¹Department of Emergency Medicine, Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Kaohsiung, Taiwan

²Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan

³Department of Pediatrics, Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Kaohsiung, Taiwan

Correspondence should be addressed to I-Min Chiu; ray1985@cgmh.org.tw and Chun-Hung Richard Lin; lin@cse.nsysu.edu.tw

Received 27 April 2021; Revised 20 August 2021; Accepted 1 September 2021; Published 20 September 2021

Academic Editor: Yue Zhang

Copyright © 2021 Chi-Yung Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Out-of-hospital cardiac arrest (OHCA) is a major health problem worldwide, and neurologic injury remains the leading cause of morbidity and mortality among survivors of OHCA. The purpose of this study was to investigate whether a machine learning algorithm could detect complex dependencies between clinical variables in emergency departments in OHCA survivors and perform reliable predictions of favorable neurologic outcomes. **Methods.** This study included adults (≥ 18 years of age) with a sustained return of spontaneous circulation after successful resuscitation from OHCA between 1 January 2004 and 31 December 2014. We applied three machine learning algorithms, including logistic regression (LR), support vector machine (SVM), and extreme gradient boosting (XGB). The primary outcome was a favorable neurological outcome at hospital discharge, defined as a Glasgow-Pittsburgh cerebral performance category of 1 to 2. The secondary outcome was a 30-day survival rate and survival-to-discharge rate. **Results.** The final analysis included 1071 participants from the study period. For neurologic outcome prediction, the area under the receiver operating curve (AUC) was 0.819, 0.771, and 0.956 in LR, SVM, and XGB, respectively. The sensitivity and specificity were 0.875 and 0.751 in LR, 0.687 and 0.793 in SVM, and 0.875 and 0.904 in XGB. The AUC was 0.766 and 0.732 in LR, 0.749 and 0.725 in SVM, and 0.866 and 0.831 in XGB, for survival-to-discharge and 30-day survival, respectively. **Conclusions.** Prognostic models trained with ML technique showed appropriate calibration and high discrimination for survival and neurologic outcome of OHCA without using prehospital data, with XGB exhibiting the best performance.

1. Introduction

Out-of-hospital cardiac arrest (OHCA) is a major public health problem worldwide, with an annual incidence of 50 to 100 per 100,000 in the general population [1]. OHCA has a high societal burden when compared to all other major causes of death, with an estimated 2.04 million years of potential life lost for men and 1.29 million years for women [2]. Despite advances in prehospital care, the prognosis for OHCA remains limited, with only 5.4%–20% [3–5] of patients surviving to hospital discharge. Neurologic injury

remains the leading cause of morbidity and mortality among survivors of OHCA, because of inadequate cerebral perfusion during cardiac arrest or reperfusion injury that occurs in the early postresuscitation phase. The Pan Asian Resuscitation Outcomes Study (PAROS) Clinical Research Network demonstrated that the survival rate with proper neurological function was only 2.7% [5].

Many prehospital factors improve survival following OHCA, including witnessed cardiac arrest, bystander cardiopulmonary resuscitation (CPR), and initial heart rhythm [6–8]. The time from collapse to initiation CPR (no-flow

interval) and the duration of CPR (low-flow interval) were also considered predictors of outcomes [9]. Severe scores were developed for predicting survival with proper neurological function at the time of ICU admission after OHCA. The OHCA score comprised five parameters, including the initial heart rhythm, no-flow interval, low-flow interval, serum creatinine, and arterial lactate [10]. The CAHP score stratified patients into three-level groups using seven variables, including age, initial heart rhythm, no-flow interval, low-flow interval, location of cardiac arrest, epinephrine dose, and arterial pH [11]. However, no-flow or low-flow intervals may be the result of inaccurate recall or recording during a highly stressful event. The updated Utstein template eliminated the necessity for recording the time of collapse, and thus, the duration of the no-flow interval could not be calculated [12].

In the past few years, machine learning (ML) techniques were used to influence clinical research and practice, such as prediction of sepsis through digital biomarker discovery [13], prediction of mortality for intensive care patients [14], and prediction of outcome in traumatic brain injury [15]. The ML algorithms outperform conventional triage tools and early warning scores in detecting patients at risk for cardiac arrest in emergency departments [16]. They can also accurately predict the need for critical care on information acquired during emergency medical services [17].

Previous studies have suggested that ML methods could predict neurologic and survival outcomes of OHCA patients [18–21]. Harford et al. found that an ML model can be used to support intervention decisions such as CPR or coronary angiography in OHCA patients [18]. However, only limited studies examined independent variables after patients arrived at the emergency department (ED). This study is aimed at investigating whether an ML algorithm could detect complex dependencies between clinical variables during ED in OHCA survivors and performing reliable predictions of the favorable neurological outcome.

2. Materials and Method

2.1. Study Setting and Variables. This was a retrospective study conducted from 1 January 2004 to 31 December 2014 in a tertiary medical center of southern Taiwan, which had 72,000 ED visits on average every year. The Ethics Committee of Chang Gung Memorial Hospital (No. 202001675B0) approved the study protocol. Because of the study's retrospective nature, informed consent from the subjects was not required.

The study included adults (≥ 18 years of age) who had a sustained return of spontaneous circulation (ROSC) after successful resuscitation from OHCA and were then admitted to ICU. The demographic characteristics, baseline comorbidities, and clinical variables were extracted from the ED electronic database. The underlying medical conditions included heart failure, cerebrovascular disease, peripheral vascular disease, diabetes mellitus, chronic obstructive pulmonary disease, chronic kidney disease, liver cirrhosis, malignancy, metastatic tumor, dementia, and moderate to severe Charlson comorbidity index

(CCI) (CCI scored ≥ 3) [22]. Tentative diagnosis of cardiac arrest causes, such as hypothermia, hyperkalemia, acidosis ($\text{pH} < 7.1$), acute myocardial infarction (AMI), pulmonary embolism, tension pneumothorax, or intoxication, at the ED was recorded. Medication administration, including epinephrine, sodium bicarbonate, dopamine, norepinephrine, amiodarone, lidocaine, and calcium use or not, was collected. Intervention at ED included percutaneous coronary intervention and extracorporeal membrane oxygenation.

The primary outcome was a favorable neurological outcome at hospital discharge, defined as a Glasgow-Pittsburgh cerebral performance category (CPC) of 1 to 2. The favorable neurological outcome included patients with full recovery or those who can independently perform daily activities but may have a minor to moderate disability. However, CPC 3–5 was categorized as a poor functional outcome, which included patients dependent on others, in a coma or vegetative state, and who are dead [23, 24]. In this study, CPC scores were collected retrospectively using electronic medical records and physical examinations by a consensus of neurologists who were blinded to the study. The secondary outcome was the 30-day survival rate and survival-to-discharge rate.

2.2. Stepwise Feature Selection and ML Algorithms. To detect the model performance between features and subsequently select the best performing subset, all collected features were subjected to stepwise feature selection. The stepwise approach started with the evaluation of each individual feature based on forward feature selection and then checked for elimination. In each step, a variable was considered for addition to or subtraction from the set of explanatory variables based on mean accuracy.

We applied three ML algorithms including logistic regression (LR), support vector machine (SVM), and extreme gradient boosting (XGB). LR is a supervised classification algorithm. It transforms its output using a sigmoid function to return a probability value, which can then be mapped to two or more discrete classes. SVM belongs to the supervised learning technique for classification, increasingly used in many data mining and bioinformatics applications. SVM constructs a hyperplane based on the support vectors and maximizes the gap width between the two categories [25, 26]. XGB is a gradient boosted tree algorithm used for regression, binary and multiclass classification, and ranking problems. XGB is a robust and supervised learning algorithm capable of handling various data types, relationships, distributions, and hyperparameters that can be fine-tuned by users [27].

2.3. Outcome Prediction and Statistical Analysis. Categorical data are expressed as counts and proportions, and continuous data are expressed as means and standard deviations. The patients enrolled were randomly separated into the training set (90%) and test set (10%) for independent performance measurement of the model's generalizability. The training set was randomly divided into ten equal-sized groups for cross-validation during model development. We

TABLE 1: Characteristics of the patients at baseline.

Variables	All patients ($n = 1071$)
<i>Demographic characteristics</i>	
Age (years), mean \pm SD	66.2 \pm 16.8
Sex, male, n (%)	596 (55.6)
<i>Underlying medical conditions, n (%)</i>	
Heart failure	161 (15.0)
Cerebrovascular disease	248 (23.2)
Peripheral vascular disease	37 (3.5)
Diabetes mellitus	244 (22.8)
Chronic obstructive pulmonary disease	247 (23.1)
Chronic kidney disease	232 (21.7)
Liver cirrhosis	15 (1.4)
Malignancy	146 (13.6)
Tumor metastasis	23 (2.1)
Dementia	100 (9.3)
CCI scored ≥ 3	715 (61.8)
<i>Laboratory data, mean \pm SD</i>	
White blood cell (1,000/ μ L)	13.651 \pm 7.4871
Segmented neutrophils (%)	53.05 \pm 19.671
Band neutrophils (%)	2.36 \pm 4.143
Hemoglobin (g/dL)	11.056 \pm 2.9859
Creatinine (mg/dL)	2.570 \pm 2.90
Alanine aminotransferase (ALT) (U/L)	248.97 \pm 560.968
Na (mEq/L)	138.97 \pm 7.935
K (mEq/L)	5.029 \pm 1.588
Troponin I (ng/mL)	0.801 \pm 5.364
pH	7.165 \pm 0.226
<i>ED diagnosis, n (%)</i>	
Hypothermia	5 (0.5)
Hyperkalemia	216 (20.2)
Acidosis	722 (67.4)
Acute myocardial infarction	140 (13.1)
Pulmonary embolism	4 (0.4)
Tension pneumothorax	3 (0.3)
Toxin	30 (2.8)
Diabetes ketoacidosis	27 (2.5)
<i>Medication and intervention</i>	
Epinephrine use, n (%)	1050 (98.0)
Epinephrine dose, mean \pm SD	5.35 \pm 4.917
Sodium bicarbonate use, n (%)	690 (64.4)
Dopamine use, n (%)	655 (61.2)
Norepinephrine use, n (%)	212 (19.8)
Amiodarone use, n (%)	179 (16.7)
Lidocaine use, n (%)	38 (3.5)
Calcium use, n (%)	196 (18.3)
Defibrillation at ED, n (%)	93 (8.7)
PCI, n (%)	86 (8.0)
ECMO, n (%)	18 (1.7)

TABLE 1: Continued.

Variables	All patients ($n = 1071$)
<i>Outcome, n (%)</i>	
CPC class 1 or 2	86 (8.0)
Survival-to-discharge	216 (20.2)
30-day survival	249 (23.2)

CCI: Charlson comorbidity index; PCI: percutaneous coronary intervention; ECMO: extracorporeal membrane oxygenation; CPC: cerebral performance category.

examined the area under the receiver operating characteristic curve (AUC) for performance measurement and plotted the receiver operating characteristic (ROC) curve using sensitivity against (1-specificity) [28]. We also compared positive predictive value (PPV) (true positives/(true positives+false positives)), sensitivity (true positives/(true positives+false negatives)), and specificity (true negatives/(-true negatives+false positives)) between each prognostic model. The ML models were performed using Scikit-learn (version 0.22.2) with Python (version 3.8).

3. Experiment and Result

3.1. Dataset Description. For the study period, although there were 1076 patients, 1071 were included in our study for the final analysis. Five patients were excluded due to missing values. The mean age of the 1071 patients was 66.2 \pm 16.8 years. The dataset included 596 (55.6%) males. There were 86 (8%) patients with favorable neurological outcomes after discharge. Furthermore, the dataset had 249 (23.2%) patients with 30-day survival and 216 (20.2%) patients survived to discharge. The other population characteristics were categorized and presented as underlying disease, laboratory data, medication, and intervention at ED. ED diagnosis is demonstrated in Table 1.

3.2. Feature Engineering. All 42 variables were subjected to stepwise feature selection based on their individual importance and their effect on the mean accuracy to create the best performing subset prediction model. Figure 1 depicts the results of stepwise feature selection for the three ML models. Table 2 ranks the results of variables by importance. We used 10, 12, and 11 parameters for model training in the LR, SVM, and XGB algorithms, respectively. The parameters ranked by LR were PCI, DM, hemoglobin, troponin I, dementia, CCI, norepinephrine use, liver cirrhosis, hypokalemia, and tumor metastasis. For SVR, the features were troponin I, CCI, dementia, DKA, PCI, norepinephrine use, ECMO, pulmonary embolism, amiodarone use, pneumothorax, tumor metastasis, and acidosis. For XGB, the features were troponin I, epinephrine dose, heart failure, PCI, amiodarone use, calcium use, dementia, sodium bicarbonate use, band neutrophil, malignancy, and AMI.

3.3. Prediction. Table 3 demonstrates the comparison of prediction ability for neurological outcomes between the three ML models. The AUC was 0.819, 0.771, and 0.956 in LR, SVM, and XGB, respectively. The sensitivity and specificity

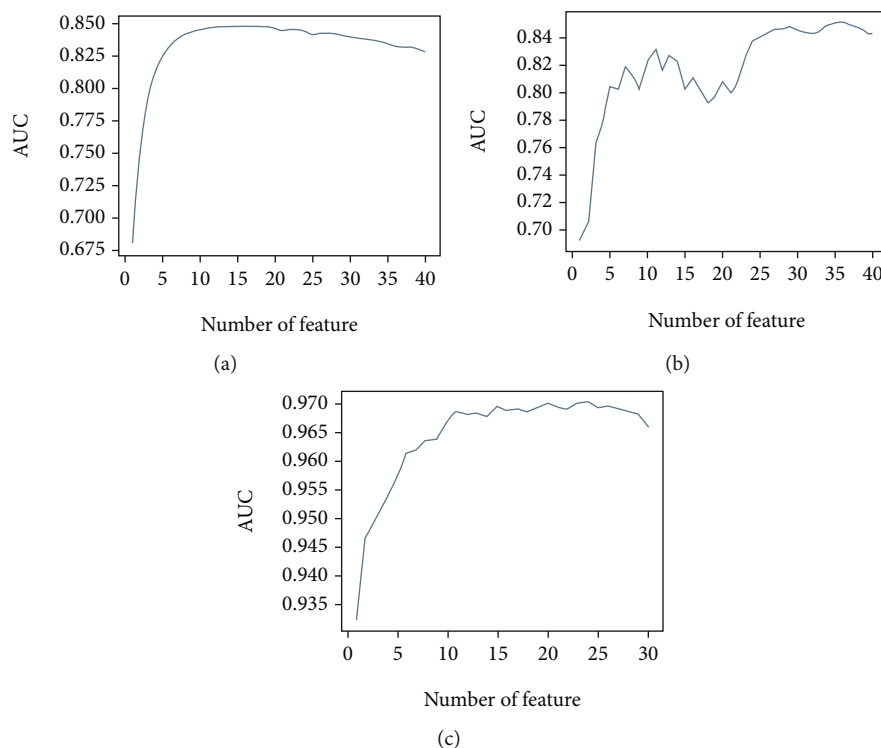


FIGURE 1: Forward stepwise feature selection of machine learning models based on AUC: (a) logistic regression; (b) support vector machine; (c) extreme gradient boosting.

TABLE 2: Rank of parameter importance after stepwise parameter selection.

Rank	LR	SVM	XGB
1st	PCI	Troponin I	Troponin I
2nd	Diabetes mellitus	CCI	Total epinephrine dose
3rd	Hemoglobin	Dementia	Heart failure
4th	Troponin I	Diabetes ketoacidosis	PCI
5th	Dementia	PCI	Amiodarone use
6th	CCI	Norepinephrine use	Calcium use
7th	Norepinephrine use	ECMO	Dementia
8th	Liver cirrhosis	Pulmonary embolism	Sodium bicarbonate use
9th	Hypokalemia	Amiodarone use	Band neutrophil
10th	Tumor metastasis	Pneumothorax	Malignancy
11th		Tumor metastasis	Acute myocardial infarction
12th		Acidosis	

LR: logistic regression; SVM: support vector machine; XGB: extreme gradient boosting; PCI: percutaneous coronary intervention; CCI: Charlson comorbidity index; ECMO: extracorporeal membrane oxygenation.

were 0.875 and 0.751 in LR, 0.687 and 0.793 in SVM, and 0.875 and 0.904 in XGB. Table 4 presents the comparison of prediction ability for survival-to-discharge and 30-day survival. The AUC was 0.766 and 0.732 in LR, 0.749 and 0.725 in SVM, and 0.866 and 0.831 in XGB, for survival-to-discharge and 30-day survival, respectively. Figure 2 depicts the ROC curve for the prediction performance of the three ML models.

4. Discussion

Using in-hospital data available within ED, we developed and validated different ML algorithms to stratify neurological outcomes after cardiac arrest. The AUC was 0.819, 0.771, and 0.956 in LR, SVM, and XGB, respectively. The sensitivity and specificity were 0.875 and 0.751 in LR, 0.687 and 0.793 in SVM, and 0.875 and 0.904 in XGB. The

TABLE 3: Area under the receiver operating curve, positive predictive value, sensitivity, and specificity between different machine learning models for neurologic outcome.

	LR	SVM	XGB
AUC	0.819 ± 0.017	0.771 ± 0.017	0.956 ± 0.003
PPV	0.229 ± 0.021	0.220 ± 0.044	0.437 ± 0.029
Sensitivity	0.875 ± 0.036	0.687 ± 0.005	0.875 ± 0.030
Specificity	0.751 ± 0.010	0.793 ± 0.004	0.904 ± 0.005

LR: logistic regression; SVM: support vector machine; XGB: extreme gradient boosting; AUC: area under the receiver operating curve; PPV: positive predictive value.

TABLE 4: Area under the receiver operating curve, positive predictive value, sensitivity, and specificity between different machine learning models for survival-to-discharge and 30-day survival.

	LR		SVM		XGB	
	Discharge	30 days	Discharge	30 days	Discharge	30 days
AUC	0.766 ± 0.020	0.732 ± 0.009	0.749 ± 0.013	0.725 ± 0.010	0.866 ± 0.006	0.831 ± 0.006
PPV	0.345 ± 0.016	0.354 ± 0.010	0.404 ± 0.018	0.368 ± 0.014	0.600 ± 0.029	0.564 ± 0.020
Sensitivity	0.780 ± 0.047	0.762 ± 0.019	0.720 ± 0.029	0.593 ± 0.021	0.840 ± 0.026	0.745 ± 0.018
Specificity	0.637 ± 0.012	0.579 ± 0.013	0.740 ± 0.009	0.692 ± 0.016	0.862 ± 0.005	0.825 ± 0.007

LR: logistic regression; SVM: support vector machine; XGB: extreme gradient boosting; AUC: area under the receiver operating curve; PPV: positive predictive value.

ML algorithm possessed suitable calibration and high discrimination in predicting favorable neurologic outcomes. For survival-to-discharge and 30-day survival prediction, the AUC was 0.766 and 0.732 in LR, 0.749 and 0.725 in SVM, and 0.866 and 0.831 in XGB, respectively. With acceptable outcome prediction ability, ML approaches are expected to improve clinician prognosis, earlier identification of outliers, information provision assistance, and physician-family communication.

In most of the current outcome prediction score and ML algorithms for OHCA, prehospital data are often implanted for predicting the variation in survival-to-discharge. The OHCA score, composed of five parameters, including no-flow and low-flow intervals, achieved an AUC of 0.82 in the development cohort and 0.88 in the validation cohort for neurological recovery outcome prediction [10]. Aschauer et al. discovered that using 21 variables, an LR model obtained an average AUC of 0.827 for survival probability, with key predictors being prehospital variables, such as the number of minutes to sustained restoration of spontaneous circulation and the first rhythm [29]. Another study cohort with 2639 patients, comparing several ML models (including decision tree, random forest (RF), *k*-nearest neighbors, XGB, light gradient boosting machine (GBM), and neural networks), stated that an embedded fully convolutional network model has the best average class sensitivity of 0.825 for neurological outcome prediction [18]. However, the above models required knowledge of the periods of time with circulatory no-flow and low-flow, limiting its use when prehospital data are unknown or recalled incorrectly. In our ML models, XGB exhibited the best performance with AUC of 0.956 for neurological outcome prediction, 0.866 for survival-to-discharge, and 0.831 for 30-day survival. The LR and XGB obtained a sensitivity of 0.875 for neurological out-

come prediction. Without using prehospital data, the result of XGB was not inferior to previous models.

Nanayakkara et al.'s study from the Australian and New Zealand Intensive Care Society included 39,566 OHCA cases without prehospital data, and five ML approaches (GBM, SVM, RF, artificial neural network, and an ensemble) were compared for predicting mortality. With a combination of demographic, physiologic, and biochemical information, an ensemble and GBM could reach AUC of 0.87 (95% CI 0.86–0.88) for predicting in-hospital mortality [30]. Similarly, the AUC for XGB reached 0.866 and 0.831 for survival-to-discharge and 30-day survival prediction in our study, respectively. However, Nanayakkara et al.'s study did not discriminate survival from neurological outcomes. In contrast, we also found that XGB exhibited satisfactory performance in neurological outcome prediction. To our knowledge, this is the first study using ML models to predict functional neurological outcomes post-OHCA using only in-hospital variables.

We determined the order of importance among features and the best subsets of features using forward stepwise regression. A forward selection begins with no explanatory features and then adds features alternately, in each step, based on which feature is the most statistically significant, until all statistically significant features have been tested. The process selects explanatory variables for multiple regression models and develops the best combination of feature subsets. Although it has been criticized for misapplying single-step statistical tests to a multistep procedure, stepwise regression is efficient at narrowing down a long list of plausible explanatory variables to a manageable number of predictors [31]. Although different ML models disagreed on feature importance in our study, troponin I and PCI remained among the top five features among all three

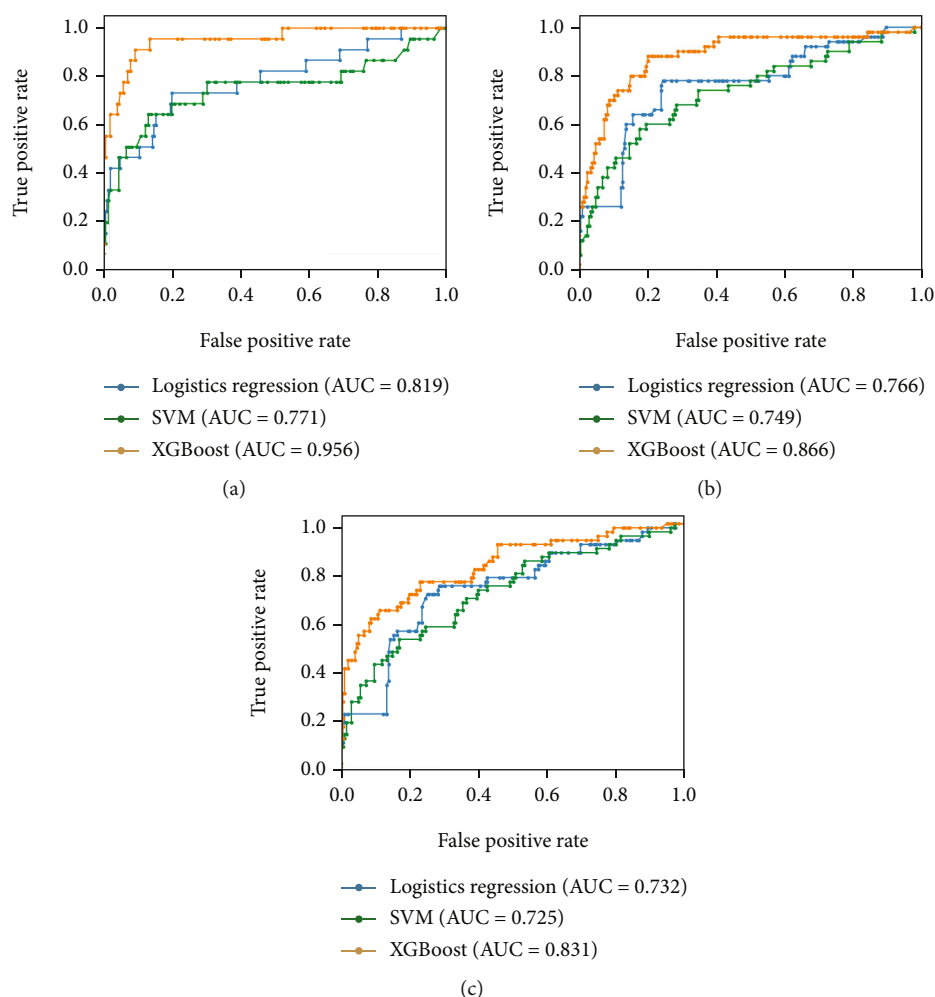


FIGURE 2: Receiver operating characteristic curve of three machine learning models: (a) favorable neurologic outcome; (b) survival-to-discharge; (c) 30-day survival.

models. Because AMI is a common cause of OHCA, some studies have demonstrated that short-term outcomes after OHCA due to AMI can be better than that due to other causes of OHCA [32, 33].

Furthermore, our study faced several limitations. First, we did not include prehospital features in our study. Although many prehospital factors can improve survival following OHCA [6–8], the ML algorithms incorporate the result of mediation before the time when measurements were taken. In other words, the models had computed a vector component triggered by earlier intervention. Second, the dataset used in this study only included patients from a tertiary medical center in southern Taiwan. The findings of this study must be validated in a different region with a more ethnically diverse patient population.

5. Conclusion

Prognostic models trained using ML technique demonstrated appropriate calibration and high discrimination for survival and neurological outcome of OHCA, without the use of prehospital data, with XGB providing the best performance.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' Contributions

CYC conceived the manuscript, performed the analyses, and wrote the manuscript. CMT contributed to data collection and measurements. WHZ was involved mainly in data analysis and quality management. IMC and CHRL provided overall supervision, edited the manuscript, and undertook the responsibility of submitting the manuscript for publication. IMC and CHRL are cocorresponding authors. All authors read and approved the final manuscript.

Acknowledgments

We gratefully acknowledge all persons who helped in this research. This study was supported by the research deputy of Kaohsiung Chang Gung Memorial Hospital.

References

- [1] C. X. Wong, A. Brown, D. H. Lau et al., "Epidemiology of sudden cardiac death: global and regional perspectives," *Heart, Lung & Circulation*, vol. 28, no. 1, pp. 6–14, 2019.
- [2] E. C. Stecker, K. Reinier, E. Marijon et al., "Public health burden of sudden cardiac death in the United States," *Circulation. Arrhythmia and Electrophysiology*, vol. 7, no. 2, pp. 212–217, 2014.
- [3] J. P. Pell, J. M. Sirel, A. K. Marsden, I. Ford, N. L. Walker, and S. M. Cobbe, "Presentation, management, and outcome of out of hospital cardiopulmonary arrest: comparison by underlying aetiology," *Heart*, vol. 89, no. 8, pp. 839–842, 2003.
- [4] W. T. Mathiesen, C. A. Bjørshol, J. T. Kvaløy, and E. Søreide, "Effects of modifiable prehospital factors on survival after out-of-hospital cardiac arrest in rural versus urban areas," *Critical Care*, vol. 22, no. 1, p. 99, 2018.
- [5] M. E. Ong, S. D. Shin, N. N. de Souza et al., "Outcomes for out-of-hospital cardiac arrests across 7 countries in Asia: the Pan Asian Resuscitation Outcomes Study (PAROS)," *Resuscitation*, vol. 96, pp. 100–108, 2015.
- [6] A. Hamilton, J. Steinmetz, M. Wissenberg et al., "Association between prehospital physician involvement and survival after out-of-hospital cardiac arrest: A Danish nationwide observational study," *Resuscitation*, vol. 108, pp. 95–101, 2016.
- [7] H. Sjøholm, C. Hassager, F. Lippert et al., "Factors associated with successful resuscitation after out-of-hospital cardiac arrest and temporal trends in survival and comorbidity," *Annals of Emergency Medicine*, vol. 65, no. 5, 2015.
- [8] M. Hara, K. Hayashi, S. Hikoso, Y. Sakata, and T. Kitamura, "Different impacts of time from collapse to first cardiopulmonary resuscitation on outcomes after witnessed out-of-hospital cardiac arrest in adults," *Circulation. Cardiovascular Quality and Outcomes*, vol. 8, no. 3, pp. 277–284, 2015.
- [9] F. Adnet, M. N. Triba, S. W. Borron et al., "Cardiopulmonary resuscitation duration and survival in out-of-hospital cardiac arrest patients," *Resuscitation*, vol. 111, pp. 74–81, 2017.
- [10] C. Adrie, A. Cariou, B. Mourvillier et al., "Predicting survival with good neurological recovery at hospital admission after successful resuscitation of out-of-hospital cardiac arrest: the OHCA score," *European Heart Journal*, vol. 27, no. 23, pp. 2840–2845, 2006.
- [11] C. Maupain, W. Bougouin, L. Lamhaut et al., "The CAHP (cardiac arrest hospital prognosis) score: a tool for risk stratification after out-of-hospital cardiac arrest," *European Heart Journal*, vol. 37, no. 42, pp. 3222–3228, 2016.
- [12] G. D. Perkins, I. G. Jacobs, V. M. Nadkarni et al., "Cardiac arrest and cardiopulmonary resuscitation outcome reports: update of the Utstein resuscitation registry templates for out-of-hospital cardiac arrest," *Circulation*, vol. 132, no. 13, pp. 1286–1300, 2015.
- [13] G. Holmgren, P. Andersson, A. Jakobsson, and A. Frigyesi, "Artificial neural networks improve and simplify intensive care mortality prognostication: a national cohort study of 217,289 first-time intensive care unit admissions," *Journal of Intensive Care*, vol. 7, no. 1, 2019.
- [14] R. Raj, T. Luostarinen, E. Pursiainen et al., "Machine learning-based dynamic mortality prediction after traumatic brain injury," *Scientific Reports*, vol. 9, no. 1, article 17672, 2019.
- [15] M. Moor, B. Rieck, M. Horn, C. R. Jutzeler, and K. Borgwardt, "Early prediction of sepsis in the ICU using machine learning: a systematic review," *Frontiers in Medicine*, vol. 8, article 607952, 2021.
- [16] D. H. Jang, J. Kim, Y. H. Jo et al., "Developing neural network models for early detection of cardiac arrest in emergency department," *The American Journal of Emergency Medicine*, vol. 38, no. 1, pp. 43–49, 2020.
- [17] D. Y. Kang, K. J. Cho, O. Kwon et al., "Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 28, no. 1, p. 17, 2020.
- [18] S. Harford, H. Darabi, M. del Rios et al., "A machine learning based model for out of hospital cardiac arrest outcome classification and sensitivity analysis," *Resuscitation*, vol. 138, pp. 134–140, 2019.
- [19] T. Seki, T. Tamura, M. Suzuki, and SOS-KANTO 2012 Study Group, "Outcome prediction of out-of-hospital cardiac arrest with presumed cardiac aetiology using an advanced machine learning technique," *Resuscitation*, vol. 141, pp. 128–135, 2019.
- [20] J. Johnsson, O. Björnsson, P. Andersson et al., "Artificial neural networks improve early outcome prediction and risk classification in out-of-hospital cardiac arrest patients admitted to intensive care," *Critical Care*, vol. 24, no. 1, 2020.
- [21] Y. Hirano, Y. Kondo, K. Sueyoshi, K. Okamoto, and H. Tanaka, "Early outcome prediction for out-of-hospital cardiac arrest with initial shockable rhythm using machine learning models," *Resuscitation*, vol. 158, pp. 49–56, 2021.
- [22] Y. Q. Huang, R. Gou, Y. S. Diao et al., "Charlson comorbidity index helps predict the risk of mortality for patients with type 2 diabetic nephropathy," *Journal of Zhejiang University. Science. B*, vol. 15, no. 1, pp. 58–66, 2014.
- [23] N. A. Blondin and D. M. Greer, "Neurologic prognosis in cardiac arrest patients treated with therapeutic hypothermia," *The Neurologist*, vol. 17, no. 5, pp. 241–248, 2011.
- [24] R. Phelps, F. Dumas, C. Maynard, J. Silver, and T. Rea, "Cerebral performance category and long-term prognosis following out-of-hospital cardiac arrest," *Critical Care Medicine*, vol. 41, no. 5, pp. 1252–1257, 2013.
- [25] Y. Baştanlar and M. Ozuysal, "Introduction to machine learning," *Methods in Molecular Biology*, vol. 1107, pp. 105–128, 2014.
- [26] Y. Lee, "Support vector machines for classification: a statistical portrait," *Methods in Molecular Biology*, vol. 620, pp. 347–368, 2010.
- [27] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, San Francisco California USA, 2016.
- [28] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [29] S. Aschauer, G. Dorffner, F. Sterz, A. Erdogmus, and A. Lagner, "A prediction tool for initial out-of-hospital cardiac arrest survivors," *Resuscitation*, vol. 85, no. 9, pp. 1225–1231, 2014.
- [30] S. Nanayakkara, S. Fogarty, M. Tremeer et al., "Characterising risk of in-hospital mortality following cardiac arrest using

machine learning: a retrospective international registry study,” *PLoS Medicine*, vol. 15, no. 11, article e1002709, 2018.

- [31] Z. Zhang, “Variable selection with stepwise and best subset approaches,” *Annals of Translational Medicine*, vol. 4, no. 7, 2016.
- [32] J. M. Larsen and J. Ravkilde, “Acute coronary angiography in patients resuscitated from out-of-hospital cardiac arrest—A systematic review and meta-analysis,” *Resuscitation*, vol. 83, no. 12, pp. 1427–1433, 2012.
- [33] G. Sideris, S. Voicu, D. Yannopoulos et al., “Favourable 5-year postdischarge survival of comatose patients resuscitated from out-of-hospital cardiac arrest, managed with immediate coronary angiogram on admission,” *European Heart Journal Acute Cardiovascular Care*, vol. 3, no. 2, pp. 183–191, 2014.

Research Article

Identification of Novel Choroidal Neovascularization-Related Genes Using Laplacian Heat Diffusion Algorithm

Minjie Sheng ¹, Haiying Cai,¹ Ming Cheng,¹ Jing Li ¹, Jian Zhang ^{2,3,4,5,6}
and Lihua Liu ¹

¹Department of Ophthalmology, Yangpu Hospital, Tongji University School of Medicine, 450 Tengyue Road, Shanghai 200090, China

²Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China

³Shanghai Key Laboratory of Ocular Fundus Diseases, Shanghai 200080, China

⁴Shanghai Engineering Center for Visual Science and Photomedicine, Shanghai 200080, China

⁵National Clinical Research Center for Eye Diseases, Shanghai 20080, China

⁶Shanghai Engineering Center for Precise Diagnosis and Treatment of Eye Diseases, Shanghai 20080, China

Correspondence should be addressed to Jian Zhang; natalieeilatan@126.com and Lihua Liu; lh_liu@yeah.net

Received 9 May 2021; Accepted 20 August 2021; Published 7 September 2021

Academic Editor: Qiushi Zhao

Copyright © 2021 Minjie Sheng 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.

Choroidal neovascularization (CNV) is a type of eye disease that can cause vision loss. In recent years, many studies have attempted to investigate the major pathological processes and molecular pathogenic mechanisms of CNV. Because many diseases are related to genes, the genes associated with CNV need to be identified. In this study, we proposed a network-based approach for identifying novel CNV-associated genes. To execute such method, we first employed a protein-protein interaction network reported in STRING. Then, we applied a network diffusion algorithm, Laplacian heat diffusion, on this network by selecting validated CNV-related genes as the seed nodes. As a result, some novel genes that had unknown but strong relationships with validated genes were identified. Furthermore, we used a screening procedure to extract the most essential genes. Eleven latent CNV-related genes were finally obtained. Extensive analyses were performed to confirm that these genes are novel CNV-related genes.

1. Introduction

Choroidal neovascularization (CNV) is a typical pathogenic process that refers to the abnormal creation of blood vessels specifically in the choroid layer of the eye. As a severe pathogenesis of one subtype of age-related macular degeneration (AMD), CNV can be clinically concomitant with various ocular symptoms such as extreme myopia and malignant myopic degeneration. According to the recent epidemiological statistics provided by *Lancet*, more than 6 million people around the world suffered from AMD in 2015 [1]. Based on another independent survey, the prevalence of CNV-associated AMD was found to be 1.2% of all adults aged 43–86 years [2], indicating that CNV may be one of the major causes of vision loss.

As mentioned above, CNV is a major threat to visual health, especially in elderly people around the world. Therefore, for centuries, scientists have attempted to determine the major pathological processes and molecular pathogenic mechanisms of CNV [3, 4]. However, the detailed and comprehensive mechanisms of CNV have not been fully elucidated. According to existing literatures, the major pathogenic mechanisms of CNV can be attributed to the imbalance of antiangiogenic factors and angiogenic factors [5, 6]. The imbalance of these factors in the choroid may promote vasculogenesis and angiogenesis pathologically related to CNV [6]. In terms of regulators, PEDF (pigment epithelium-derived factor) [6] and VEGF (vascular endothelial growth factor), which are antiangiogenic and typical angiogenic factors, respectively [5] have both been confirmed

to contribute to the initiation and progression of CNV. However, the factors or initiators that drive the abnormal biological functions of PEDF and VEGF have not been confirmed. Hypoxia [7], high glucose [8], protein kinase C activation [9], advanced glycation end products [10], reactive oxygen species [11], activated oncogenes [12], and abnormal cytokine production [13] may all contribute to the pathogenesis and clinical symptoms of CNV.

Although the pathogenesis of the diseases we have discussed is complicated, we can still simply cluster all the potential pathogenic factors into two groups: genetic factors and environmental factors [3, 4]. In this study, we computationally investigated the genetic pathogenesis of CNV. According to recent publications, various genetic factors have been confirmed to contribute to CNV. Abnormal angiogenesis and antiangiogenesis are two major pathogenic processes in such disease [5, 6]. Recent publications revealed that various genes related to angiogenesis and antiangiogenesis may directly participate in the pathogenesis of CNV. *VEGF* [14] and *FGF2* [15] are two typical genes associated with angiogenesis. In 2009, these genes have been confirmed to be related to CNV and regulate its rate of progression [16]. Besides these genes, another functional gene called *CFI*, which is related to extreme myopia, has also been reported to contribute to CNV pathogenesis [17, 18], revealing the complicated genetic basis of CNV. Other functional genes associated with cell proliferation, such as *RELA* [19], *NFKB1* [20, 21], and *RELB* [19], have all been reported to promote abnormal angiogenesis during the initiation and progression of CNV.

For decades, scientists have attempted to reveal the comprehensive genetic background of CNV. However, identifying and validating CNV-associated genes one by one is quite expensive and time consuming. In recent years, with the development of high-throughput sequencing, bioinformatics algorithms have provided us a novel and more effective approach for identifying CNV-associated genes. In 2016, a systematic prediction [22] based on all the identified CNV-related genes, protein-protein interaction (PPI) network, and shortest path algorithm identified various genes associated with CNV, including *ANK1*, *ITGA4*, and *CD44*. Most of these genes have already been identified to contribute to abnormal angiogenesis or antiangiogenesis in the choroid [22], validating the efficacy and accuracy of computational prediction on disease-associated genes. Therefore, in this study, we introduced a novel computational method called *Laplacian heat diffusion* (LHD) [23] to further explore the pathogenic factors of CNV. This study not only identified potential CNV-associated genes but also revealed the detailed pathogenesis of CNV.

2. Materials and Methods

2.1. CNV-Associated Genes. Genes associated with AMD were first obtained from a previous study [24]. In detail, we downloaded the “Additional file 3” in such study, which contained these genes. Then, according to “Additional file 5”, genes in CNV up or CNV down modules were picked up, accessing 37 CNV-associated genes (Table S1). These

genes were further converted to Ensembl gene IDs to be consistent with the protein IDs in the PPI network from the STRING database [25]. These genes comprised a seed gene set S .

2.2. PPIs. In general, proteins interact with each other to regulate biological process; thus, they share similar biological functions. Based on this assumption, many studies have been devoted to infer protein functions. Therefore, potential CNV-associated genes can be identified from the known CNV-associated genes and their interaction network.

We downloaded 4,274,001 human protein-protein interactions (PPIs) for 19,247 proteins from STRING (<https://www.string-db.org/>, version 10) [26]. These interactions were derived from genomic context predictions, high-throughput lab experiments, (conserved) coexpression, automated text mining, and previous knowledge in databases. Thus, PPIs reported in STRING can widely measure the associations of proteins compared with those in some other databases [27, 28], in which PPIs were only determined by solid experiments. For each PPI, both proteins are represented by Ensembl IDs, and a score ranging from 150 to 999 is assigned. A high score indicates that the corresponding interaction is supported by high-quality evidence. The interaction score between two proteins (P_1 and P_2) was denoted as $I(P_1, P_2)$. Using the abovementioned data, we can construct a PPI network consisting of 19,247 nodes and 4,274,001 edges, which connects two nodes with interaction score as the weight if and only if two proteins interact. The PPI network is denoted as G . Such PPI network has been widely used in many researches [29–37].

2.3. Laplacian Heat Diffusion. Nowadays, network methods are more and more popular to deal with different biological and medical problems [30, 32, 36, 38–41]. This study also adopted a powerful network method, LHD algorithm. As a type of network diffusion method, heat diffusion follows some rules to transmit heat on the seed nodes to surrounding nodes in the network. The heat on a node indicates its connections to seed nodes. In this study, the LHD algorithm [23] was applied to search for novel CNV-related genes, which was a heat diffusion process on a Laplacian matrix constructed from protein-protein network.

Given a PPI network G , we can first construct its adjacent matrix A based on the edge weights. Then, we normalize it column wisely as follows:

$$A'[i, j] = \frac{A[i, j]}{\sum_{k=1}^n A[k, j]}, \quad (1)$$

where i is the column index of 19,247 nodes in G and j is the index of CNV-related genes. Each column in A' was a 19,247-dimensional vector. Each element was the heat of a node in the network G . Initially, the component in A' corresponding to 37 CNV-related genes was configured to be $1/37$; other components were set to 0. Then, the values of each vector were updated as follows:

$$H_i[i] = H_0[i] \exp(-\lambda_i t), \quad (2)$$

where H_t is the heat distribution at time t and λ_i is the i th eigenvalue of matrix A' . We updated the vectors until the heat distribution vectors at two consecutive time points change as small as a defined threshold. After the diffusion process, each node was assigned a heat value. A larger heat value indicates that the node is more important. Thus, we selected nodes with heat values greater than the defined cut-off and mapped those nodes back to the corresponding genes.

In this study, we used the LHD algorithm (<https://CRAN.R-project.org/package=diffusr>) to perform the analysis with default parameters on the PPI network G .

2.4. Postprocessing of CNV-Related Candidate Genes. According to the LHD-based method, we can obtain a large number of candidate CNV-related genes. However, some of them are essential genes, while others are nonessential genes. A three-stage method was applied to select the essential genes by integrating other biological information: (1) Z -score based on permutation test to exclude false positives, (2) maximum interaction score (MIS) based on PPI information to exclude genes with few connections to the validated CNV-related genes, and (3) maximum function score (MFS) based on biological function annotation information to filter functional genes.

2.4.1. Z -Score. To evaluate the significance of the produced heat values, we randomly sampled 1000 gene sets and calculated the mean and standard deviation of these heat values. Then, we calculated the Z -score for all CNV-related candidate genes. In detail, 1000 gene sets with a size of 37 were randomly generated. For each gene set, we performed the LHD algorithm on the PPI network G by using it as the seed set. Then, each gene g was assigned a heat value. The above process was run for the produced 1000 gene sets. Each g received 1000 heat values and a real heat value based on 37 validated CNV-related genes. We calculated the measurement Z -score as follows:

$$Z\text{-score}(g) = \frac{h - \bar{h}}{\text{sd}}, \quad (3)$$

where h is the real heat value of gene g and \bar{h} and sd are the mean and standard deviation of 1000 heat values of the 1000 randomly produced gene sets, respectively. The higher the Z -score of one gene is, the more likely it is a real CNV-related gene. In this study, we selected genes with Z -score greater than 1.96.

2.4.2. MIS. After the permutation test, some CNV-related candidate genes were further verified to have strong associations with the validated CNV-related genes. In general, interacting proteins always exhibit similar functions. Based on this observation, we calculated MIS as follows:

$$\text{MIS}(g) = \max \left\{ I(g, g') \mid g' \text{ is a validated CNV-related gene} \right\}, \quad (4)$$

where $I(g, g')$ is the interaction score between two genes from the STRING database. A high MIS value indicates that this gene is strongly connected to the validated CNV-related genes; thus, it is more likely to be true CNV-associated gene. Here, we set a threshold of 900 (the highest confidence score in the STRING database) to filter out genes with low MIS values.

2.4.3. MFS. To be CNV-related genes, they must highly contribute to certain biological processes involved in CNV. To further select more reliable CNV-related candidate genes, Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) [42] pathways were used. We extracted important candidate CNV-related genes with similar GO terms and KEGG pathways to validate CNV-related genes. The enrichment theory [43, 44] was applied to estimate the relationships between genes and GO/KEGG pathways. It encodes a gene as a vector. The relationship between two genes can be calculated as follows:

$$Q(g, g') = \frac{E(g) \cdot E(g')}{\|E(g)\| \cdot \|E(g')\|}, \quad (5)$$

where $E(g)$ is the column vector obtained according to enrichment theory.

Similarly, for each gene, $\text{MFS}(g)$ was calculated as follows:

$$\text{MFS}(g) = \max \left\{ Q(g, g') \mid g' \text{ is a validated CNV-related gene} \right\}. \quad (6)$$

The higher the MFS of one gene is, the more GO/KEGG pathways it shares with the validated CNV-related genes. The final candidate CNV-related genes were extracted with an MFS value greater than a defined cutoff of 0.9.

3. Results

In this study, we presented a computational approach to infer novel CNV-associated genes using the LHD-based method. The entire procedures are illustrated in Figure 1. This approach collected verified CNV-related genes, which were extrapolated to identify novel candidate genes on the PPI network using Laplacian heat diffusion. Next, these identified candidate genes were further screened to filter out false positive genes that are not associated with any CNV-related biological process.

We first selected genes with a heat value $> e - 10$, and a total of 19,218 genes were obtained. Then, these genes were evaluated using the permutation test with 1000 randomly generated sets. We selected genes with a Z -score greater than 1.96 and obtained a list of 153 genes. We further filtered out genes with fewer connections to the validated CNV-related genes by MIS score. We kept genes with MIS value greater than 900 and obtained 27 genes. Finally, for each of the 27 genes, we calculated the MFS and selected genes with an MFS value greater than 0.9, resulting in a final list of 11

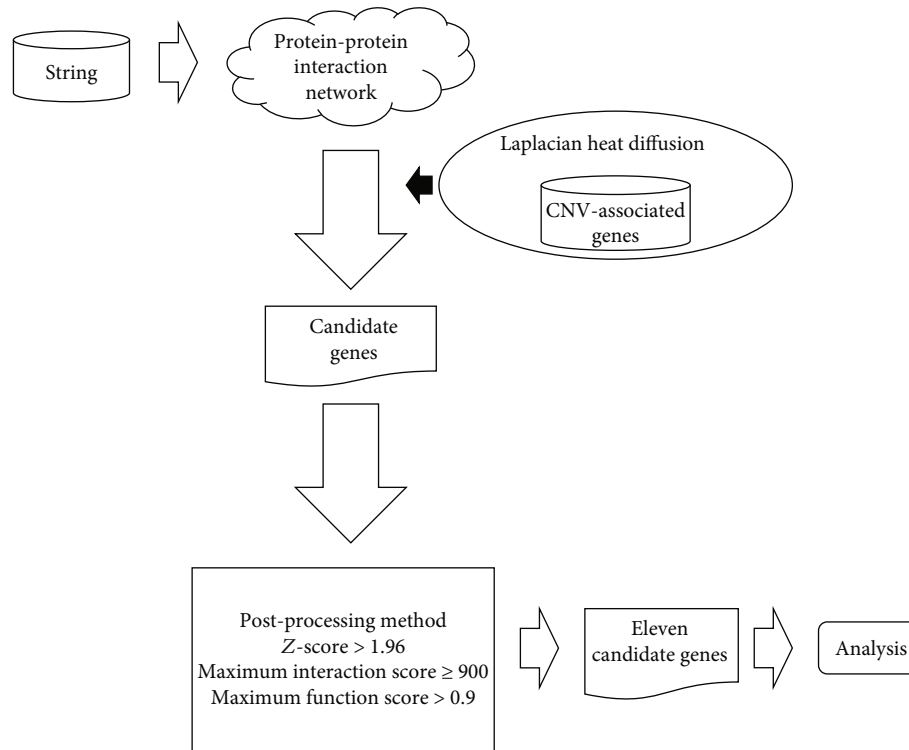


FIGURE 1: Entire procedures to identify novel choroidal neovascularization- (CNV-) related genes. A protein-protein interaction network reported in STRING is employed. Laplacian heat diffusion (LHD) with validated CNV-related genes as seed nodes is applied to such network for extracting raw candidate genes. They are further filtered by a postprocessing method, resulting in eleven candidate genes. These genes were extensively analyzed.

CNV-related genes, which is totally different from previous discoveries [22]. The selected numbers of putative genes in different steps of LHD are shown in Table 1, and the detailed information of the 11 final candidate CNV-associated genes is listed in Table 2. The interaction network between the 11 candidate genes and 37 verified CNV-related genes is shown in Figure 2. All measurements mentioned above are listed in sheets 1–4 of Table S2.

4. Discussion

As we have analyzed above, we applied a novel computational method named *Laplacian heat diffusion* [45] to identify potential CNV-related genes based on the existing PPI network provided in STRING [25]. According to such algorithm and the database, we screened out eleven functional genes that may directly or indirectly participate in the pathogenesis of CNV. To validate the efficacy and accuracy of our newly applied computational method, we performed a systematic datamining on the biological functions and CNV relevance of all predicted genes. The predicted genes have been validated by recent publications. The detailed analysis on each gene can be seen below. For a clear description, we classified these genes into some classes, which is illustrated in Figure 3.

4.1. Matrix Metalloproteinases (MMPs). *MMP3* (ENSP0000299855), which ranks the highest in the prediction list, has been predicted to be related to the pathogenesis of

TABLE 1: Number of candidate CNV-related genes in different stages of LHD-based method.

Method	Network diffusion algorithm	Z-score	MIS	MFS
LHD-based method	19,218	153	27	11

CNV. Generally, it has been widely reported to contribute to the activation of procollagenase [46] and matrix remodeling [47]. In terms of its potential pathogenic functions in CNV, this gene has been confirmed to act abnormally in the choroidal neovascular membranes, implying its pathogenic potential [48]. Further studies on the contribution of hypoxia to CNV confirmed that our predicted gene *MMP3* may contribute to hypoxia-induced apoptosis and secretion of proangiogenic factors in the choroid under hypoxia microenvironment, which further initiates CNV [49]. Therefore, our predicted gene *MMP3* may functionally be a potential driving factor for CNV, demonstrating the accuracy of our prediction result. Apart from *MMP3*, three other components of the MMP family, namely, *MMP13* (ENSP00000260302), *MMP7* (ENSP00000260227), and *MMP10* (ENSP00000279441), have also been predicted to contribute to the pathogenesis of CNV in our prediction list with a high rank. With similar biological functions as *MMP3*, all of such three genes (*MMP13*, *MMP7*, and *MMP10*) have been reported to participate in the abnormal

TABLE 2: Eleven candidate genes yielded by LHD-based method.

Ensemble ID	Gene symbol	Description	Heat	Z-score	MIS	MES
ENSP00000299855	MMP3	Matrix metalloproteinase 3	$8.79E - 05$	2.0806	999	0.9761
ENSP00000230990	HBEGF	Heparin-binding EGF-like growth factor	$9.01E - 05$	3.0964	989	0.9608
ENSP00000260302	MMP13	Matrix metalloproteinase 13	$1.54E - 04$	3.9922	964	0.9600
ENSP00000260227	MMP7	Matrix metalloproteinase 7	$1.25E - 04$	3.8042	975	0.9580
ENSP00000262768	TIMP2	TIMP metalloproteinase inhibitor 2	$1.20E - 04$	3.2461	994	0.9569
ENSP00000222390	HGF	Hepatocyte growth factor	$1.09E - 04$	4.0224	922	0.9487
ENSP00000304408	COL3A1	Collagen type III alpha 1 chain	$1.40E - 04$	2.6753	951	0.9461
ENSP00000279441	MMP10	Matrix metalloproteinase 10	$1.31E - 04$	3.2940	977	0.9186
ENSP00000347665	COL18A1	Collagen type XVIII alpha 1 chain	$1.25E - 04$	2.5902	991	0.9183
ENSP00000277480	LCN2	Lipocalin 2	$9.96E - 05$	2.1251	985	0.9141
ENSP00000297904	VEGFD	Vascular endothelial growth factor D	$2.13E - 04$	6.6122	939	0.9078

angiogenesis of choroidal tissues, validating their specific contribution to CNV. In 2011, a study [50] on CNV in a mouse model confirmed that the deficiency of *MMP13* contributed to the impairment of neovascularization formation in choroid tissues, and such pathogenesis could be restored by injecting mesenchymal cells secreting *MMP13*, validating the specific role of this gene during CNV initiation and progression. As for *MMP7*, basal laminar and linear deposits are typical complications of CNV, contributing to the constitution of the CNV microenvironment [51, 52]. A recent study [53] on the typical basal laminar and linear deposits of CNV confirmed that *MMP7* together with its homologue *MMP13* may contribute to CNV by regulating the inflammatory processes in the microenvironment of choroidal tissues. Furthermore, *MMP10* has also been validated by recent publications. Although no reports connected *MMP10* and CNV directly, the specific contribution of all metalloproteinases including *MMP10* on choroidal microenvironment remodeling and inflammation mediation implies the specific biological function of *MMP10* during the progression of CNV [54].

4.2. Growth Factors. *HBEGF* (ENSP00000230990) has also been predicted to contribute to the progression of CNV. As a typical growth factor, *HBEGF* participates in the ERBB2 signaling pathway and interacts with functional genes such as EGFR and ERBB4 [55, 56]. A recent study confirmed that *HBEGF* may affect the production and biological functions of *VEGF* in CNV [57]. Therefore, although no direct reports confirmed the detailed biological function of *HBEGF* in CNV, this gene may interact with *VEGF* and play a crucial pathogenic role during the progression of CNV. Another functional growth factor encoding gene *HGF* (ENSP00000222390) has also been predicted to contribute to the pathogenesis of CNV. Generally, the binding of *HGF* to its target receptor (hepatocyte growth factor) contributes to the regulation of cell growth, cell motility, and morphogenesis in various cell and tissue subtypes [58, 59]. As for its unique pathogenic contribution to CNV, a paired experimental study [60] on CNV confirmed that compared with normal tissues, the pathogenic tissues of the choroid during CNV

initiation and progression have different expression profiling of growth factors including *VEGF*, *HGF*, and *FGF*, implying the potential pathogenic role of *HGF* in such disease. In 2011, a specific study on the biological and pathogenic functions of cytokines in CNV confirmed that *HGF* has a mitogenic effect on choroidal cells, promoting neovascularization processes [61]. Therefore, such gene may be a potential CNV-associated gene. As the next predicted growth factor in the predicted list of genes, *VEGFD* (ENSP00000297904) has been widely reported to be a member of the platelet-derived growth factor family. This gene has been reported to promote angiogenesis [62], lymph angiogenesis [62], and endothelial cell growth [63]. As the homologue of the identified key driver gene of CNV (*VEGF*) generated by differential alternative splicing, *VEGFD* directly participates in the pathogenesis of CNV, regulating the same biological processes of *VEGF* [64]. Recent clinical studies [65] confirmed that *VEGFD* may also be a candidate marker for the diagnosis and treatment of CNV, and drugs that target *VEGF* to relieve symptoms may also target the products of *VEGFD*.

4.3. MMP Inhibitors. *TIMP2* (ENSP00000262768) has been widely reported to act as a natural inhibitor for MMPs [66]. With a specific expression pattern in vitreous and sub-retinal fluid, this gene has been found to be expressed in choroid tissues [67] and directly contribute to the activation of the hypoxia-induced VEGF signaling pathway and MMP regulation [68]. Considering the irreplaceable role of *VEGF* in CNV, *TIMP2* may be a potential CNV-associated gene.

4.4. Collagens. Based on our newly presented computational methods, we also obtained two collagen coding genes that may contribute to the pathogenesis of CNV, namely, *COL3A1* (ENSP00000304408) and *COL18A1* (ENSP00000347665). *COL3A1* encodes the pro-alpha1 chain of type III collagen, a fibrillary collagen. Based on existing literatures, this gene contributes to the regulation of cortical development together with type I collagen in soft connective tissues [69, 70]. As for its specific pathogenic contribution to CNV, a specific study [71] confirmed that *COL3A1* may

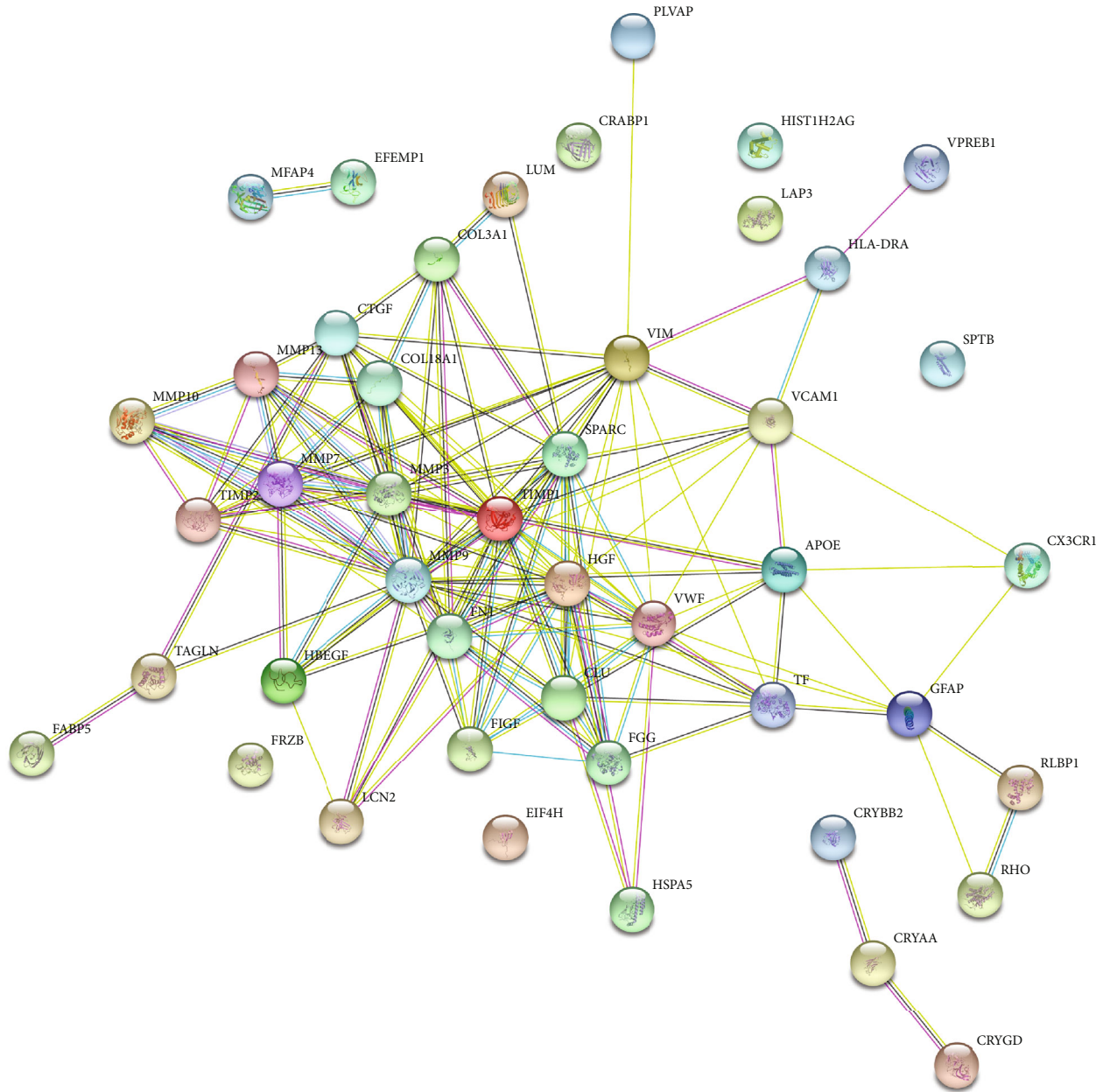


FIGURE 2: Interaction network between the 11 candidate genes and verified 37 CNV-related genes.

contribute to actin cytoskeleton remodeling and affect the specific lesion size and fibrosis of CNV. Similarly, the next predicted gene (*COL18A1*) has also been reported to participate in collagen-associated CNV pathogenesis [72]. Currently, no direct pathogenic experiment has confirmed that *COL18A1* can induce the progression of CNV independently. Other studies on collagen families including collagen XVIII [48, 73, 74] in CNV and their respective angiogenic functions have validated the potential pathogenic role of our predicted collagen encoding genes.

4.5. Lipocalins. Apart from MMPs, collagen, and growth factor-associated genes, we also obtained a specific lipocalin encoding gene, namely, *LCN2* (ENSP00000277480). Generally, this gene has been identified in the lung, breast [75],

and eye secretions [76] and contribute to the transport of hydrophobic ligands [77]. As for its specific contribution to CNV, this gene may promote angiogenesis and neovascularization under pathogenic conditions [78, 79]. With a high-expression pattern in choroid tissues [80] and its interaction with MMPs [81], *LCN2* has been confirmed to participate in the pathogenic activation of the *AKT2-NF- κ B-lipocalin-2* axis in CNV [82].

Taken together, the predicted functional genes are enriched in MMP-, growth factor-, collagen-, and lipocalin-related genes, implying the specific role of such components during the initiation and progression of CNV. The predicted genes have all been confirmed by recent publications as we have described above. Therefore, the computational approach in this study may be quite effective and accurate for

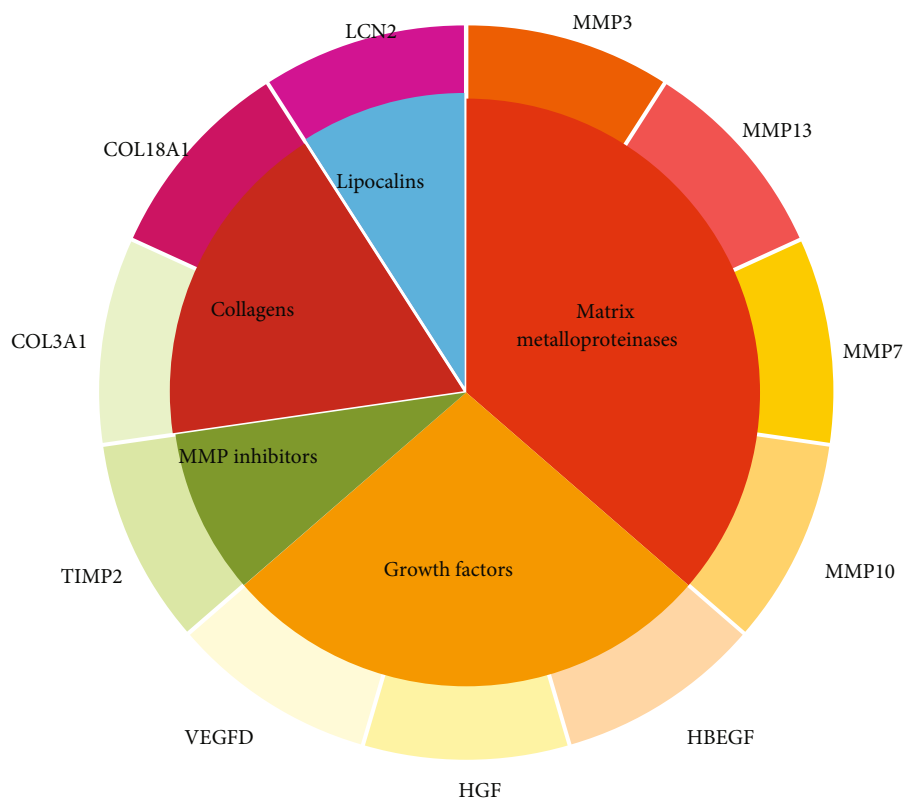


FIGURE 3: Classification of eleven candidate CNV-associated genes.

identifying CNV-associated genes. This study not only identified a group of functional CNV-associated genes and potential related biological processes but also contributed to the improvement of current computational prediction approaches on the genetic background of diseases.

5. Conclusions

This study employed a powerful network diffusion method to identify possible CNV-related genes in a PPI network. To obtain reliable genes, a three-stage method followed to screen out key latent CNV-related genes. The analysis on final obtained genes indicate that they can be novel CNV-related genes with high likelihood. It is hopeful that the new findings reported in this study can provide new insights for investigating CNV.

Data Availability

The data used to support the findings of this study are included within the supplementary information files.

Conflicts of Interest

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

Authors' Contributions

Minjie Sheng and Haiying Cai contributed equally to this work.

Supplementary Materials

Table S1: curated CNV-associated genes. Table S2: measurements of genes in each step. (*Supplementary Materials*)

References

- [1] G. B. D. Disease, I. Injury, and C. Prevalence, "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015," *Lancet*, vol. 388, no. 10053, pp. 1545-1602, 2016.
- [2] R. Klein, B. E. Klein, and K. L. Linton, "Prevalence of Age-related Maculopathy: The Beaver Dam Eye Study," *Ophthalmology*, vol. 99, no. 6, pp. 933-943, 1992.
- [3] E. D'Ambrosio, P. Tortorella, and L. Iannetti, "Management of uveitis-related choroidal neovascularization: from the pathogenesis to the therapy," *Journal of Ophthalmology*, vol. 2014, Article ID 450428, 6 pages, 2014.
- [4] D. V. Do, "Detection of new-onset choroidal neovascularization," *Current Opinion in Ophthalmology*, vol. 24, no. 3, pp. 244-247, 2013.
- [5] K. Kinnunen and S. Yla-Herttuala, "Vascular endothelial growth factors in retinal and choroidal neovascular diseases," *Annals of Medicine*, vol. 44, no. 1, pp. 1-17, 2012.

- [6] J. P. Tong and Y. F. Yao, "Contribution of VEGF and PEDF to choroidal angiogenesis: a need for balanced expressions," *Clinical Biochemistry*, vol. 39, no. 3, pp. 267–276, 2006.
- [7] S. Takata, T. Masuda, S. Nakamura et al., "The effect of triamcinolone acetonide on laser-induced choroidal neovascularization in mice using a hypoxia visualization bio-imaging probe," *Scientific Reports*, vol. 5, no. 1, p. 9898, 2015.
- [8] Y. Cai, X. Li, Y. S. Wang et al., "Hyperglycemia promotes vasculogenesis in choroidal neovascularization in diabetic mice by stimulating VEGF and SDF-1 expression in retinal pigment epithelial cells," *Experimental Eye Research*, vol. 123, pp. 87–96, 2014.
- [9] Y. Saishin, R. L. Silva, Y. Saishin et al., "Periocular injection of microspheres containing PKC412 inhibits choroidal neovascularization in a porcine model," *Investigative Ophthalmology & Visual Science*, vol. 44, no. 11, pp. 4989–4993, 2003.
- [10] L. Sun, T. Huang, W. Xu, J. Sun, Y. Lv, and Y. Wang, "Advanced glycation end products promote VEGF expression and thus choroidal neovascularization via Cyr61-PI3K/AKT signaling pathway," *Scientific Reports*, vol. 7, no. 1, p. 14925, 2017.
- [11] R. S. Eshaq, W. S. Wright, and N. R. Harris, "Oxygen delivery, consumption, and conversion to reactive oxygen species in experimental models of diabetic retinopathy," *Redox Biology*, vol. 2, pp. 661–666, 2014.
- [12] K. Nochioka, H. Okuda, K. Tatsumi, S. Morita, N. Ogata, and A. Wanaka, "Hedgehog signaling components are expressed in choroidal neovascularization in laser-induced retinal lesion," *Acta Histochemica et Cytochemica*, vol. 49, no. 2, pp. 67–74, 2016.
- [13] H. Yin, X. Fang, J. Ma et al., "Idiopathic choroidal neovascularization: intraocular inflammatory cytokines and the effect of intravitreal ranibizumab treatment," *Scientific Reports*, vol. 6, no. 1, p. 31880, 2016.
- [14] Y. Zhang, Q. Han, Y. Ru, Q. Bo, and R. H. Wei, "Anti-VEGF treatment for myopic choroid neovascularization: from molecular characterization to update on clinical application," *Drug Design, Development and Therapy*, vol. 9, pp. 3413–3421, 2015.
- [15] J. Kusari, E. Padillo, S. X. Zhou et al., "Effect of brimonidine on retinal and choroidal neovascularization in a mouse model of retinopathy of prematurity and laser-treated rats," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 8, pp. 5424–5431, 2011.
- [16] K. Nakai, M. S. Rogers, T. Baba et al., "Genetic loci that control the size of laser-induced choroidal neovascularization," *FASEB Journal*, vol. 23, no. 7, pp. 2235–2243, 2009.
- [17] K. Fang, P. Gao, J. Tian et al., "Joint effect of CFH and ARMS2/HTRA1 polymorphisms on neovascular age-related macular degeneration in Chinese population," *Journal of Ophthalmology*, vol. 2015, Article ID 821918, 8 pages, 2015.
- [18] C. Saade, B. Ganti, M. Marmor, K. B. Freund, and R. T. Smith, "Risk characteristics of the combined geographic atrophy and choroidal neovascularisation phenotype in age-related macular degeneration," *The British Journal of Ophthalmology*, vol. 98, no. 12, pp. 1729–1732, 2014.
- [19] K. Izumi-Nagai, N. Nagai, K. Ohgami et al., "Macular pigment lutein is antiinflammatory in preventing choroidal neovascularization," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 27, no. 12, pp. 2555–2562, 2007.
- [20] M. Hirasawa, K. Takubo, H. Osada et al., "Angiopoietin-like Protein 2 Is a Multistep Regulator of Inflammatory Neovascularization in a Murine Model of Age-related Macular Degeneration," *The Journal of Biological Chemistry*, vol. 291, no. 14, pp. 7373–7385, 2016.
- [21] L. Paneghetti and Y. S. Ng, "A novel endothelial-derived anti-inflammatory activity significantly inhibits spontaneous choroidal neovascularisation in a mouse model," *Vascular Cell*, vol. 8, no. 1, p. 2, 2016.
- [22] J. Zhang, Y. Suo, Y. H. Zhang et al., "Mining for genes related to choroidal neovascularization based on the shortest path algorithm and protein interaction information," *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1860, no. 11, pp. 2740–2749, 2016.
- [23] D. E. Carlin, B. Demchak, D. Pratt, E. Sage, and T. Ideker, "Network propagation in the cytoscape cyberinfrastructure," *PLoS Computational Biology*, vol. 13, no. 10, article e1005598, 2017.
- [24] A. M. Newman, N. B. Gallo, L. S. Hancox et al., "Systems-level analysis of age-related macular degeneration reveals global biomarkers and phenotype-specific functional networks," *Genome Medicine*, vol. 4, no. 2, p. 16, 2012.
- [25] D. Szklarczyk, A. Franceschini, S. Wyder et al., "STRING v10: protein-protein interaction networks, integrated over the tree of life," *Nucleic Acids Research*, vol. 43, no. D1, pp. D447–D452, 2015.
- [26] D. Szklarczyk, A. Franceschini, S. Wyder et al., "STRING v10: protein-protein interaction networks, integrated over the tree of life," *Nucleic Acids Research*, vol. 43, no. D1, pp. D447–D452, 2014.
- [27] I. Xenarios, D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte, and D. Eisenberg, "DIP: the database of interacting proteins," *Nucleic Acids Research*, vol. 28, no. 1, pp. 289–291, 2000.
- [28] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "BioGRID: a general repository for interaction datasets," *Nucleic Acids Research*, vol. 34, no. 9, pp. D535–D539, 2006.
- [29] J. Gao, B. Hu, and L. Chen, "A path-based method for identification of protein phenotypic annotations," *Current Bioinformatics*, 2021.
- [30] H. Liu, B. Hu, L. Chen, and L. Lu, "Identifying protein subcellular location with embedding features learned from networks," *Current Proteomics*, vol. 17, 2021.
- [31] X. Zhang, L. Chen, Z. H. Guo, and H. Liang, "Identification of human membrane protein types by incorporating network embedding methods," *IEEE Access*, vol. 7, pp. 140794–140805, 2019.
- [32] X. L. Zhang and L. Chen, "Prediction of membrane protein types by fusing protein-protein interaction and protein sequence information," *Biochimica Et Biophysica Acta-Proteins and Proteomics*, vol. 1868, no. 12, article 140524, 2020.
- [33] Y. Zhang, T. Zeng, L. Chen, S. J. Ding, T. Huang, and Y. D. Cai, "Identification of COVID-19 infection-related human genes based on a random walk model in a virus-human protein interaction network," *BioMed Research International*, vol. 2020, Article ID 4256301, 7 pages, 2020.
- [34] X. Pan, H. Li, T. Zeng et al., "Identification of protein subcellular localization with network and functional embeddings," *Frontiers in Genetics*, vol. 11, article 626500, 2021.
- [35] Y. H. Zhang, T. Zeng, L. Chen, T. Huang, and Y. D. Cai, "Determining protein-protein functional associations by functional rules based on gene ontology and KEGG pathway,"

- Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1869, no. 6, p. 140621, 2021.
- [36] R. Zhao, L. Chen, B. Zhou, Z. H. Guo, S. Wang, and Aorigele, "Recognizing novel tumor suppressor genes using a network machine learning strategy," *IEEE Access*, vol. 7, pp. 155002–155013, 2019.
- [37] R. Zhao, B. Hu, L. Chen, and B. Zhou, "Identification of latent oncogenes with a network embedding method and random forest," *BioMed Research International*, vol. 2020, Article ID 5160396, 11 pages, 2020.
- [38] H. Liang, L. Chen, X. Zhao, and X. Zhang, "Prediction of drug side effects with a refined negative sample selection strategy," *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 1573543, 16 pages, 2020.
- [39] H. Y. Liang, B. Hu, L. Chen, S. Wang, and Aorigele, "Recognizing novel chemicals/drugs for anatomical therapeutic chemical classes with a heat diffusion algorithm," *Biochimica et Biophysica Acta-Molecular Basis of Disease*, vol. 1866, no. 11, article 165910, 2020.
- [40] Y. Zhu, B. Hu, L. Chen, and Q. Dai, "iMPTCE-Hnetwork: A Multilabel Classifier for Identifying Metabolic Pathway Types of Chemicals and Enzymes with a Heterogeneous Network," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 6683051, 12 pages, 2021.
- [41] J.-P. Zhou, L. Chen, and Z.-H. Guo, "iATC-NRAKEL: an efficient multi-label classifier for recognizing anatomical therapeutic chemical classes of drugs," *Bioinformatics*, vol. 36, no. 5, pp. 1391–1396, 2020.
- [42] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [43] J. Yang, L. Chen, X. Kong, T. Huang, and Y. D. Cai, "Analysis of tumor suppressor genes based on Gene Ontology and the KEGG Pathway," *PLoS One*, vol. 9, no. 9, article e107202, 2014.
- [44] P. Carmona-Saez, M. Chagoyen, F. Tirado, J. M. Carazo, and A. Pascual-Montano, "GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists," *Genome Biology*, vol. 8, no. 1, p. R3, 2007.
- [45] S. Köhler, S. Bauer, D. Horn, and P. N. Robinson, "Walking the interactome for prioritization of candidate disease genes," *The American Journal of Human Genetics*, vol. 82, no. 4, pp. 949–958, 2008.
- [46] Z. Yu, R. Visse, M. Inouye, H. Nagase, and B. Brodsky, "Defining Requirements for Collagenase Cleavage in Collagen Type III Using a Bacterial Collagen System," *The Journal of Biological Chemistry*, vol. 287, no. 27, pp. 22988–22997, 2012.
- [47] X. Ji, L. Wang, B. Wu et al., "Associations of MMP1, MMP2 and MMP3 genes polymorphism with coal workers' pneumoconiosis in Chinese Han population," *International Journal of Environmental Research and Public Health*, vol. 12, no. 11, pp. 13901–13912, 2015.
- [48] B. Steen, S. Sejersen, L. Berglin, S. Seregard, and A. Kvanta, "Matrix metalloproteinases and metalloproteinase inhibitors in choroidal neovascular membranes," *Investigative Ophthalmology & Visual Science*, vol. 39, no. 11, pp. 2194–2200, 1998.
- [49] J. Zhang, J. Zhao, Y. Bai, L. Huang, W. Yu, and X. Li, "Effects of p75 neurotrophin receptor on regulating hypoxia-induced angiogenic factors in retinal pigment epithelial cells," *Molecular and Cellular Biochemistry*, vol. 398, no. 1-2, pp. 123–134, 2015.
- [50] J. Lecomte, K. Louis, B. Detry et al., "Bone marrow-derived mesenchymal cells and MMP13 contribute to experimental choroidal neovascularization," *Cellular and Molecular Life Sciences*, vol. 68, no. 4, pp. 677–686, 2011.
- [51] S. Sarks, S. Cherepanoff, M. Killingsworth, and J. Sarks, "Relationship of basal laminar deposit and membranous debris to the clinical presentation of early age-related macular degeneration," *Investigative Ophthalmology & Visual Science*, vol. 48, no. 3, pp. 968–977, 2007.
- [52] V. K. Katsi, M. E. Marketou, D. A. Vrachatis et al., "Essential hypertension in the pathogenesis of age-related macular degeneration: a review of the current evidence," *Journal of Hypertension*, vol. 33, no. 12, pp. 2382–2388, 2015.
- [53] A. Lommatzsch, P. Hermans, K. D. Müller, N. Bornfeld, A. C. Bird, and D. Pauleikhoff, "Are low inflammatory reactions involved in exudative age-related macular degeneration? Morphological and immunohistochemical analysis of AMD associated with basal deposits," *Graefes' Archive for Clinical and Experimental Ophthalmology*, vol. 246, no. 6, pp. 803–810, 2008.
- [54] M. Singh and S. C. Tyagi, "Metalloproteinases as mediators of inflammation and the eyes: molecular genetic underpinnings governing ocular pathophysiology," *International Journal of Ophthalmology*, vol. 10, no. 8, pp. 1308–1318, 2017.
- [55] K. J. Thornton, E. Kamange-Sollo, M. E. White, and W. R. Dayton, "Role of G protein-coupled receptors (GPCR), matrix metalloproteinases 2 and 9 (MMP2 and MMP9), heparin-binding epidermal growth factor-like growth factor (hbEGF), epidermal growth factor receptor (EGFR), erbB2, and insulin-like growth factor 1 receptor (IGF-1R) in trenbolone acetate-stimulated bovine satellite cell proliferation," *Journal of Animal Science*, vol. 93, no. 9, pp. 4291–4301, 2015.
- [56] R. Roskoski Jr., "The ErbB/HER family of protein-tyrosine kinases and cancer," *Pharmacological Research*, vol. 79, pp. 34–74, 2014.
- [57] K. Nakai, K. Yoneda, T. Moriue, J. Igarashi, H. Kosaka, and Y. Kubota, "HB-EGF-induced VEGF production and eNOS activation depend on both PI3 kinase and MAP kinase in HaCaT cells," *Journal of Dermatological Science*, vol. 55, no. 3, pp. 170–178, 2009.
- [58] H. Ota, A. Itaya-Hironaka, A. Yamauchi et al., "Pancreatic β cell proliferation by intermittent hypoxia via up-regulation of Reg family genes and HGF gene," *Life Sciences*, vol. 93, no. 18-19, pp. 664–672, 2013.
- [59] C. C. Chang, J. J. Chiu, S. L. Chen et al., "Activation of HGF/c-Met signaling by ultrafine carbon particles and its contribution to alveolar type II cell proliferation," *American Journal of Physiology. Lung Cellular and Molecular Physiology*, vol. 302, no. 8, pp. L755–L763, 2012.
- [60] W. Hu, M. H. Criswell, S. L. Fong et al., "Differences in the temporal expression of regulatory growth factors during choroidal neovascular development," *Experimental Eye Research*, vol. 88, no. 1, pp. 79–91, 2009.
- [61] J. R. de Oliveira Dias, E. B. Rodrigues, M. Maia, O. Magalhaes, F. M. Penha, and M. E. Farah, "Cytokines in neovascular age-related macular degeneration: fundamentals of targeted combination therapy," *The British Journal of Ophthalmology*, vol. 95, no. 12, pp. 1631–1637, 2011.
- [62] N. I. Bower, A. J. Vogrin, L. le Guen et al., "VEGFD modulates both angiogenesis and lymphangiogenesis during zebrafish embryonic development," *Development*, vol. 144, no. 3, pp. 507–518, 2017.

- [63] T. Duong, K. Koltowska, C. Pichol-Thievend et al., "VEGFD regulates blood vascular development by modulating SOX18 activity," *Blood*, vol. 123, no. 7, pp. 1102–1112, 2014.
- [64] R. M. Hussain and T. A. Ciulla, "Emerging vascular endothelial growth factor antagonists to treat neovascular age-related macular degeneration," *Expert Opinion on Emerging Drugs*, vol. 22, no. 3, pp. 235–246, 2017.
- [65] T. Cabral, L. H. Lima, L. G. M. Mello et al., "Bevacizumab injection in patients with neovascular age-related macular degeneration increases angiogenic biomarkers," *Ophthalmol Retina*, vol. 2, no. 1, pp. 31–37, 2018.
- [66] K. Nakamura, K. Shinozuka, and N. Yoshikawa, "Anticancer and antimetastatic effects of cordycepin, an active component of *Cordyceps sinensis*," *Journal of Pharmacological Sciences*, vol. 127, no. 1, pp. 53–56, 2015.
- [67] T. Matsuo, Y. Okada, F. Shiraga, and T. Yanagawa, "TIMP-1 and TIMP-2 levels in vitreous and subretinal fluid," *Japanese Journal of Ophthalmology*, vol. 42, no. 5, pp. 377–380, 1998.
- [68] P. Ottino, J. Finley, E. Rojo et al., "Hypoxia activates matrix metalloproteinase expression and the VEGF system in monkey choroid-retinal endothelial cells: involvement of cytosolic phospholipase A2 activity," *Molecular Vision*, vol. 10, pp. 341–350, 2004.
- [69] S. J. Jeong, S. Li, R. Luo, N. Strokes, and X. Piao, "Loss of Col3a1, the gene for Ehlers-Danlos syndrome type IV, results in neocortical dyslamination," *PLoS One*, vol. 7, no. 1, article e29767, 2012.
- [70] D. Horn, E. Siebert, U. Seidel et al., "Biallelic COL3A1 mutations result in a clinical spectrum of specific structural brain anomalies and connective tissue abnormalities," *American Journal of Medical Genetics. Part A*, vol. 173, no. 9, pp. 2534–2538, 2017.
- [71] S. Caballero, R. Yang, M. B. Grant, and B. Chaqour, "Selective blockade of cytoskeletal actin remodeling reduces experimental choroidal neovascularization," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 5, pp. 2490–2496, 2011.
- [72] Y. Bai, M. Zhao, C. Zhang et al., "Anti-angiogenic effects of a mutant endostatin: a new prospect for treating retinal and choroidal neovascularization," *PLoS One*, vol. 9, no. 11, article e112448, 2014.
- [73] M. Aikio, M. Hurskainen, G. Brideau et al., "Collagen XVIII short isoform is critical for retinal vascularization, and overexpression of the Tsp-1 domain affects eye growth and cataract formation," *Investigative Ophthalmology & Visual Science*, vol. 54, no. 12, pp. 7450–7462, 2013.
- [74] K. S. Moulton, B. R. Olsen, S. Sonn, N. Fukai, D. Zurakowski, and X. Zeng, "Loss of collagen XVIII enhances neovascularization and vascular permeability in atherosclerosis," *Circulation*, vol. 110, no. 10, pp. 1330–1336, 2004.
- [75] L. Colina-Vegas, W. Villarreal, M. Navarro et al., "Cytotoxicity of Ru(II) piano-stool complexes with chloroquine and chelating ligands against breast and lung tumor cells: interactions with DNA and BSA," *Journal of Inorganic Biochemistry*, vol. 153, pp. 150–161, 2015.
- [76] W. Tang, J. Ma, R. Gu et al., "Lipocalin 2 suppresses ocular inflammation by inhibiting the activation of NF- κ B pathway in endotoxin-induced uveitis," *Cellular Physiology and Biochemistry*, vol. 46, no. 1, pp. 375–388, 2018.
- [77] B. Wu, C. Li, Z. Du et al., "Network based analyses of gene expression profile of LCN2 overexpression in esophageal squamous cell carcinoma," *Scientific Reports*, vol. 4, p. 5403, 2014.
- [78] M. Shen, Y. Tao, Y. Feng, X. Liu, F. Yuan, and H. Zhou, "Quantitative proteomic analysis of mice corneal tissues reveals angiogenesis-related proteins involved in corneal neovascularization," *Biochimica et Biophysica Acta*, vol. 1864, no. 7, pp. 787–793, 2016.
- [79] L. Wu, Y. Du, J. Lok, E. H. Lo, and C. Xing, "Lipocalin-2 enhances angiogenesis in rat brain endothelial cells via reactive oxygen species and iron-dependent mechanisms," *Journal of Neurochemistry*, vol. 132, no. 6, pp. 622–628, 2015.
- [80] S. D. Mesquita, A. C. Ferreira, A. M. Falcao et al., "Lipocalin 2 modulates the cellular response to amyloid beta," *Cell Death and Differentiation*, vol. 21, no. 10, pp. 1588–1599, 2014.
- [81] K. M. Rood, I. A. Buhimschi, K. Rodewald Millen et al., "Evidence for participation of neutrophil gelatinase-associated lipocalin/matrix metalloproteinase-9 (NGAL•MMP-9) complex in the inflammatory response to infection in pregnancies complicated by preterm birth," *American Journal of Reproductive Immunology*, vol. 76, no. 2, pp. 108–117, 2016.
- [82] S. Ghosh, P. Shang, M. Yazdankhah et al., "Activating the AKT2-nuclear factor- κ B-lipocalin-2 axis elicits an inflammatory response in age-related macular degeneration," *The Journal of Pathology*, vol. 241, no. 5, pp. 583–588, 2017.

Research Article

A New Method for Syndrome Classification of Non-Small-Cell Lung Cancer Based on Data of Tongue and Pulse with Machine Learning

Yu-lin Shi ¹, Jia-yi Liu,¹ Xiao-juan Hu,² Li-ping Tu,¹ Ji Cui,¹ Jun Li,¹ Zi-juan Bi,¹ Jia-cai Li,¹ Ling Xu ³ and Jia-tuo Xu ¹

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

²Shanghai Innovation Center of TCM Health Service, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Pudong, Shanghai, China

³Shanghai University of Traditional Chinese Medicine Yueyang Hospital of Integrated Traditional Chinese Medicine and Western Medicine, 110 Ganhe Road, Hongkou, Shanghai, China

Correspondence should be addressed to Ling Xu; xulq67@aliyun.com and Jia-tuo Xu; xjt@fudan.edu.cn

Received 20 April 2021; Revised 12 July 2021; Accepted 23 July 2021; Published 11 August 2021

Academic Editor: Yue Zhang

Copyright © 2021 Yu-lin Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. To explore the data characteristics of tongue and pulse of non-small-cell lung cancer with Qi deficiency syndrome and Yin deficiency syndrome, establish syndrome classification model based on data of tongue and pulse by using machine learning methods, and evaluate the feasibility of syndrome classification based on data of tongue and pulse. **Methods.** We collected tongue and pulse of non-small-cell lung cancer patients with Qi deficiency syndrome ($n = 163$), patients with Yin deficiency syndrome ($n = 174$), and healthy controls ($n = 185$) using intelligent tongue diagnosis analysis instrument and pulse diagnosis analysis instrument, respectively. We described the characteristics and examined the correlation of data of tongue and pulse. Four machine learning methods, namely, random forest, logistic regression, support vector machine, and neural network, were used to establish the classification models based on symptom, tongue and pulse, and symptom and tongue and pulse, respectively. **Results.** Significant difference indices of tongue diagnosis between Qi deficiency syndrome and Yin deficiency syndrome were TB-a, TB-S, TB-Cr, TC-a, TC-S, TC-Cr, perAll, and the tongue coating texture indices including TC-CON, TC-ASM, TC-MEAN, and TC-ENT. Significant difference indices of pulse diagnosis were t_4 and t_5 . The classification performance of each model based on different datasets was as follows: tongue and pulse < symptom < symptom and tongue and pulse. The neural network model had a better classification performance for symptom and tongue and pulse datasets, with an area under the ROC curves and accuracy rate which were 0.9401 and 0.8806. **Conclusions.** It was feasible to use tongue data and pulse data as one of the objective diagnostic basis in Qi deficiency syndrome and Yin deficiency syndrome of non-small-cell lung cancer.

1. Introduction

Lung cancer is a common malignant tumor of the lung and is a major cause of morbidity and mortality. It is estimated that the number of deaths from lung cancer accounts for about 24% of all cancer deaths in the United States [1, 2]. An organization report shows that lung cancer causes approximately 1.76 million deaths worldwide each year, accounting for 18.7% of all cancer deaths [3]. Non-small-cell lung cancer (NSCLC) is the most common histological type of lung can-

cer, accounting for more than 80% of primary lung cancers [4]. Sixty percent of NSCLC cases have metastasized at the time of diagnosis. The 5-year survival rate for advanced NSCLC is lower than 5%, and early diagnosis of lung cancer is an important opportunity to reduce mortality [5, 6]. The current treatment methods for NSCLC mainly include surgery, radiotherapy, chemotherapy, and targeted therapy [7, 8]. Chemotherapy is the most common treatment. However, patients with poor health often have a low tolerance to conventional treatment with a tendency of drug resistance [9].

Traditional Chinese medicine (TCM) has a long history and rich experience in the treatment of lung cancer, which is one of the main methods of comprehensive treatment of lung cancer in China. Systematic evaluation of TCM shows that TCM combined with radiotherapy and chemotherapy and targeted therapy had certain advantages in alleviating symptoms, stabilizing tumors, improving life quality, and prolonging survival period [10]. TCM has been proved to be an effective method for the treatment of advanced lung cancer. On the basis of accurate syndrome differentiation, TCM plays an active role in each stage of the occurrence and development of lung cancer [11, 12].

Syndrome differentiation and treatment is the basic principle of TCM to diagnose and deal with diseases. It is a process of comprehensive judgment on the four types of diagnostic information of patients based on the theory of TCM combined with the doctor's experience [13]. Accurate syndrome differentiation is able to provide a basis for the treatment of diseases and is the foundation of clinical efficacy. Traditional syndrome differentiation and treatment inevitably suffer from subjectivity and ambiguity, which actually hinders the development of TCM. Microsyndrome differentiation is a method of using modern advanced technology to go deep into the body's microcosmic level to understand and differentiate syndromes on the basis of macroscopic syndrome differentiation. Microsyndrome differentiation can be used to guide disease differentiation and syndrome differentiation, explore the cause and pathogenesis, and evaluate the efficacy and guide the prognosis of the disease [14]. Previous studies have verified that there is a close relationship between different syndromes and physical and chemical indices. A combination of microindex and macrosymptom can assist syndrome differentiation effectively.

With the rapid development of modern research on tongue and pulse diagnoses, a variety of tongue and pulse diagnoses instruments are widely used in clinical practice. This has generated a large number of objective data of tongue and pulse diagnoses, which are also microscopic indices in a sense. In recent years, studies based on data of tongue and pulse diagnoses have been increasing, with many researchers applying machine learning and data mining methods to the fields of image recognition, target detection, natural language processing, and others [15–18]. In addition, studies have demonstrated that accurate detection, identification, and multidimensional quantitative analysis based on tongue data and pulse data have been gradually applied to disease diagnosis. By constructing the diagnostic relationship between tongue and pulse and health status, it not only saves medical resources but also greatly improves diagnosis efficiency and treatment [19–22]. Qi deficiency syndrome and Yin deficiency syndrome are the two main common syndromes of NSCLC. When the symptoms are not obvious, the traditional symptom-based syndrome differentiation cannot be carried out. The modern study of tongue and pulse diagnoses research provides a good data basis for TCM syndrome differentiation.

Tongue data and pulse data are the most representative data of four diagnoses of TCM. The data collected and ana-

lyzed under the standardized condition has a high level of stability, which provides reliable objective data for intelligent syndrome differentiation. Among all kinds of syndromes, tongue and pulse are related to some extent, but the traditional syndrome differentiation cannot be clearly explained due to the lack of accurate data. With the development of diagnosis technology, the analysis and interpretation of the relationship between tongue and pulse can be realized more clearly. In this study, two common syndromes of NSCLC were selected to explore the differences of tongue data and pulse data and quantitatively analyze the data correlation of tongue and pulse, using machine learning methods to establish syndrome classification models based on macrosymptom, objective tongue and pulse, and macrosymptom and objective tongue and pulse, and evaluate the contribution rate of the objective data of tongue and pulse to syndrome differentiation.

2. Materials and Methods

2.1. Study Design and Subjects. We selected a total of 337 patients from the oncology department of Yueyang Hospital of Integrated Traditional Chinese and Western Medicine from January 2018 to October 2020, including 163 patients with Qi deficiency syndrome and 174 patients with Yin deficiency syndrome. All patients were pathologically or cytologically confirmed to be NSCLC. We additionally selected a total of 184 healthy people from Shuguang Hospital of Shanghai University of Traditional Chinese Medicine from January 2018 to October 2020 as the healthy controls. The flowchart is shown in Figure 1.

2.2. Diagnostic Criteria. Diagnostic criteria of Western medicine: according to the clinical practice guidelines for lung cancer screening issued by the National Comprehensive Cancer Network (NCCN) [23] and the fourth edition lung cancer histological classification standards of "Classification of Lung Tumors" [24, 25] issued by the World Health Organization.

TCM Syndrome Differentiation Standard: according to the "Technical Guidelines for Clinical Research of New Drugs of Syndromes" [26] and the Syndrome Part of TCM Clinical Diagnosis and Treatment Terms [27] and textbooks of Common Diseases and Symptoms in Internal Medicine of Traditional Chinese Medicine.

The main manifestations of Qi deficiency syndrome are cough, white or foamy phlegm, small amount of hemoptysis, chest tightness, shortness of breath, low fever, spontaneous sweating, lack of energy, pale complexion, poor appetite, loose stools, pale red tongue with tooth marks, thin white coating, and thin pulse. The main manifestations of Yin deficiency syndrome are cough without phlegm, or less but sticky phlegm, phlegm with blood, shortness of breath and dull chest pain, low fever, dry mouth, night sweat, upset and insomnia, red tongue, little or bare without tongue coating, and thin and rapid pulse. The syndrome was determined by at least three senior physicians to ensure the consistency and authenticity of syndrome differentiation.

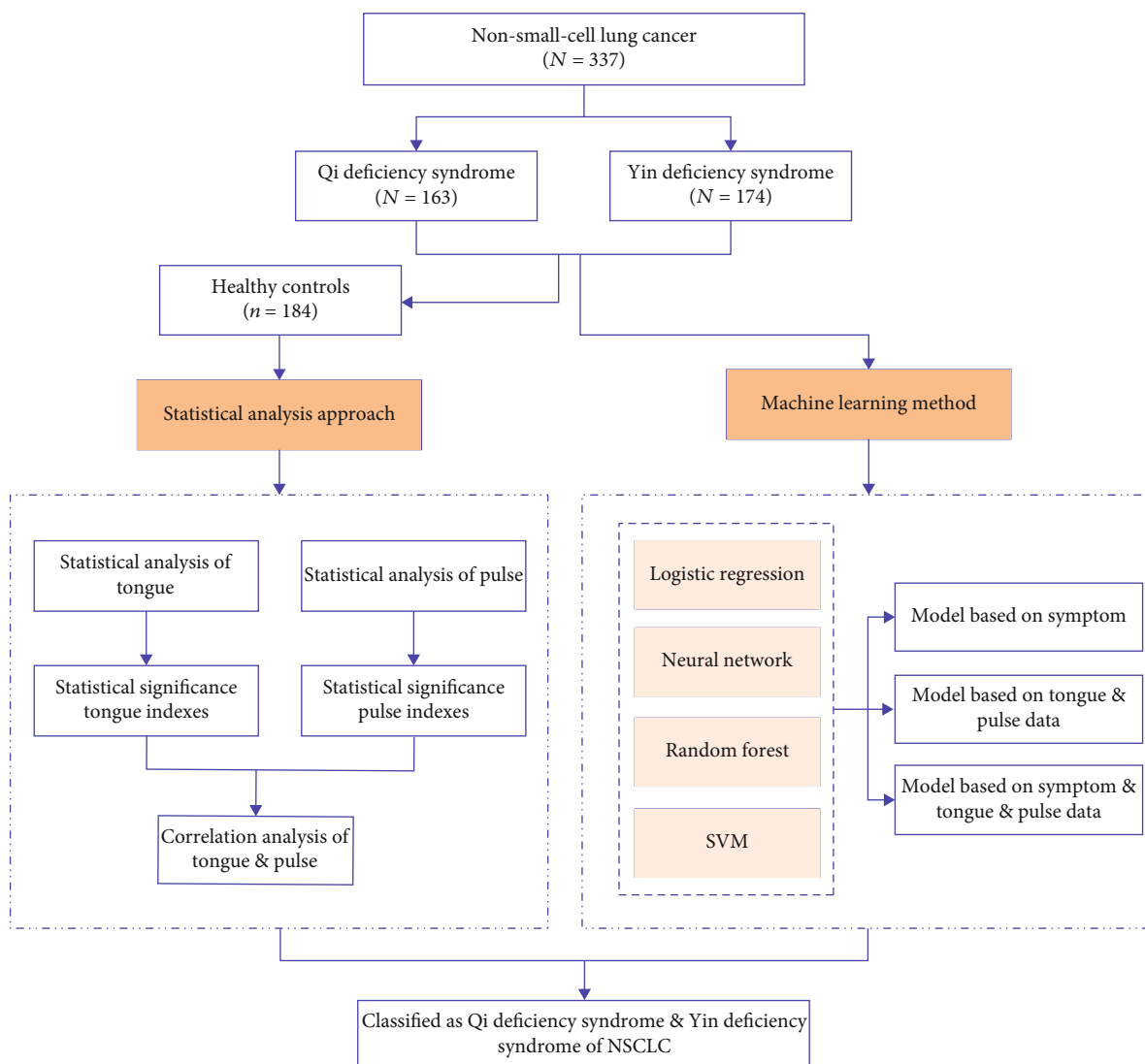


FIGURE 1: Flowchart.

2.3. *Inclusion and Exclusion Criteria.* The inclusion criteria are as follows: (1) meeting the above diagnostic criteria, (2) confirmed by pathology or cytology, (3) no serious liver or kidney damage, and (4) know and sign informed consent.

The exclusion criteria are as follows: (1) those who did not meet the inclusion criteria for NSCLC, (2) patients with Qi deficiency syndrome combined with Yin deficiency syndrome, (3) patients with severe primary diseases such as cardiovascular, cerebrovascular, liver, kidney, and blood system, (4) pregnant or lactating women, (5) psychopath, and (6) patients who were unable to cooperate with research work due to subjective and objective reasons and who had poor compliance.

2.4. *Collecting Clinical Data of Tongue and Pulse.* We used TFDA-1 digital tongue diagnosis instrument and PDA-1 digital pulse diagnosis instrument developed by the National Key Research and Development Program to collect tongue and pulse diagnostic data of patients, respectively. We used the Information Record Form of TCM Clinical Four Diagnostics (Copyright No.: 2016Z11L025702) developed by our

research group to record the symptoms of patients [28]. All the work of tongue and pulse diagnoses collection and inquiry were completed by professional personnel of TCM or integrated TCM and western medicine who had received standardized training. Each patient was consulted by at least two professional researchers, and the syndromes of all patients were judged by three senior doctors to ensure the consistency and authenticity of data collection and interpretation and minimize deviation.

TFDA-1 digital tongue diagnosis instrument and tongue diagnosis analysis system (TDAS v2.0) are shown in Figures 2 and 3. The tongue was imaged by a video camera (Nikon 1 J5) with a fixed-focal lens which has 12 megapixels, and the picture resolution is 5568 * 3712. TFDA-1 digital tongue diagnosis instrument uses LED light sources, and a curved reflector is set in front of the light sources to ensure the uniformity of illumination in all parts when the tongue image is collected. The color rendering index of light source is 96, and color temperature is around 5,000–6,500 K. Parameters of the TFDA-1 digital tongue diagnosis instrument are



FIGURE 2: TFDA-1 digital tongue diagnosis instrument: (a) front view; (b) profile view.

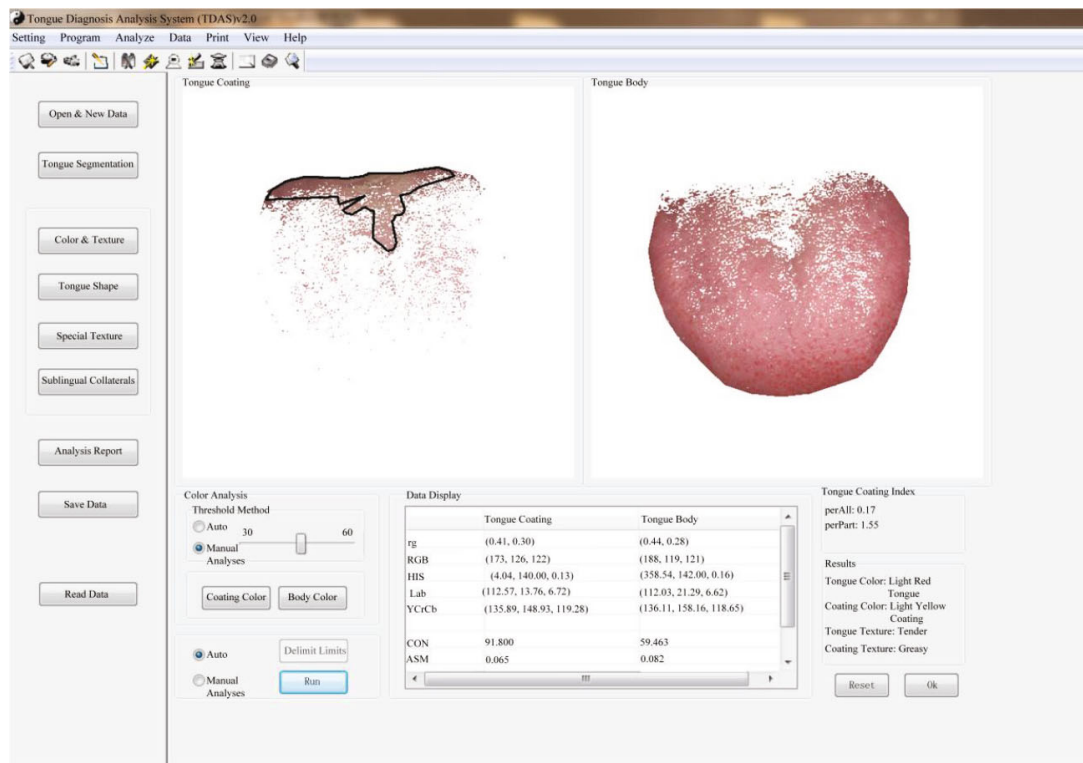


FIGURE 3: Tongue diagnosis analysis system (TDAS v2.0) of TFDA-1 digital tongue diagnosis instrument.

as follows: white balance, center-weighted metering, M mode, shutter speed of 1/125, aperture value of F6.3, and ISO sensitivity of 200.

PDA-1 digital pulse diagnosis instrument and its corresponding sphygmogram are shown in Figure 4. The PDA-1 pulse diagnosis instrument uses a pressure sensor. Place the probe at the guan place of the patient's left hand, fix the strap, and adjust the tightness of the strap so that the sphygmogram reaches the best peak (the peak value of the main sphygmogram is 2 grids and above). Collect 30 s after the waveform is stable.

Tongue indices can be divided into two categories: tongue body (TB) index and tongue coating (TC) index which

mainly come from the three color spaces of Lab, HIS, and YCrCb [29–32]. Each parameter of tongue diagnosis and pulse diagnosis has its corresponding medical significance [32–34]. In tongue indices, they are R (Red), G (Green), B (Blue), H (Hue), S (Saturation), I (Intensity) and L (Light), a (red-green axis), b (yellow-blue axis), Y (brightness), Cr (difference between red signal and brightness), Cb (difference between blue signal and brightness), texture indices include CON (Contrast), ASM (Angular Second Moment), ENT (Entropy), MEAN (Mean), and tongue coating indices include perAll and perPart. perAll represents the ratio of coated tongue area to total tongue area, and perPart represents the ratio of coated tongue area to noncoated tongue

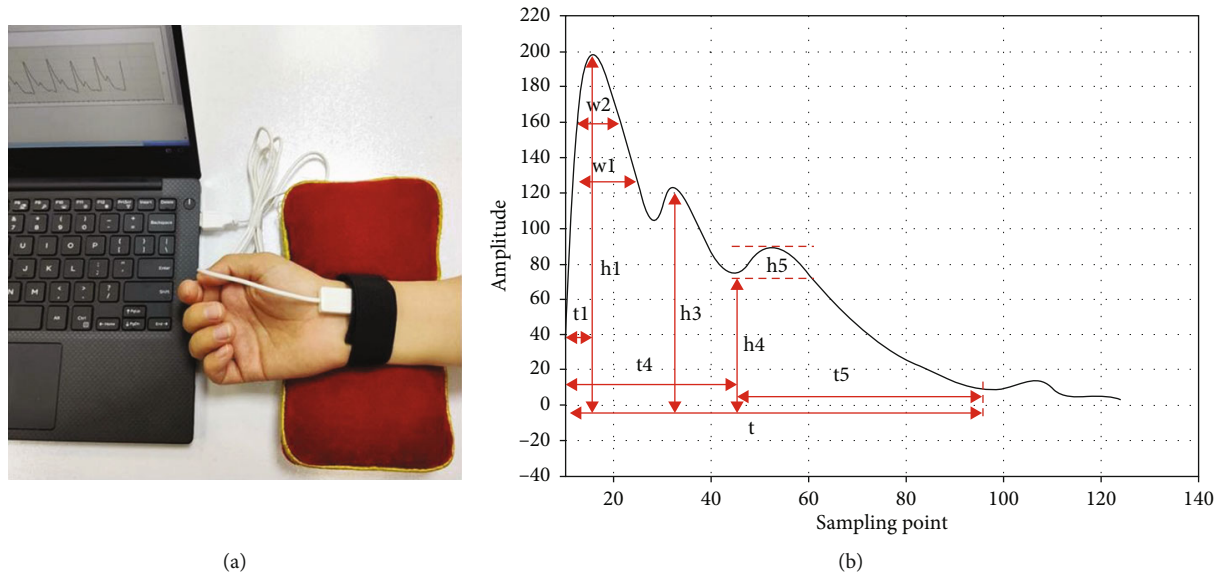


FIGURE 4: PDA-1 digital pulse diagnosis instrument and its corresponding sphygmogram: (a) PDA-1 digital pulse diagnosis instrument; (b) sphygmogram.

area. In pulse indices, h_1-h_5 mainly represent the amplitude height. h_1 is the main wave amplitude, h_3 is heavy wave front wave amplitude, h_3/h_1 is the ratio of heavy wave front wave amplitude to the amplitude of the main wave, h_4 is the dicrotic notch amplitude, h_4/h_1 is the ratio of the dicrotic notch amplitude to the amplitude of the main wave, h_5 is the gravity wave amplitude, and h_5/h_1 is the ratio of gravity wave amplitude to the amplitude of the main wave. t represents a complete pulse cycle, and t_1 is the time value from the start point to the crest point of the main wave on the sphygmogram. t_4 is the time value from the start point to the dicrotic notch on the sphygmogram, and t_5 is the time value from the dicrotic notch to the end point on the sphygmogram. w_1 is the width at 1/3 of the main wave, and w_2 is the width at 1/5 of the main wave. All the tongue and pulse indices are extracted by special tongue analysis software (TDAS v2.0) and pulse analysis software (PulseCol).

2.5. Statistical Analysis. SPSS 26.0 was used for statistical analysis. Categorical variables were expressed as percentages (%). Continuous variables were expressed as mean \pm standard deviation (SD) for those with normal distribution or median (interquartile range) for those with skewed distribution. Continuous variables were compared with analysis of variance (ANOVA) or rank-sum test (Kruskal-Wallis H test), and the correlation heat map was made by GraphPad Prism 8.0. A two-sided P value < 0.05 was considered statistically significant.

2.6. Classification by Machine Learning Approach. We used four machine learning methods, namely, neural network, random forest, support vector machine (SVM), and logistic regression to set the ratio of training set to test set at 8:2 using Orange (3.26.0) software. We used adjusted parameters of each model to establish classification and diagnosis models of Qi deficiency syndrome and Yin deficiency syndrome of

NSCLC based on “symptom,” “tongue and pulse,” and “symptom and tongue and pulse”, respectively. We used accuracy, precision, F1-score (F1), sensitivity, specificity, and area under the curve (AUC) as evaluation indices to evaluate the predictive performance. AUC was the area under the ROC curve. The larger the value, the better the classification effect of the classifier. The calculation formula of each index was as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%, \quad (2)$$

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

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

$$F1 = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}. \quad (5)$$

In the above statements, True Positive (TP) was the positive sample predicted by the model as the positive category. True Negative (TN) was the negative sample predicted by the model as the negative category. False Positives (FP) was the negative sample predicted by the model as the positive category. False Negative (FN) was the positive sample predicted by the model as the negative category.

3. Results

3.1. Characteristics of Participants. The basic statistical analysis result of the three groups is shown in Table 1.

The result showed that people with Qi deficiency syndrome and Yin deficiency syndrome had a statistically

TABLE 1: Basic statistical analysis.

Characteristic		Healthy controls ($n = 184$)	Qi deficiency syndrome ($n = 163$)	Yin deficiency syndrome ($n = 174$)
Sex, n (%)	Male	96 (52.17)	72 (44.17)	89 (51.15)
	Female	88 (47.83)	91 (55.83)	85 (48.85)
Age, years		27.00 (29.00-24.25)	67.00 (59.00-71.00)**	67.00 (60.00-72.00)**

vs. healthy controls, ** $P < 0.01$.

TABLE 2: Statistical analysis of tongue diagnosis data (mean (SD), median (P_{25} , P_{75})).

Domain	Color space	Index	Healthy controls ($n = 184$)	Qi deficiency syndrome ($n = 163$)	Yin deficiency syndrome ($n = 174$)
TB	Lab	TB-L	103.99 (100.81-108.79)	96.31 (75.15-102.89)**	99.83 (80.51-103.24)**
		TB-a	19.98 \pm 2.82	19.31 \pm 3.81	21.06 \pm 4.23*##
		TB-b	4.76 (0.82-7.00)	7.04 (5.47-8.28)**	7.04 (5.47-8.28)**
		TB-H	176.22 (168.50-180.95)	180.00 (177.98-182.83)**	180.00 (177.98-182.83)**
	HIS	TB-S	0.17 (0.16-0.20)	0.17 (0.15-0.19)	0.17 (0.15-0.19)*##
		TB-I	117.00 (108.00-132.00)	116.00 (109.00-126.00)**	116.00 (109.00-126.00)
		TB-Y	114.98 (107.03-126.56)	114.35 (106.900-123.72)	114.35 (106.900-123.72)*
	YCrCb	TB-Cr	151.41 \pm 3.05	152.29 \pm 3.89	154.15 \pm 4.44***
		TB-Cb	121.61 (119.75-124.82)	119.84 (118.53-120.99)**	119.27 (118.09-120.57)**
		TB-CON	71.47 (46.96-99.54)	74.56 (48.28-94.64)	60.96 (45.32-86.08)
	Texture index	TB-ASM	0.08 (0.07-0.10)	0.07 (0.07-0.09)	0.09 (0.07-0.10)
		TB-MEAN	0.03 (0.02-0.03)	0.03 (0.02-0.03)	0.02 (0.02-0.03)
		TB-ENT	1.21 (1.11-1.28)	1.22 (1.12-1.28)	1.17 (1.10-1.25)
		TC-L	109.24 (104.97-113.54)	89.38 (76.22-104.87)**	95.35 (82.53-105.08)**
TC	Lab	TC-a	12.31 \pm 2.69	12.75 \pm 3.21	14.25 \pm 3.78***
		TC-b	2.71(-1.16-5.32)	5.59 (4.24-6.62)**	5.86 (4.35-7.26)**
		TC-H	176.70 (162.43-183.25)	183.00 (180.00-186.35)**	182.58 (178.64-185.72)**
	HIS	TC-S	0.11 (0.09-0.13)	0.12 (0.10-0.14)*	0.13 (0.11-0.17)***
		TC-I	130.00 (117.00-142.75)	119.00 (99.00-135.00)**	115.00 (92.75-133.00)**
		TC-Y	126.78 (115.63-137.70)	118.65 (99.72-132.07)**	114.02 (95.19-129.53)**
	YCrCb	TC-Cr	142.89 (140.89-145.181)	143.97 (142.27-146.51)**	145.49 (143.00-148.79)***
		TC-Cb	123.90 (121.54-127.61)	121.36 (120.34-122.81)**	121.35 (120.01-122.67)**
	Area index	perAll	0.54 (0.43-0.69)	0.44 (0.34-0.50)**	0.38 (0.21-0.50)*##
		perPart	1.09 (1.02-1.22)	1.24 (1.11-1.42)**	1.28 (1.11-1.57)**
Texture index	TC-CON	89.27 (62.31-124.17)	83.13 (63.82-123.30)	71.53 (44.56-115.98)***	
	TC-ASM	0.07 (0.06-0.08)	0.07 (0.06-0.08)	0.08 (0.06-0.10)***	
	TC-MEAN	0.03 (0.02-0.03)	0.03 (0.02-0.03)	0.03 (0.02-0.03)***	
	TC-ENT	1.26 (1.18-1.34)	1.25 (1.18-1.34)	1.21 (1.09-1.31)***	

vs. healthy controls, * $P < 0.05$, vs. healthy controls, ** $P < 0.01$. vs. Qi deficiency syndrome, # $P < 0.05$, vs. Qi deficiency syndrome, ## $P < 0.01$.

significantly higher age than healthy controls. However, there was no difference in age between people with Qi deficiency syndrome and Yin deficiency syndrome.

3.2. *Statistical Analysis of Tongue Data.* Statistical analysis result of tongue diagnosis data in the three groups is shown in Table 2.

The result showed that (1) compared with Qi deficiency syndrome, there were more significant differences between Yin deficiency syndrome and the healthy controls. (2) In

the significant difference indices between Yin deficiency syndrome and healthy controls, except for the texture index of tongue coating, the changes of tongue body index of Yin deficiency syndrome were more significant than that of tongue coating index. (3) Significant difference tongue indices between Qi deficiency syndrome and Yin deficiency syndrome were TB-a, TB-S, TB-Cr, TC-a, TC-S, TC-Cr, perAll, and TC-CON, TC-ASM, TC-MEAN, and TC-ENT; among them, TB-a, TB-Cr, TC-a, TC-S, TC-Cr, and TC-ASM of Yin deficiency syndrome were higher than those of Qi

TABLE 3: Statistical analysis of pulse diagnosis data (mean (SD), median (P_{25} , P_{75})).

Index	Healthy controls ($n = 184$)	Qi deficiency syndrome ($n = 163$)	Yin deficiency syndrome ($n = 174$)
t_1 (s)	0.13 (0.12-0.14)	0.14 (0.13-0.15)**	0.14 (0.13-0.14)**
t_4 (s)	0.34 (0.32-0.36)	0.37 (0.35-0.39)**	0.37 (0.34-0.39)**#
t_5 (s)	0.41 (0.39-0.42)	0.43 (0.41-0.46)**	0.42 (0.40-0.44)**##
t (s)	0.80 (0.75-0.88)	0.86 (0.76-0.97)**	0.84 (0.72-0.94)
h_1 (mv)	13.89 (11.53-16.41)	10.99 (7.62-15.42)**	11.56 (8.86-16.51)**
h_3 (mv)	8.48 (6.56-10.59)	6.64 (4.38-10.07)**	7.18 (4.85-10.12)**
h_4 (mv)	5.21 (4.18-6.32)	2.18 (1.37-3.24)**	2.53 (1.44-3.50)**
h_5 (mv)	0.50 (0.15-0.95)	0.23 (0.05-0.69)**	0.21 (0.05-0.60)**
h_3/h_1	0.62 (0.52-0.70)	0.61 (0.53-0.71)	0.60 (0.49-0.73)
h_1/t_1	4.43 (3.49-5.35)	3.22 (2.26-4.57)**	3.45 (2.68-4.82)**
h_4/h_1	0.38 (0.32-0.43)	0.21 (0.12-0.31)**	0.21 (0.14-0.28)**
t_1/t	0.16 (0.14-0.17)	0.16 (0.14-0.19)	0.17 (0.14-0.19)
t_4/t_5	0.83 (0.80-0.88)	0.86 (0.82-0.91)**	0.87 (0.82-0.91)**
w_1/t	0.20 (0.15-0.23)	0.21 (0.19-0.23)**	0.21 (0.19-0.23)**
w_2/t	0.12 (0.10-0.16)	0.15 (0.13-0.18)**	0.15 (0.13-0.18)**

vs. healthy controls, * $P < 0.05$, vs. healthy controls, ** $P < 0.01$. vs. Qi deficiency syndrome, # $P < 0.05$, vs. Qi deficiency syndrome, ## $P < 0.01$.

deficiency syndrome, while perAll, TC-CON, and TC-ENT of Yin deficiency syndrome were lower than those of Qi deficiency syndrome.

3.3. Statistical Analysis of Pulse Data. Statistical analysis result of pulse diagnosis data in the three groups is shown in Table 3.

The result showed that (1) the pulse parameters t_1 , t_4 , t_5 , h_1 , h_3 , h_4 , h_5 , h_1/t_1 , h_4/h_1 , t_4/t_5 , w_1/t , and w_2/t of Qi deficiency syndrome and Yin deficiency syndrome had statistical significance compared with those of healthy controls. (2) Only two parameters, t_4 and t_5 , showed statistically significant differences between Qi deficiency syndrome and Yin deficiency syndrome.

3.4. Correlation Analysis of Tongue Data and Pulse Data. Tongue data and pulse data were statistically significantly correlated among people with Qi deficiency syndrome and Yin deficiency syndrome (Figure 5 and Table 4). Heat map result of Qi deficiency syndrome is shown in Figure 5.

Correlation analysis result of tongue data and pulse data between Qi deficiency syndrome is shown in Table 4.

The result showed that (1) there was a strong correlation between the tongue coating texture parameters, and the color space parameters of the tongue coating and the tongue body were also correlated. The correlation between the tongue coating texture parameters and the color space parameters was weaker than the correlation of the pulse parameters. (2) There was a definite correlation between pulse parameters t_4 and tongue parameters TC-ASM, TC-ENT, and TC-MEAN, with a correlation coefficient of -0.18, 0.18, and 0.18, respectively. (3) There was a weak correlation between t_5 and TB-Cr with a correlation coefficient of -0.16 ($P < 0.05$).

The heat map result of Yin deficiency syndrome is shown in Figure 6.

The correlation analysis result of tongue data and pulse data between Yin deficiency syndrome is shown in Table 5.

The result showed that (1) similar to Qi deficiency syndrome, the tongue coating texture parameters of Yin deficiency syndrome had a strong correlation, and the color space parameters of the tongue coating and tongue body were also strongly correlated. The correlation between tongue coating texture parameters and color space parameters was weaker than that of pulse parameters. (2) There was a certain correlation between pulse parameters t_4 and tongue parameters TC-ASM and TC-a. Both of the correlation coefficients were -0.14, but the difference was not statistically significant ($P > 0.05$). (3) t_5 was strongly correlated with TB-a, TC-S, TC-Cr, and TB-a, and the correlation coefficients were -0.33, -0.27, -0.23, and -0.23, respectively ($P < 0.01$). The correlation coefficients of t_5 with TB-Cr, TB-S, and TC-ASM were -0.21, -0.20, and -0.20, respectively ($P < 0.01$).

The correlation analysis result showed that correlation intensity of tongue and pulse in Yin deficiency syndrome was significantly stronger than that in Qi deficiency syndrome, and compared with Qi deficiency syndrome, the correlation between t_4 and tongue indices in Yin deficiency syndrome was significantly reduced, while the correlation between t_5 and tongue indices was significantly increased.

3.5. Machine Learning Results. Based on neural network, random forest, SVM, and logistic regression four machine learning methods, the modeling result of Qi deficiency syndrome and Yin deficiency syndrome based on symptom, tongue and pulse, and symptom and tongue and pulse is shown in Table 6.

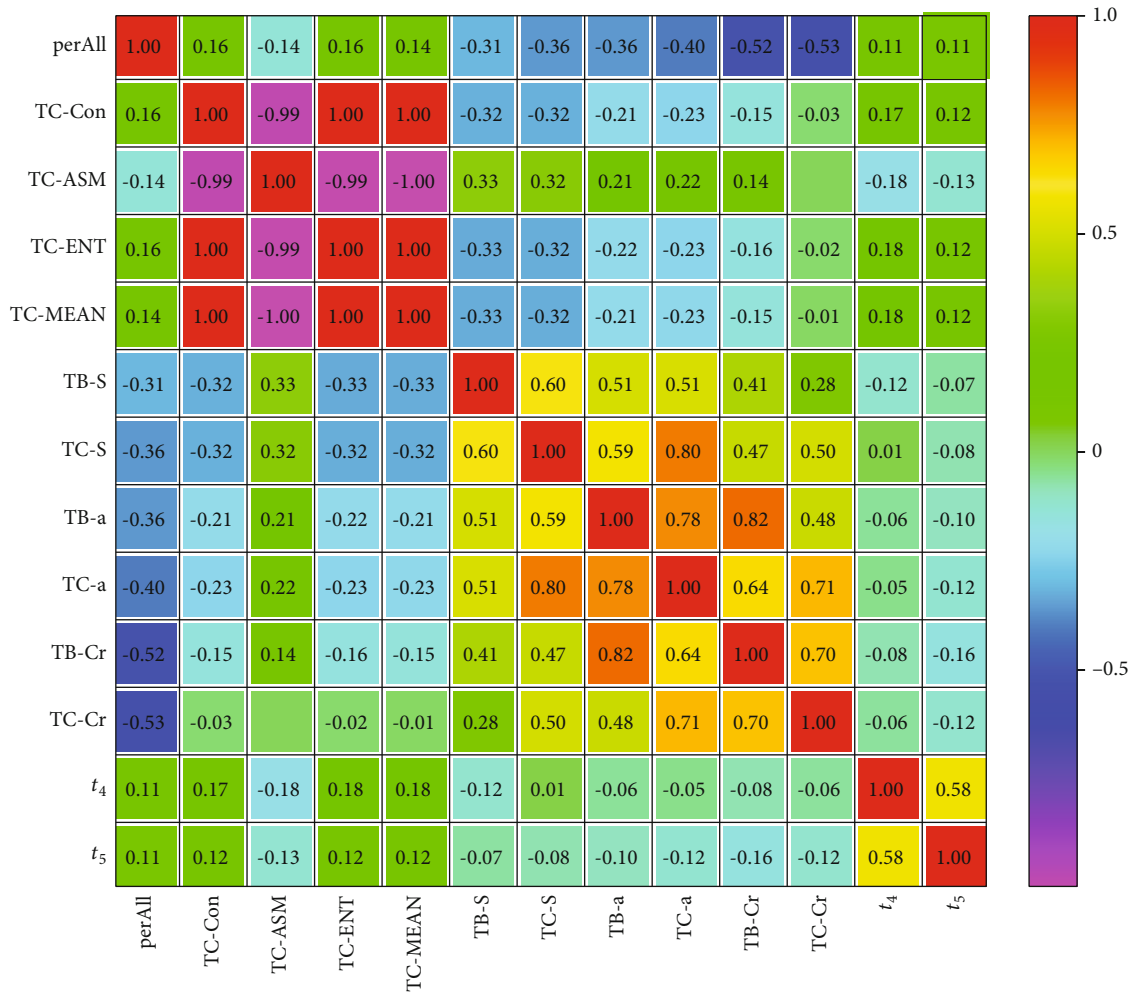


FIGURE 5: Heat map of tongue and pulse correlation analysis of Qi deficiency syndrome.

TABLE 4: Correlation analysis of tongue data and pulse data of Qi deficiency syndrome.

Index	perAll	TC-CON	TC-ASM	TC-ENT	TC-MEAN	TB-S	TC-S	TB-a	TC-a	TB-Cr	TC-Cr	t ₄	t ₅
perAll	1.00												
TC-CON	0.16*	1.00											
TC-ASM	-0.14	-0.99**	1.00										
TC-ENT	0.16*	1.00**	-0.99**	1.00									
TC-MEAN	0.14	1.00**	-1.00**	1.00**	1.00								
TB-S	-0.31**	-0.32**	0.33**	-0.33**	-0.33**	1.00							
TC-S	-0.36**	-0.32**	0.32**	-0.32**	-0.32**	0.60**	1.00						
TB-a	-0.36**	-0.21**	0.21**	-0.22**	-0.21**	0.51**	0.59**	1.00					
TC-a	-0.40**	-0.23**	0.22**	-0.23**	-0.23**	0.51**	0.80**	0.78**	1.00				
TB-Cr	-0.52**	-0.15	0.14	-0.16*	-0.15	0.41**	0.47**	0.82**	0.64**	1.00			
TC-Cr	-0.53**	-0.03	-0.00	-0.02	-0.01	0.28**	0.50**	0.48**	0.71**	0.70**	1.00		
t ₄	0.11	0.17*	-0.18*	0.18*	0.18*	-0.12	0.01	-0.06	-0.05	-0.08	-0.06	1.00	
t ₅	0.11	0.12	-0.13	0.12	0.12	-0.07	-0.08	-0.10	-0.13	-0.16*	-0.12	0.58**	1.00

*P < 0.05, **P < 0.01.

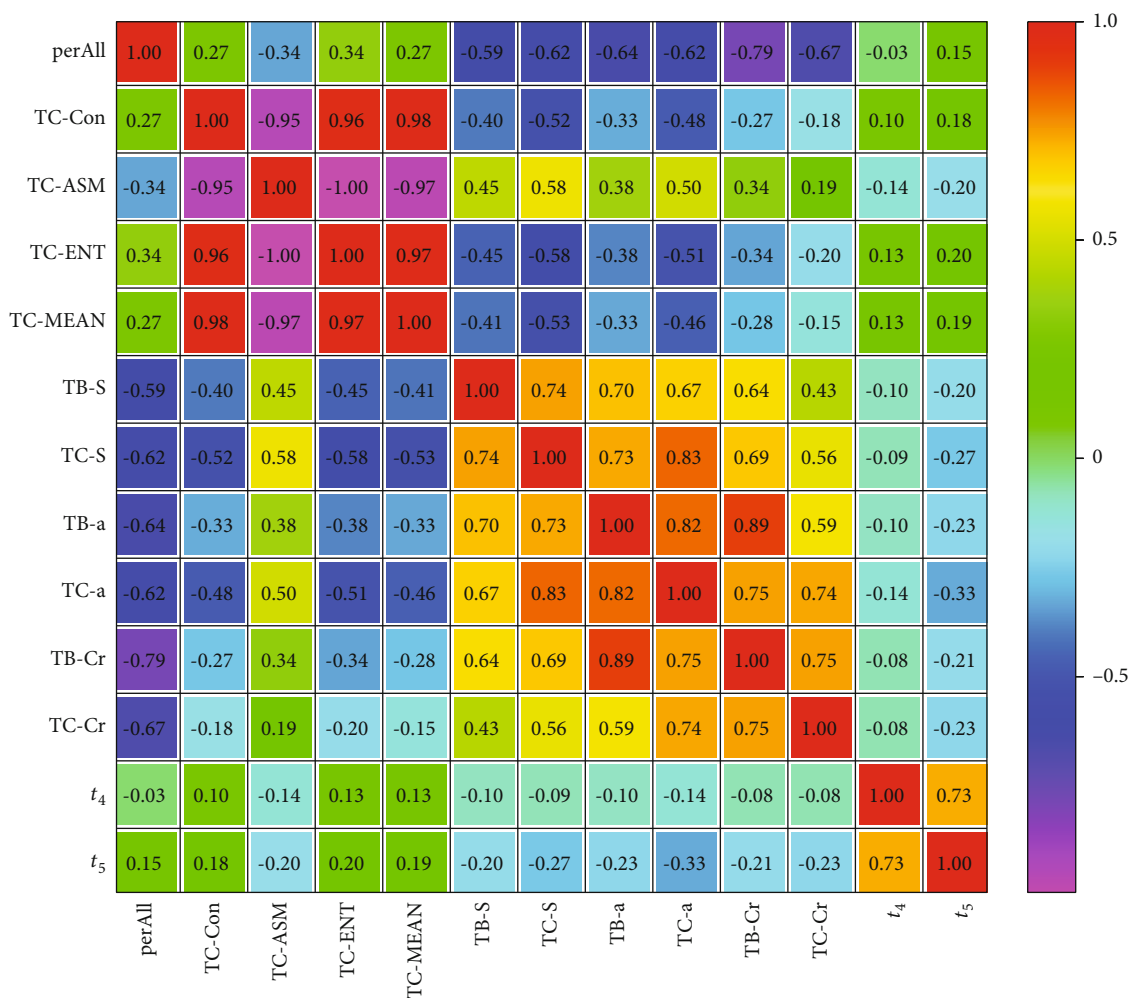


FIGURE 6: Heat map of tongue and pulse correlation analysis of Yin deficiency syndrome.

TABLE 5: Correlation analysis of tongue data and pulse data of Yin deficiency syndrome.

Index	perAll	TC-CON	TC-ASM	TC-ENT	TC-MEAN	TB-S	TC-S	TB-a	TC-a	TB-Cr	TC-Cr	t ₄	t ₅
perAll	1.00												
TC-CON	0.27**	1.00											
TC-ASM	-0.34**	-0.95**	1.00										
TC-ENT	0.34**	0.96**	-1.00**	1.00									
TC-MEAN	0.27**	0.98**	-0.97**	0.97**	1.00								
TB-S	-0.59**	-0.40**	0.45**	-0.45**	-0.41**	1.00							
TC-S	-0.62**	-0.52**	0.57**	-0.58**	-0.53**	0.75**	1.00						
TB-a	-0.64**	-0.33**	0.38**	-0.39**	-0.33**	0.70**	0.73**	1.00					
TC-a	-0.62**	-0.48**	0.50**	-0.51**	-0.46**	0.67**	0.83**	0.82**	1.00				
TB-Cr	-0.79**	-0.27**	0.34*	-0.34**	-0.28**	0.64**	0.69**	0.89**	0.75**	1.00			
TC-Cr	-0.67**	-0.180*	0.19*	-0.20**	-0.15	0.43**	0.56**	0.59**	0.74**	0.75**	1.00		
t ₄	-0.03	0.10	-0.14	0.13	0.13	-0.10	-0.09	-0.10	-0.14	-0.08	-0.08	1.00	
t ₅	0.15	0.18	-0.20**	0.20	0.19*	-0.21**	-0.27**	-0.23**	-0.33**	-0.21**	-0.23**	0.73**	1.00

P < 0.05, *P < 0.01.

TABLE 6: Performance of models for detecting Qi deficiency syndrome of NSCLC based on different datasets.

Datasets	Model	AUC	Sensitivity	Specificity	F1	Precision	Accuracy
Symptom	Neural network	0.9223	0.9063	0.8286	0.8657	0.8286	0.8657
	SVM	0.9321	0.8750	0.8857	0.8750	0.8750	0.8806
	Logistic regression	0.9000	0.8125	0.8286	0.8125	0.8125	0.8209
	Random forest	0.9116	0.7813	0.8571	0.8065	0.8333	0.8209
Tongue & pulse	Neural network	0.7677	0.6316	0.6897	0.6761	0.7273	0.6567
	SVM	0.7455	0.6842	0.6552	0.7027	0.7222	0.6716
	Logistic regression	0.8022	0.6842	0.8276	0.7536	0.8387	0.7463
	Random forest	0.7314	0.5263	0.8621	0.6452	0.8333	0.6716
Symptom & tongue & pulse	Neural network	0.9401	0.9310	0.8421	0.8710	0.8182	0.8806
	SVM	0.9328	0.6552	0.9737	0.7755	0.9500	0.8358
	Logistic regression	0.9301	0.7931	0.8684	0.8070	0.8214	0.8358
	Random forest	0.9229	0.8966	0.8421	0.8525	0.8125	0.8657

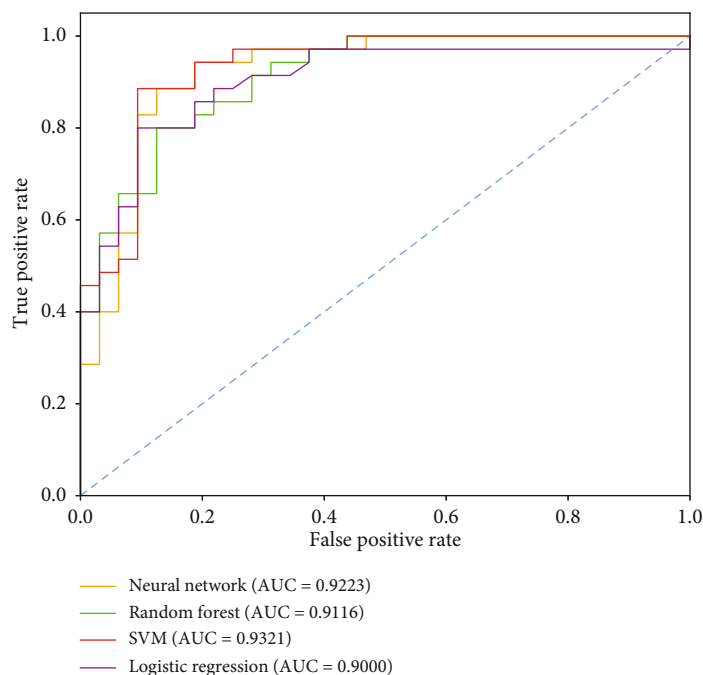


FIGURE 7: ROC curves of Qi deficiency syndrome model based on symptom.

The ROC curves of the models based on symptom, tongue and pulse, and symptom and tongue and pulse are shown in Figures 7–9, respectively.

According to the above modeling results, the classification efficiency of each model based on different datasets had the following order: tongue and pulse < symptom < symptom and tongue and pulse. Among them, the SVM model had a better classification performance for symptom datasets, and the area under the ROC curve was 0.9321. The logistic regression model had a better classification performance for tongue and pulse datasets, with an area under the ROC curve of 0.9401. The neural network model had a better classification performance for the symptom and tongue and pulse datasets, with an area under the ROC curve of 0.9401.

4. Discussion

Treatment based on syndrome differentiation is the basic principle of TCM to recognize and treat diseases. It runs through the whole process of prevention and rehabilitation of medical care practices. Syndrome differentiation is used to recognize the disease and determine the syndrome, and treatment is to establish treatment methods and prescription drugs based on the results of syndrome differentiation. Syndrome differentiation is the prerequisite and basis for treatment. Accurate syndrome differentiation results in a good therapeutic effect. Qi deficiency syndrome and Yin deficiency syndrome are two common syndromes in TCM. According to the basic theory of TCM syndrome differentiation, Qi deficiency syndrome refers to the lack of vitality of the body and

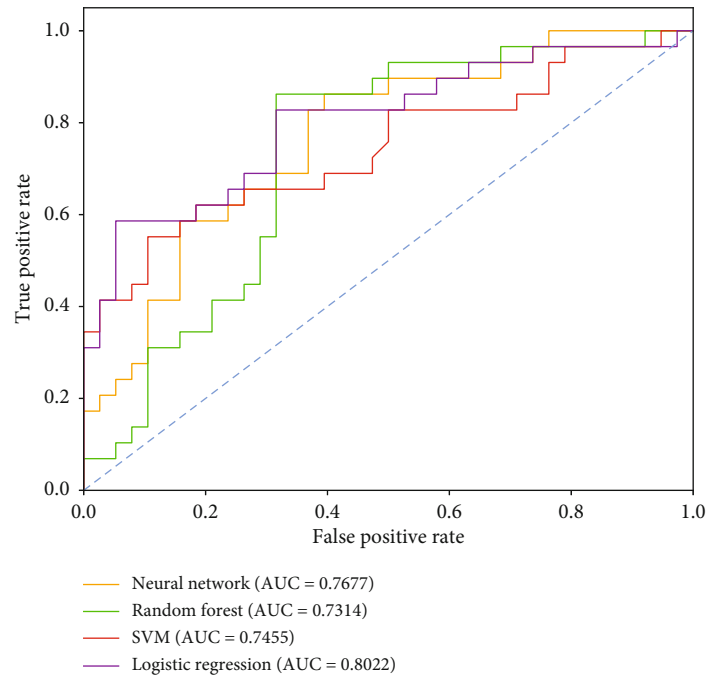


FIGURE 8: ROC curves of Qi deficiency syndrome model based on tongue and pulse.

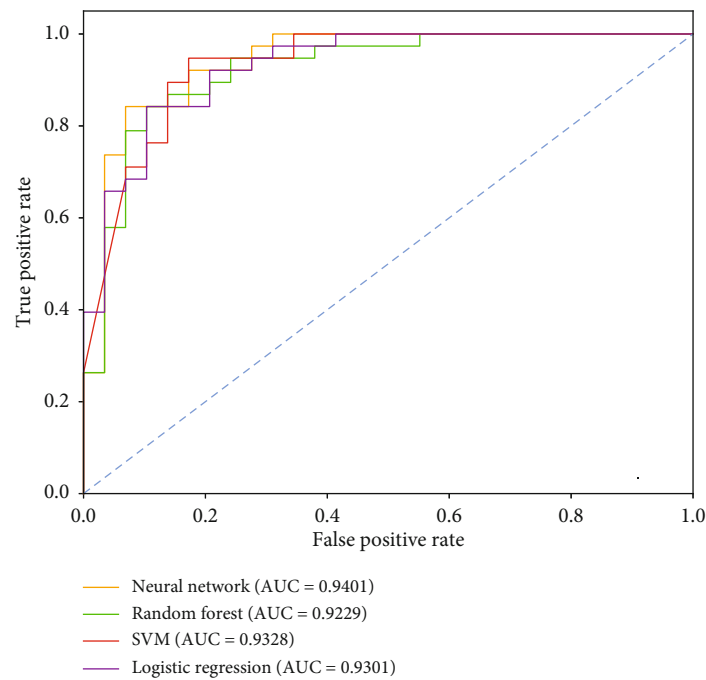


FIGURE 9: ROC curves of Qi deficiency syndrome model based on syndrome and tongue and pulse.

the decreased function of visceral organs. The main manifestations are fatigue, lack of energy, lazy speech, and weak pulse. Yin deficiency syndrome refers to the lack of Yin fluid in human body, its nourishing and nourishing functions are reduced, or Yin does not control Yang, and Yang is too hyperactive. The main manifestations are dry mouth and pharynx, dysphoria in chestpalms-soles, tidal fever, and night sweating. According to the principle of TCM syndrome differentiation and treatment, the principle and treatment method of Qi defi-

ciency syndrome are to invigorate the spleen and replenish Qi, and the corresponding prescription is Sijunzi decoction. The principle and treatment method of Yin deficiency syndrome are to nourish Yin and clear lung, and the corresponding prescription is Shashen Maidong decoction.

4.1. Statistical Analysis of Tongue Data and Pulse Data of Qi Deficiency Syndrome and Yin Deficiency Syndrome. TCM is a promising and effective adjuvant therapy in the treatment of

lung cancer. Compared with chemotherapy and radiotherapy, it has the advantages of availability, effectiveness, and low toxicity [35], although its various mechanisms deserve further study [36, 37]. In this study, the tongue parameters, including TB-a, TC-a, TB-Cr, and TC-Cr of Qi deficiency syndrome and Yin deficiency syndrome, represent the red value of tongue body and tongue coating. Larger values of tongue parameters reflect the redder or magenta tongue. In Yin deficiency syndrome, TB-a, TC-a, TB-Cr, and TC-Cr were all higher than those in Qi deficiency syndrome, indicating that the tongue of Yin deficiency syndrome was redder or magenta. S stands for saturation, and the higher the value of S, the brighter the tongue color will be. TC-S in Yin deficiency syndrome was higher than that in Qi deficiency syndrome, indicating that the tongue color of Yin deficiency syndrome was brighter. perAll is the ratio of tongue coating area to total tongue area. perAll has a higher diagnostic value for thick coating, and the higher the value, the thicker the tongue coating. perAll in Yin deficiency syndrome was lower than that in Qi deficiency syndrome, indicating that the tongue coating was thinner in Yin deficiency syndrome. Among the four parameters of texture parameters CON, ASM, ENT, and MEAN, the smaller the value of CON, ENT, and MEAN, the larger the ASM, reflecting that the more delicate the tongue texture or the more greasy the tongue coating. In this study, TC-CON and TC-ENT of Yin deficiency syndrome were significantly lower than those of Qi deficiency syndrome, while TC-ASM was higher than that of Qi deficiency syndrome, indicating that the tongue coating of Yin deficiency syndrome was greasier.

In the pulse parameters, t_4 is the time value from the starting point to the descending isthmus of the sphygmogram, corresponding to the systolic period of left ventricle, and t_5 is the time value from the dicrotic notch to the end point of the sphygmogram, corresponding to the diastolic period of left ventricle. t_4 and t_5 of Yin deficiency syndrome were smaller than those of Qi deficiency syndrome, indicating that the time of systole and diastole of Yin deficiency syndrome was shorter than those of Qi deficiency syndrome, and the pulsation cycle t of Yin deficiency syndrome also showed a decreasing trend, indicating that the pulse wave velocity of Yin deficiency syndrome was slightly higher. In addition, there was a phenomenon of elevation of Yin deficiency syndrome in dicrotic notch h_4 . Furthermore, indicrotic notch h_4 in Yin deficiency syndrome was elevated. In the Qi deficiency syndrome, h_3/h_1 , h_1/t_1 , and t_1 were prolonged, reflecting that the pulse force of the Qi deficiency syndrome was soft and weak, the amplitude of the main wave h_1 was reduced, and the area under the sphygmogram was smaller, indicating that the pulse shape was thin and small. All in all, the tongue of Qi deficiency syndrome was pale and the pulse was weak, while the tongue body of Yin deficiency syndrome was more red or crimson, more brighter in tongue color, thinner and greasy in tongue coating, and more fine in pulse.

4.2. Modeling Analysis of Qi Deficiency Syndrome and Yin Deficiency Syndrome Based on Data of Tongue and Pulse. In

recent years, with the rapid development of computer technology, different recognition algorithms and machine learning methods, such as logical regression [38], SVM [22, 39], random forest [40], and neural network [15, 41], and other data mining technologies have been widely used in medical research. The quantitative diagnosis of diagnostic information through various mathematical models has promoted the development of TCM informatization. In this study, symptom and tongue and pulse data were used to classify syndromes. The results showed that the classification efficiency of models based on different datasets was as follows: tongue and pulse < symptom < symptom and tongue and pulse, indicating that tongue and pulse data contributed to the classification of syndrome to some extent. Therefore, when faced with a complicated quantitative and qualitative, subjective and objective, determine and fuzzy, and massive TCM data combining linear and nonlinear, TCM syndrome associated with complex multidimensional characteristics and associated with multiple microindex, especially when symptoms were not evident, to explore the relationship between different syndromes and physical and chemical indices can effectively assist in syndrome differentiation. Research also showed that it was very reasonable to combine microindex with macrosymptoms. Using machine learning or data mining methods to build TCM syndrome or disease diagnosis model can make the process of syndrome differentiation and treatment more objective, standardized, and intelligent [42–44].

4.3. Limitations and Future Work. This research is based on the real-world investigation, and the results basically conform to the syndrome distribution feature of NSCLC in the clinic. However, there are also some limitations in the study. First of all, due to the limitation of time and place, the sample size of this study is not large enough. Secondly, the basic data statistics of subjects are not comprehensive enough, and there is a lack of statistics on height, weight, body mass index (BMI), history of present illness, past medical history, etc., which may affect the data results. Last but not the least, this study mainly focused on the common NSCLC syndrome of Qi deficiency and Yin deficiency, lacking more syndromes to explore. In the future, a large-scale and multicenter epidemiological investigation should be combined, the collection of four diagnostic information and basic characteristics needs to be more standardized and complete, and further researches based on more comprehensive syndrome differentiation results need to be carried out.

5. Conclusions

In conclusion, objective tongue and pulse data of NSCLC are useful for the classification of TCM syndrome, which can improve the accuracy of TCM syndrome classification to a certain extent. Tongue and pulse diagnosis parameters can provide new ideas and methods for TCM syndrome differentiation of Qi deficiency syndrome and Yin deficiency syndrome of NSCLC.

Data Availability

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

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

YS and JL drafted the initial manuscript and performed classification models; XH, LT, and JC performed data statistical analyses; JX and LX developed the study concept and design. JL, ZB, and JL provided critical revisions. All authors read and approved the final manuscript.

Acknowledgments

This research was funded by the National Key Research and Development Program of China (2017YFC1703301), the National Natural Science Foundation of China (81873235, 81973750, and 81904094), and the 1226 Major Project (BWS17J028).

References

- [1] L. A. Torre, R. L. Siegel, E. M. Ward, and A. Jemal, "Global cancer incidence and mortality rates and trends—an update," *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, vol. 25, no. 1, pp. 16–27, 2016.
- [2] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA: a Cancer Journal for Clinicians*, vol. 65, no. 2, pp. 87–108, 2015.
- [3] World Health Organization Cancer, 2020.
- [4] G. Liu, F. Pei, F. Yang et al., "Role of autophagy and apoptosis in non-small-cell lung cancer," *International Journal of Molecular Sciences*, vol. 18, no. 2, 2017.
- [5] W. D. Travis, E. Brambilla, M. Noguchi et al., "International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma," *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, vol. 6, no. 2, pp. 244–285, 2011.
- [6] "Chinese expert consensus on antiangiogenic drugs for advanced non-small cell lung cancer (2020 Edition)," *Zhonghua zhong liu za zhi Chinese journal of oncology*, vol. 42, no. 12, pp. 1063–1077, 2020.
- [7] K. D. Miller, L. Nogueira, A. B. Mariotto et al., "Cancer treatment and survivorship statistics, 2019," *CA: a Cancer Journal for Clinicians*, vol. 69, no. 5, pp. 363–385, 2019.
- [8] J. R. Choo and R. A. Soo, "Lorlatinib for the treatment of ALK-positive metastatic non-small cell lung cancer," *Expert Review of Anticancer Therapy*, vol. 20, no. 4, pp. 233–240, 2020.
- [9] K. M. Islam, T. Anggondowati, P. E. Deviany et al., "Patient preferences of chemotherapy treatment options and tolerance of chemotherapy side effects in advanced stage lung cancer," *BMC Cancer*, vol. 19, no. 1, p. 835, 2019.
- [10] S. Chen, Y. Bao, J. Xu et al., "Efficacy and safety of TCM combined with chemotherapy for SCLC: a systematic review and meta-analysis," *Journal of Cancer Research and Clinical Oncology*, vol. 146, no. 11, pp. 2913–2935, 2020.
- [11] R. Liu, S. L. He, Y. C. Zhao et al., "Chinese herbal decoction based on syndrome differentiation as maintenance therapy in patients with extensive-stage small-cell lung cancer: an exploratory and small prospective cohort study," *Evidence-based Complementary and Alternative Medicine: Ecam*, vol. 2015, article 601067, 12 pages, 2015.
- [12] S. Chen, A. Flower, A. Ritchie et al., "Oral Chinese herbal medicine (CHM) as an adjuvant treatment during chemotherapy for non-small cell lung cancer: a systematic review," *Lung Cancer (Amsterdam, Netherlands)*, vol. 68, no. 2, pp. 137–145, 2010.
- [13] M. Jiang, C. Lu, C. Zhang et al., "Syndrome differentiation in modern research of traditional Chinese medicine," *Journal of Ethnopharmacology*, vol. 140, no. 3, pp. 634–642, 2012.
- [14] S. Xia, J. Zhang, G. Du et al., "A microcosmic syndrome differentiation model for metabolic syndrome with multilabel learning," *Evidence-based Complementary and Alternative Medicine: Ecam*, vol. 2020, article 9081641, 10 pages, 2020.
- [15] X. Li, Y. Zhang, Q. Cui, X. Yi, and Y. Zhang, "Tooth-marked tongue recognition using multiple instance learning and CNN features," *IEEE transactions on cybernetics*, vol. 49, no. 2, pp. 380–387, 2019.
- [16] B. Qin, L. Liang, J. Wu, Q. Quan, Z. Wang, and D. Li, "Automatic identification of Down syndrome using facial images with deep convolutional neural network," *Diagnostics (Basel, Switzerland)*, vol. 10, no. 7, 2020.
- [17] Z. Pan, Z. Shen, H. Zhu et al., "Clinical application of an automatic facial recognition system based on deep learning for diagnosis of Turner syndrome," *Endocrine*, 2020.
- [18] X. Xia, X. Chen, G. Wu et al., "Three-dimensional facial-image analysis to predict heterogeneity of the human ageing rate and the impact of lifestyle," *Nature Metabolism*, vol. 2, no. 9, pp. 946–957, 2020.
- [19] F. Bing and L. Shaozi, "Unsupervised clustering analysis of human-pulse signal in traditional Chinese medicine," *CAAI Transactions on Intelligent Systems*, vol. 13, no. 4, pp. 564–570, 2018.
- [20] X. Wang, B. Zhang, Z. Yang, H. Wang, and D. Zhang, "Statistical analysis of tongue images for feature extraction and diagnostics," *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, vol. 22, no. 12, pp. 5336–5347, 2013.
- [21] N. D. Kamarudin, C. Y. Ooi, T. Kawanabe, H. Odaguchi, and F. Kobayashi, "A fast SVM-based tongue's colour classification aided by k-means clustering identifiers and colour attributes as computer-assisted tool for tongue diagnosis," *Journal of health-care engineering*, vol. 2017, Article ID 7460168, 13 pages, 2017.
- [22] J. Zhang, J. Xu, X. Hu et al., "Diagnostic method of diabetes based on support vector machine and tongue images," *BioMed Research International*, vol. 2017, Article ID 7961494, 9 pages, 2017.
- [23] D. E. Wood, "National Comprehensive Cancer Network (NCCN) clinical practice guidelines for lung cancer screening," *Thoracic Surgery Clinics*, vol. 25, no. 2, pp. 185–197, 2015.

- [24] E. Brambilla, W. D. Travis, T. V. Colby, B. Corrin, and Y. Shimamoto, "The new World Health Organization classification of lung tumours," *The European Respiratory Journal*, vol. 18, no. 6, pp. 1059–1068, 2001.
- [25] P. Mücke, J. S. Mattsson, D. Djureinovic et al., "The impact of the fourth edition of the WHO classification of lung tumours on histological classification of resected pulmonary NSCCs," *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, vol. 11, no. 6, pp. 862–872, 2016.
- [26] H B, "Technical guidelines for clinical research of new drugs of syndrome," *Journal of Traditional Chinese Medicine*, vol. 26, no. 21, p. 107, 2018.
- [27] Institute of TCM Diagnosis HUoT, Department of Internal Medicine CAoT, Department of Surgery CAoT et al., *Clinic terminology of traditional Chinese medical diagnosis and treatment—syndromes*, vol. GB/T 16751.2-1997, State Bureau of Technical Supervision, 1997.
- [28] Z. Jian-feng, X. Jia-tuo, T. Li-ping et al., "Study on the characteristics of sub-health symptoms and TCM syndrome patterns distribution in 1 754 non-disease population," *Chinese Journal of Integrative Medicine*, vol. 37, no. 8, pp. 934–938, 2017.
- [29] F. Schiller, M. Valsecchi, and K. R. Gegenfurtner, "An evaluation of different measures of color saturation," *Vision Research*, vol. 151, pp. 117–134, 2018.
- [30] X. Sun, J. Young, J. H. Liu et al., "Prediction of pork color attributes using computer vision system," *Meat Science*, vol. 113, pp. 62–64, 2016.
- [31] Z. Qi, L. P. Tu, J. B. Chen, X. J. Hu, J. T. Xu, and Z. F. Zhang, "The classification of tongue colors with standardized acquisition and ICC profile correction in traditional Chinese medicine," *BioMed Research International*, vol. 2016, Article ID 3510807, 9 pages, 2016.
- [32] Z. Y. Luo, J. Cui, X. J. Hu et al., "A study of machine-learning classifiers for hypertension based on radial pulse wave," *BioMed Research International*, vol. 2018, Article ID 2964816, 12 pages, 2018.
- [33] Y. Shi, X. Hu, J. Cui et al., "Clinical data mining on network of symptom and index and correlation of tongue-pulse data in fatigue population," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 72, 2021.
- [34] J. Li, P. Yuan, X. Hu et al., "A tongue features fusion approach to predicting prediabetes and diabetes with machine learning," *Journal of Biomedical Informatics*, vol. 115, p. 103693, 2021.
- [35] Y. Xiang, Z. Guo, P. Zhu, J. Chen, and Y. Huang, "Traditional Chinese medicine as a cancer treatment: modern perspectives of ancient but advanced science," *Cancer Medicine*, vol. 8, no. 5, pp. 1958–1975, 2019.
- [36] X. L. Su, J. W. Wang, H. Che et al., "Clinical application and mechanism of traditional Chinese medicine in treatment of lung cancer," *Chinese Medical Journal*, vol. 133, no. 24, pp. 2987–2997, 2020.
- [37] J. Leng, L. Lei, S. F. Lei, Z. Zhu, A. Ocampo, and F. Gany, "Use of traditional Chinese herbal medicine concurrently with conventional cancer treatment among Chinese cancer patients," *Journal of Immigrant and Minority Health*, vol. 22, no. 6, pp. 1240–1247, 2020.
- [38] K. Zhang, W. Geng, and S. Zhang, "Network-based logistic regression integration method for biomarker identification," *BMC Systems Biology*, vol. 12, Suppl 9, p. 135, 2018.
- [39] C. Liu and Y. Cheng, "An application of the support vector machine for attribute-by-attribute classification in cognitive diagnosis," *Applied Psychological Measurement*, vol. 42, no. 1, pp. 58–72, 2018.
- [40] Y. Kong and T. Yu, "A deep neural network model using random forest to extract feature representation for gene expression data classification," *Scientific Reports*, vol. 8, no. 1, p. 16477, 2018.
- [41] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, 2017.
- [42] S. Xia, J. Cai, J. Chen et al., "Factor and cluster analysis for TCM syndromes of real-world metabolic syndrome at different age stage," *Evidence-based Complementary and Alternative Medicine: Ecam*, vol. 2020, article 7854325, 10 pages, 2020.
- [43] Y. Gu, Y. Wang, C. Ji et al., "Syndrome differentiation of IgA nephropathy based on clinicopathological parameters: a decision tree model," *Evidence-based Complementary and Alternative Medicine: Ecam*, vol. 2017, article 2697560, 11 pages, 2017.
- [44] H. Kang, Y. Zhao, C. Li et al., "Integrating clinical indexes into four-diagnostic information contributes to the traditional Chinese medicine (TCM) syndrome diagnosis of chronic hepatitis B," *Scientific Reports*, vol. 5, no. 1, p. 9395, 2015.

Research Article

Doctor Recommendation Model Based on Ontology Characteristics and Disease Text Mining Perspective

Chunhua Ju ¹ and Shuangzhu Zhang ²

¹Business Administration College, Zhejiang Gongshang University, Hangzhou, China

²School of Management Science and Engineering, Zhejiang Gongshang University, Hangzhou, China

Correspondence should be addressed to Shuangzhu Zhang; 16010000028@pop.zjgsu.edu.cn

Received 21 May 2021; Accepted 20 July 2021; Published 9 August 2021

Academic Editor: Yue Zhang

Copyright © 2021 Chunhua Ju and Shuangzhu Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Patients can access medical services such as disease diagnosis online, medical treatment guidance, and medication guidance that are provided by doctors from all over the country at home. Due to the complexity of scenarios applying medical services online and the necessity of professionalism of knowledge, the traditional recommendation methods in the medical field are confronting with problems such as low computational efficiency and poor effectiveness. At the same time, patients consulting online come from all sides, and most of them suffer from nonacute or malignant diseases, and hence, there may be offline medical treatment. Therefore, this paper proposes an online prediagnosis doctor recommendation model by integrating ontology characteristics and disease text. Particularly, this recommendation model takes full consideration of geographical location of patients. **Objective.** The recommendation model takes the real consultation data from online as the research object, fully testifying its effectiveness. Specifically, this model would make recommendation to patients on department and doctors based on patients' information of symptoms, diagnosis, and geographical location, as well as doctor's specialty and their department. **Methods.** Utilizing crawler technique, five hospital departments were selected from the online medical service platform. The names of the departments were in accordance with the standardized department names used in real hospitals (e.g., endocrinology, dermatology, gynecetrics, pediatrics, and neurology). As a result, a dataset consisting of 20000 consultation questions by patients was built. Through the application of Python and MySQL algorithms, replacing semantic dictionary retrieval or word frequency statistics, word vectors were utilized to measure similarity between patients' prediagnosis and doctors' specialty, forming a recommendation framework on medical departments or doctors based on the above-obtained sentence similarity measurement and providing recommendation advices on intentional departments and doctors. **Results.** In the online medical field, compared with the traditional recommendation method, the model proposed in the paper is of higher recommendation accuracy and feasibility in terms of department and doctor recommendation effectiveness. **Conclusions.** The proposed online prediagnosis doctor recommendation model integrates ontology characteristics and disease text mining. The model gives a relatively more accurate recommendation advice based on ontology characteristics such as patients' description texts and doctors' specialties. Furthermore, the model also gives full consideration on patients' location factors. As a result, the proposed online prediagnosis doctor recommendation model would improve patients' online consultation experience and offline treatment convenience, enriching the value of online prediagnosis data.

1. Introduction

As the emphasis of medical care gradually shifts from disease to patient, the role of patients' participation in online health improvement is becoming more prominent. The health service in the world is not only different in terms of regions but also varying in terms of online health services [1, 2]. Spe-

cifically, there exist phenomenon such as information asymmetry between doctors and patients and unequal distribution of medical resources geographically [3]. Therefore, patients registering doctors online and intelligent department recommendation have also become one of the important topics of medical informatization. According to a report released in 2019 by the Big Data Research Institute, the scale of users

in China's medical and health market was about 800 million by the end of 2018 [4]. With a large number of doctors and patients interacting online, a large amount of real consultation data has been accumulated in the online health community. Therefore, it is of important theoretical and practical value to investigate how to make full use of online data to build models to improve patients' medical treatment experience in terms of increasing the accuracy of patients' medical choice and the effectiveness of department recommendation.

The existing literature has been conducting studies from perspectives of department recommendation and doctor recommendation. The two methods of department recommendation are separately based on expert system and similarity calculation. As for department recommendation based on the expert system, on one hand, through establishment of medical knowledge base with the help from medical experts, the diagnosis process of medical experts is simulated by applying rule-based reasoning engine. As a result, patients' diseases are predicted, so as to achieve the target department recommendation for patients. Moreover, the expert-based department recommendation is built upon fuzzy logic and RBF neural network, effectively improving the recommendation accuracy [5, 6]. On the other hand, there exist many problems due to the abundant number of reasoning rules, such as low computational efficiency and high maintenance cost of knowledge base. As for department recommendation based on similarity calculation, the current literature uses various methods to measure similarities, such as similarity between patients' symptoms and disease' symptoms [7], TF-IDF sentence-based similarity and TF-IDF algorithm that is based on multiple words [8, 9], combination of focus shifting backwards, and professional medical corpus [10]. This similarity-based recommendation would, respectively, calculate the possibility of having disease and descriptive words that may correspond with certain symptoms, realizing the goal of department recommendation to patients. Research of recommendation on doctor is mainly based on the content and collaborative filtering recommendation algorithm, focusing on user keywords, browsing history, evaluation, and other data [11, 12]. The user collaborative filtering algorithm assumes that one user and other user group who share similar interest would have same product preference [13–15]. Among them, user collaborative filtering algorithm integrating projects mainly solves the problem of information overload through filtering attribute collaboratively [16]. Moreover, the application of customized relational network and tags solves the problem of data sparsity in the matrix factorization recommendation model [17, 18], and the collaborative filtering recommendation method integrates contextual perception, project similarity, and user behavior, giving recommendation results from perspectives of patients' contexts, projects, and user participation [19–21]. In addition, scholars also conducted modeling research on doctor recommendation, disease diagnosis, and medical examination [22, 23] from the perspectives of semantic characteristics of medical resources [24], user information types [25], user ratings, and comment portraits [26], as well as Bayesian algorithm [27].

The recommendation algorithms in the traditional medical field mainly have the following three problems. First, in

terms of department recommendation, the algorithm based on the expert system causes problems such as explosion of knowledge rule reasoning and high maintenance cost of knowledge base. Furthermore, the algorithm based on similarity may not effectively recognize synonyms, possibly decreasing recommendation accuracy. Second, in terms of doctor recommendation, the user-based collaborative filtering algorithm may cause problems that patients of similar symptoms would not be diagnosed with the same disease, due to complexity and diversity of diseases. What is more, because of the nonnecessary relationship among patients' etiologies, the assumption of the project-based collaborative filtering algorithm that users would choose doctors with the same research field as their previous doctors may hardly be met. Third, although relevant literatures have studied how to reduce data sparsity [28–30], the collaborative filtering recommendation algorithm still cannot completely avoid the performance problems caused by data sparsity.

Based on the above theorization, it can be concluded that the existing recommendation algorithms cannot fully meet requirements with regard to recommendation in the context of the Internet medical field. Patients can access medical services provided by doctors in the online health community all over the country online without going out, including disease diagnosis, medical treatment guidance, and medication guidance. Meanwhile, patients consulting online come from far and near and may involve situations of offline medical treatment, making it necessary to take into account the factor of patients' location. Therefore, this paper proposes an online prediagnosis doctor recommendation model that integrates ontology characteristics and disease text mining, improving both the effectiveness of doctor recommendation within the environment of online medical service and the convenience of offline medical treatment for patients.

2. Research on the Doctor Recommendation Model

The doctor recommendation model is mainly divided into three steps. Step 1: data preprocessing. Perform word segmentation and stop word removal with regard to patient's input of natural language. Step 2: hospital department recommendation. After screening patients' query data, create the most similar sentence set based on key parts of word vector or the similarity measurement for symptom descriptions, so as to achieve department recommendation. Step 3: doctor recommendation. Use SQL sentence query in the MYSQL database to complete doctor recommendation (Figure 1).

3. Data Cleaning Process

There are mainly two aspects of data that are available online. The first aspect of data is patients' online consultation regarding disease symptom. This source of information mainly covers age, gender, symptom description, and other data. The second aspect of data is doctors' information online, including doctors' names, titles, hospitals, departments, and their specialties as shown in Table 1. All data is in structured form, and information such as disease

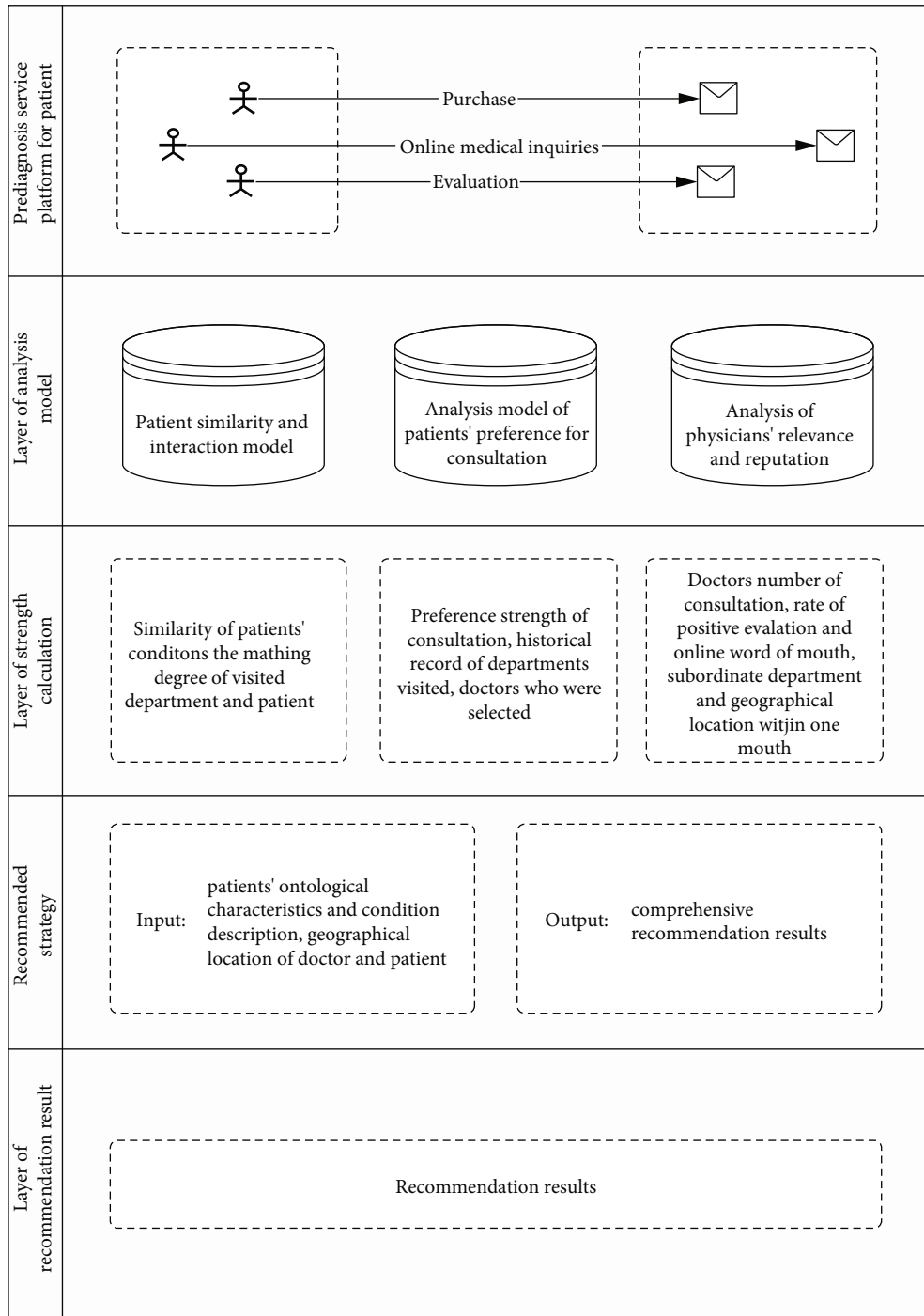


FIGURE 1: Prediagnosis doctor recommendation model integrating ontology characteristics and disease text mining.

description, prediagnosis, and specialties are stored in text form. Then, model will be built after word segmentation and keyword extraction (Figure 2).

4. Data on Ontology Characteristics of Doctors and Patients

The doctor-patient demographic data obtained from WeiYi platform are mostly well-organized semistructured textual data. The first step is to transform unstructured text data into

structured text data through named entity recognition and information extraction. Organization names, people's names, and location names can be recognized by applying multiple open source Chinese language processing tools [31], such as fudanNLP developed by Fudan University [32], NLPiR word segmentation system developed by Chinese Academy of Sciences [33], and LTP Chinese natural language processing platform of Harbin Institute of Technology [34]. In addition, delete the missing value and duplicated information. And, for the problem of different doctors sharing one same name, use

TABLE 1: Data sample on patients and doctors online.

Patient ID	Gender	Age	Province/city	Main complaint	Initial consultation department online
8070844	Female	65	Jiangsu	Menstruation keeps coming. B-ultrasound result shows that my endometrium is thick. I ate progesterone and did curettage. For now, I have been taking medicines for 10 days. 3 days after progesterone, I still had large amount of blood flow, and my stomach ached. I am wondering what is wrong with me.	Gynecology
81305510	Female	42	Guangdong	Bilateral hydrosalpinx. I never had abortion history. I want to be pregnant now, what should I do now?	Gynecology
12031251	Female	43	Heilongjiang	43-year-old, irregular menstruation for many years, 3 times for 2 months, the period was long for 7/8 days, the amount is little, and the color is dark brown. What medicine should I take?	Gynecology
57715499	Female	37	Henan	Just had miscarriage a month ago; yet, I got pregnant in confinement. Can I keep the child?	Gynecology
72520784	Female	53	Shanghai	My mother is 53 years old. She feels nervous, unable to breathe, cannot lie down, and feels no strength.	Neurology
Doctor name	Title	Hospital	City	Specialties	Department
Niu**	Chief Physician	Ningbo First Hospital	Ningbo	Diagnosis and treatment of diabetes and thyroid disease	Endocrinology
Yang*	Associate Chief Physician	Shijiazhuang First Hospital	Shijiazhuang	Hemorrhagic cerebrovascular disease such as cerebral aneurysm, arteriovenous malformation, arteriovenous fistula, and cavernous hemangioma; ischemic cerebrovascular diseases such as carotid artery stenosis, vertebral artery stenosis, intracranial artery stenosis, and moyamoya disease	Neurosurgery
Xu**	Chief Physician	Beijing Anzhen Hospital, Capital Medical University	Beijing	Diagnosis, surgical treatment, and perioperative treatment of various congenital heart diseases	Pediatric cardiac surgery
Wang**	Associate Chief Physician	Shenzhen Bao'an People's Hospital	Shenzhen	Diagnosis and treatment of diabetes and its complications, hyperthyroidism, and hypothyroidism; use of insulin pump and dynamic blood glucose monitors	Endocrinology
Liu**	Chief Physician	Hospital of Traditional Chinese Medicine in Uygur, Xinjiang	Xinjiang	Neurology of traditional Chinese medicine	Neurology

* and ** mean one word or two words for the Chinese name.

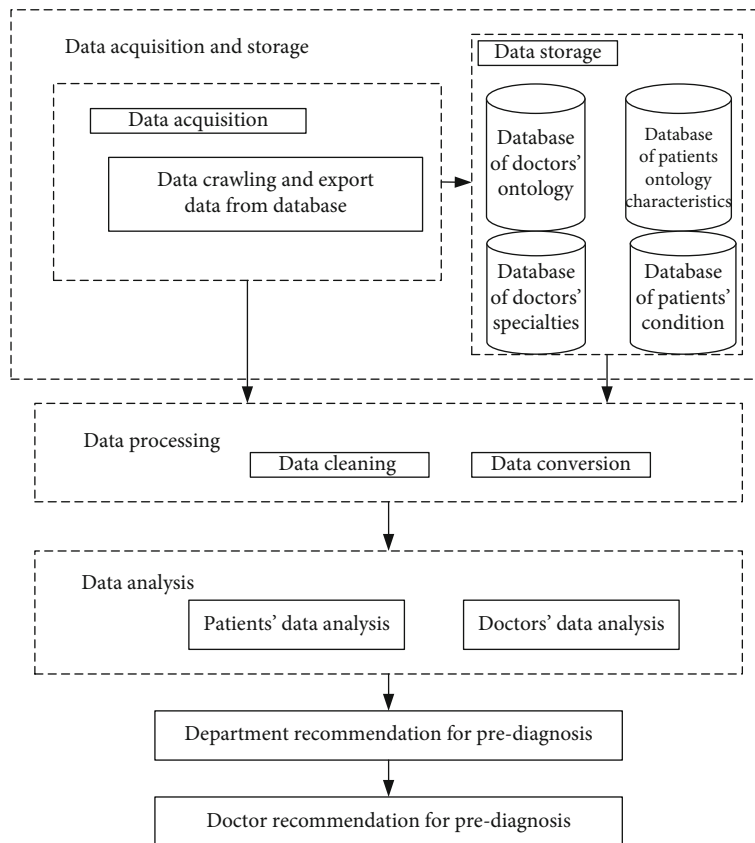


FIGURE 2: Data cleaning process.

fields such as “the hospital to which they belong” and “the department to which they belong” to restrict.

5. Data on Patients' Condition Description

Data on patients' online condition description are presented as specific evaluations expressed by patients in natural language. The data in its initial form are fulfilled with problems that the contents are nonstandardized, repetitive, short, and single [35]. The authors marked the text content by part of speech and synonyms and then use human tissue lexicon and human anatomy lexicon to match the word segmentation results so as to extract disease symptoms and keywords of human body parts. As shown in Table 1, the patient's main complaint was that “it was caused by pelvic effusion eight years ago, there was no abortion history and no pregnancy.” The common clinical symptoms that the patient did not actually have appeared in the description make it difficult to extract keywords. For example, “no abortion history” was divided into “no” and “abortion history,” resulting in the extraction of “abortion history” as the keyword; yet, the patient did not have these symptoms. To deal with situations like the abovementioned, before word segmentation, the authors would divide the description paragraph into short sentences or phrases by punctuation marks, and the stop words should be retained in word segmentation. Then, while extracting keyword, the target words cannot be consid-

ered as the real target keywords if they contain negative modifiers such as none, unaccompanied, and no.

6. Data on Doctors' Specialties

Data on doctors' specialties are structured textual data and are confronted with problems of synonymous naming and missing data. An example of synonymous naming refers to the problem that doctors in different hospitals have different naming for their fields of expertise. Specifically, synonyms for fields of expertise are specialties, being good at, specializing in, being skilled in, being professional with, medical interest, and research direction. All synonymous naming shall be integrated into the same field. As for the problem of missing data, utilize multiple data source data integration to complete improvement or deletion.

7. Doctor Recommendation

7.1. Department Recommendation. For questions input by patients, every keyword for each sentence can be obtained after word segmentation and word stopping removal. Next, the corresponding question set can be obtained by positioning question sentences that are associated with each keyword. The authors divided the question set into sample dataset and test dataset, both containing information of patients' condition description text, online prediagnosis department recommendation, etc. Then, use the word2vec library to train a

```

The code is as the follows:
with open("test.txt") as f:
    document = f.read()
    document_cut = jieba.cut(document)
    result = ".join(document_cut)
    result = result.encode("utf-8")
    with open("test2.txt", "w") as f2:
        f2.write(result)
f.close()
f2.close()

```

ALGORITHM 1: This module preprocesses the sample dataset using the following code. The aim is to segment words, remove stop words, and retain key parts or key symptoms with regard to patients' condition description online.

```

The code is as the follows:
logging.basicConfig(format = "%(asctime)s: %(levelname)s: %(message)s", level = logging.INFO)
sentences = word2vec.LineSentence("test2.txt")
model = word2vec.Word2Vec(sentences, hs = 1, min_count = 1, window = 3, size = 100)
model.save("pifuke.model")

```

ALGORITHM 2: This module used the word2VEC library to train the word vector model of dermatology on sample data such as "dermatology.XLS."

```

The code is as the follows:
#Note: Load the department's word vector model
model_1 = word2vec.Word2Vec.load("pifuke.model")
for strZhengZhuang in symptom word set of a certain patient
    try:
        sim3 = model_1.most_similar(strZhengZhuang, topn = 20)
        if sim3.__len__() > 0:
            return 1
    except:
        return 0
#Note: Number of matching words/total number of symptom words for a patient = Matching probability
probability = Words/WordsCount

```

ALGORITHM 3: The module mainly had two goals to achieve. First, preprocess the test data, including word segmentation and stop word removal, and retaining key parts or symptoms for the disease description. Second, compare the word vectors of test data and that of the training results, and the departments with high similarity were recommended to patients.

word vector model on the keywords of the sentences in the sample data set, calculate the similarity between the questions input by the patient in the test data set and the word vector model of the sample data set, and lastly select the most similar questions to the sample data set in the test dataset. Following the rule that higher similarity indicates the same one department, after screening the similarity calculation one by one, the department with the highest similarity would be the final recommendation result.

8. Doctor Recommendation

The core significance of the development of online medical and health services is to reshape the medical service process and optimize the allocation of medical resources, so as to meet the medical and health needs of individual consumers.

Due to its mobility, convenience, rapidness, personalization, and interaction, the online medical services have become the main channel for consumers to seek medical help online, having been adopted and utilized by consumers. To some extent, it alleviates the medical pressure and realizes the optimal allocation of medical resources. The patients using online medical service come from all sides, and the majority of them have conventional and chronic diseases, making it sometimes necessary for patients to confirm their diagnosis offline. Therefore, doctor recommendation that takes into account of patients' location information is particularly important to improve patients' convenience of offline medical treatment and to attract more patients to use online medical services. Based on the SQL statements query function in the MYSQL database, matching keywords with doctors' specialties, department, and region information, integrating

patients' location information, and this paper recommends local doctors that meet the requirements according to patients' region. For instance, a patient's naming Zhang San, living in Zhejiang province, with condition described as thick endometrium, heavy menstrual flow, and stomach-ache, would be recommended to see a Chief Physician from Department of Gynecology at Zheyi hospital with family name of Wang.

9. Sentence Similarity

9.1. Calculation of Similarity Based on Postcontent. After obtaining the unique d -dimensional distribution vector representation of the disease description text content, the similarity and distance between each two text contents can be obtained through similarity calculation. The author uses the cosine formula to measure the similarity between two texts and uses the Mahala Nobis distance to calculate the natural language description of the two posts. Assume that two paragraph vectors of natural language description of text content are expressed as $PV_a = (\times 11, \times 12, \dots, \times 1d)$ and $PV_b = (\times 21, \times 22, \dots, \times 2d)$, where d represents two paragraph vectors. The similarity and distance are defined as follows:

$$\begin{aligned} \text{sim}(PV_a, PV_b) &= \frac{PV_a \bullet PV_b}{\|PV_a\| \cdot \|PV_b\|}, \\ &= \frac{\sum_{i=1}^d x_{1i} x_{2i}}{\sqrt{\sum_{i=1}^d x_{1i}^2} \sqrt{\sum_{i=1}^d x_{2i}^2}}, \quad (1) \\ \text{dis}(PV_a, PV_b) &= \sqrt{(PV_a - PV_b)^T S^{-1} (PV_a - PV_b)}, \end{aligned}$$

where S is the covariance matrix of eigenvectors PV_a and PV_b .

9.2. TF-IDF Sentence Similarity Based on Co-Occurring Words. This method believes that in two sentences, the more the same vocabulary, the higher the similarity of the two sentences^[36]. Specifically,

$$\begin{aligned} \text{SimScore}(S_1, S_2) &= \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \sum_{w_i \in S_1 \cap S_2} \text{weight}(w_i), \\ \text{weight}(w_i) &= \frac{\text{Num}(w_i, k)}{N_k} \times \log \left(\frac{N_t}{\text{Num}(w_i, t) + 1} \right). \quad (2) \end{aligned}$$

Among them, $|\cdot|$ is the cardinality of the set, S_1 and S_2 are the word sets of the two sentences to be compared, w_i represents the symptom word i in the department question and answer sentence, $\text{weight}(w_i)$ is the TF-IDF^[37] weight, $\text{Num}(w_i, k)$ represents the number of sentences in which the symptom word w_i appears in the question and answer sentence set of department k , N_k represents the number of all questions and answers in department k , N_t represents the total number of questions and answers in the knowledge base, and $\text{Num}(w_i, t)$ represents the total number of ques-

TABLE 2: Summary of the characteristics of the collected data records ($N = 20000$).

Characteristic	Value, n (%)
Gender	
Male	4540 (33.7)
Female	15460 (77.3)
Age (years)	
25-30	1586 (7.9)
31-45	16800 (84.0)
46-50	1014 (5.1)
>55	600 (3.0)
Physician's professional title	
Resident physician	2670 (13.35)
Attending physician	4330 (21.65)
Associate chief physician	8040 (40.2)
Chief physician	4560 (22.8)
Other	400 (2.0)
Hospital's ranking level	
3A	19400 (97.0)
Other	600 (3.0)

tions and answers in the knowledge base. The number of sentences in which the symptom word i appears in the question. The TF-IDF sentence similarity calculation method based on co-occurring words belongs to the surface structure analysis method. It simply uses the surface information of the sentence, that is, the word frequency, part of speech, and other information of the words in the sentence to calculate the sentence similarity, without considering synonyms. This results in a decrease in the accuracy of sentence similarity.

9.3. Sentence Similarity Method Based on Word Vector. Word vector sentence similarity is mainly used indepth learning tool `word2vec`^[38] to process words into vectors and obtain the semantic similarity of sentence pairs to be compared by calculating the similarity between vectors. The specific formula is as follows:

$$\begin{aligned} \text{CosSim}(w_i, w_j) &= \frac{\sum_{i=1}^n (x_i, y_i)}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}}, \\ \text{SimScore}(S_1, S_2) &= \frac{\sum_{w \in IR} \beta w \text{MaxSimValue}(\text{CosSim}(w, IR))}{\sum_{w \in IR} \beta w}. \quad (3) \end{aligned}$$

Among them, $IR = S_1 \cup S_2$, w_i and w_j are the two words to be compared, which represent the words in sentence S_1 and the words in sentence S_2 , respectively; n represents the dimension of the word vector, and x_i and y_i represent the word vector of w_i , and the vector value of the i th dimension of the word vector of w_j ; $\text{MaxSimValue}(\text{CosSim}(w, \cdot))$ represents the maximum value of the cosine similarity between the word vector corresponding to word w and the word vector corresponding to all vocabulary of another sentence;

TABLE 3: Word vector model and keyword examples.

Word vector-based model	Keyword set	Department
	Headache, nausea, right eye, swelling, stuffy nose, right ear, tinnitus, etc.	Neurology
Keyword set	1. Migraines, nausea, loss of appetite 2. Headache, dizziness, protrusion of left eye, congestion of eyeball 3. Head distension, stuffiness, dizziness, palpitation, and restlessness 4. Palpitations and palpitations 10. Weak right hand, unable to clench a fist, palpitation, unable to breathes	

parameter β_w is The TF-IDF weight value of word w in the sentence. The greater the value of SimScore (S1, S2), the greater the similarity between the two sentences and the closer the semantics.

10. Experiment

10.1. The Data Set. To analyze the doctor recommendation method proposed in this paper, an experimental study was conducted. The data of five most common departments were crawled from the well-known domestic medical online platform-WeiYi. The names of the departments were in accordance with the standardized department names used in real hospitals (e.g., endocrinology, dermatology, gynecometrics, pediatrics, and neurology). As a result, a dataset with name of T consisting of 20000 patients' preclinical data online were built. To conduct experimentally comparative analysis of various algorithms, two widely used evaluation indexes for the recommendation performance were adopted in this paper, being accuracy rate (being P) and recall rate (being R):

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \quad (4)$$

10.2. Parameter Setting. In the experiment, the dimension parameter of the word vector was set as 100. With regard to the calculated similarity results of keyword set that would be used for department recommendation, take the top 5 questions with the highest sentence similarity as the recommended result data (top $N = \text{top } 5$), and the threshold value of keyword set similarity was set as 0.8; that is, when keyword and test set data were used for keyword similarity calculation, the result must exceed 0.8 to be included in the hospital department recommendation set. If there were 2 or more recommended hospital departments, it would be considered as no recommendation, being a special case.

11. Results and Analysis

Among the 20000 patients surveyed, 16170 were female (77.3%). This may be because women are often required to care of family health and other responsibilities in addition to work; also, women tend to pay more attention to health information than men. A total of 16800/20000 patients (84.0%) were 30 to 45 years of age. Because of old men with limited experiences in consulting physicians and obtaining medicines and children that cannot master online counseling

TABLE 4: Comparison of accuracy and recall rate.

Algorithm method	Accuracy rate (%)	Recall rate (%)
Word vector-based	74	78
Content-based	63	67
Co-occurring word-based	54	56

skills, so, old men and children may not frequently consult physicians on the internet or ask their family members to perform online inquiries. In the 20000 records, 12600 of the physicians (63.0%) are chief physicians or associate chief physicians, while 19400 hospitals (97.0%) were ranked 3A (see Table 2). In order to verify the feasibility and effectiveness of the proposed recommendation algorithms for department and doctor, the experiment was conducted to compare them with the content-based recommendation algorithm and user-based collaborative filtering algorithm. First, randomly extract 100 pieces of data from the dataset T based on the hospital department name and then perform word vector training. After the process of word segmentation and stop word removal for data of different departments, the keyword set was obtained, and the word vector model was trained using this keyword set (see Table 3). The word vector model consisted of patients' real consultation questions, and the other words excluding those questions within the group were considered as noise words, representing meaningless words unrelated to patient's consultation. Three different algorithms were all used to measure similarity for keywords to give hospital department recommendation (see results of three algorithms in Table 4).

Seen from Table 4, the proposed similarity recommendation method in this paper that incorporates ontology features and disease text data mining was the best when applied to consultation about selecting appropriate hospital department since the accuracy rate and recall rate were much higher than the other two algorithms. This is because the word vector sentence similarity measurement strategy can better measure the semantic similarity of sentences. For example, for sentence pairs "I went to the hospital to see the dentist and went home, dizzy, heavy head, runny nose" and "When I came back from the dentist, I started to feel Dizziness with symptoms of heavy head and runny nose". If a co-occurring word-based measurement method based on co-occurrence words is used, the similarity value is low, because the sentence pair contains such things as (dizziness, dizziness), (heavy head, sinking head), and (runny nose, runny nose). Synonym pairs such as clear nose) make the content-based

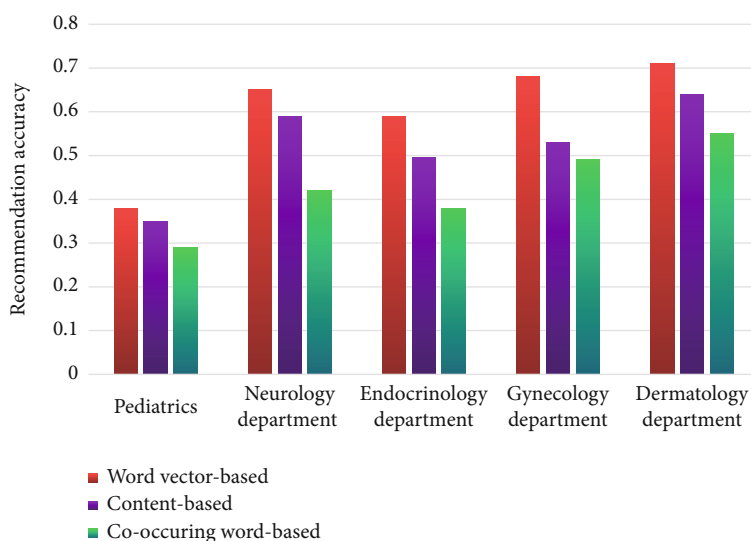


FIGURE 3: Recommendation accuracy comparison of different departments.

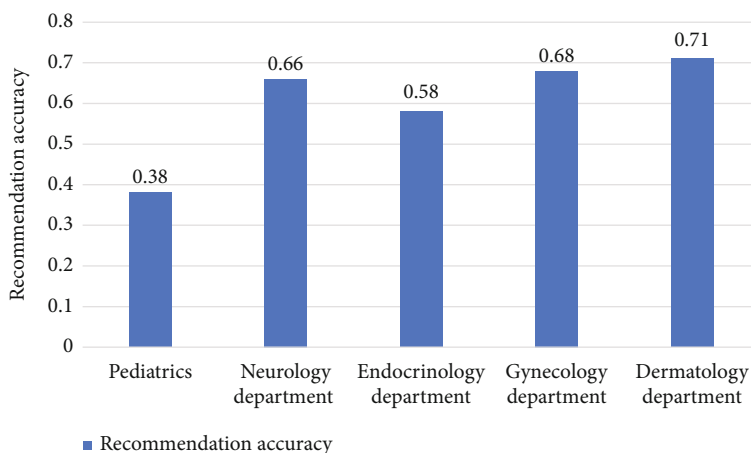


FIGURE 4: Comparison of recommendation rates of various departments.

method relatively good, and the word vector method has the best effect, indicating that it can more accurately capture the underlying semantics of the sentence. On one hand, this is because the method in this paper can measure the similarity of keywords better. For instance, keywords of “headache, palpitation, insomnia” and keywords of “head distension and restlessness” were considered as similar. The results were better than the sentence similarity measurement based on collocates. On the other hand, the proposed method in this paper took fully consideration of factors such as location information of doctors and patients, as well as doctors’ expertise field, which would not be the case for the content-based recommendation method that only takes the patient’s disease information into account.

Seen from Figures 3 and 4, the recommendation performance of the word vector method was varying for different hospital departments. The recommendation accuracy of pediatric department was below 0.5, and that of neurology, endocrinology, gynecology, and dermatology departments

were all above 0.5, among which the recommendation accuracy of gynecology was the most improved. With regard to the four departments with relatively higher recommendation accuracy, including neurology, obstetrics, gynecology, and dermatology, what they had in common was that the characteristics of the consultation questions were very typical and obvious. For example, high blood sugar, sudden weight loss, and thirst are typical for endocrinology; red rash, circular rash, redness, swelling, and itching are typical for dermatology; pregnancy and irregular menstruation are typical for gynecology. However, the situation is different for pediatric department in that if information indicating age such as baby, child, and 6 months old is not included in the consultation, it may lead to the systematic recommendation to other departments, reducing the accuracy accordingly.

Finally, The SQL statement query function in the MYSQL database used to integrate the patient’s regional factors. According to the patient’s region, we use the department and regional keyword matching and recommend the doctors

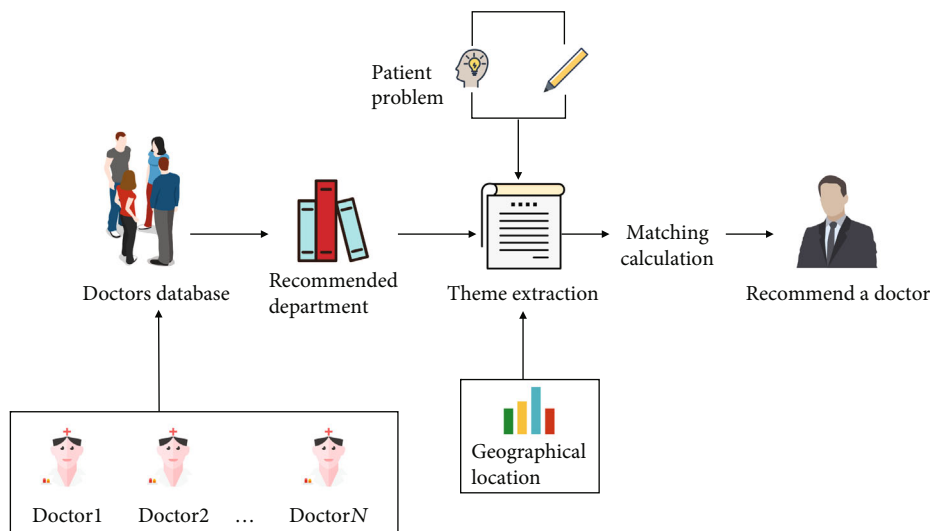


FIGURE 5: Doctor recommendation framework.

in the hospital to patient in the region that meet the needs, such as “Zhang San, from Zhejiang, the condition is described as uterus Thick intima, heavy menstrual flow, and stomachache,” and the recommended doctor is “Zhejiang First Hospital-Gynecology-Dr. Wang (Chief Physician).” The process is shown in Figure 5.

12. Conclusion

Traditional manual medical guidance is increasingly unable to meet the people’s medical needs, registration is difficult, and the problem of not finding a clinic has become increasingly prominent. Aiming at the shortcomings of traditional medical department recommendation research methods and factors such as the necessity for professional medical diagnosis expertise and information asymmetry between doctors and patients makes it impossible for patients to identify the appropriate clinic room or doctors. Once mistakes are made, online consultation time would be wasted, increasing the cost of hospitals and patients when the patient goes offline instead for medical treatment. In this paper, the proposed online prediagnosis doctor recommendation model integrates ontology characteristics and disease text mining. The experimental process uses real data on the Internet medical comprehensive website and is similar to the sentence based on content based, and based on collocate based is compared; the experiment verifies the reliability and effectiveness of the method in this paper. This provides great convenience for patients to seek medical treatment and at the same time reduces medical costs. It gives a relatively more accurate recommendation advice based on ontology characteristics such as patients’ description texts and doctors’ specialties. As a result, the proposed online prediagnosis doctor recommendation model improves patients’ online consultation experience and offline treatment convenience, enriching the value of online prediagnosis data. In addition, the primary real data from the online medical consultation platform were utilized to verify the reliability and effectiveness of the proposed method.

13. Limitations

It is not without limitation in this paper. First of all, this study was only carried out based on data from one online medical community, rendering its generalizability a question. Future study may consider collecting data from multiple online medical community platforms to verify the recommendation effect of the proposed algorithm. Second, considering that this study is solely focused on the proposed recommendation model for Chinese patients, similar studies shall be carried out in Western background in the future. Third, because of the complexity of the medical domain knowledge, follow-up researches shall not only incorporate techniques such as semantic analysis and sentiment analysis to expand the sample into general practice data but also consider introducing users’ other behavioral information to introduce the user information behavior factor optimize the target object, for intelligent department recommendation tasks, in addition to controlling data quality and deep learning algorithms such as LSTM shall be applied to improve model accuracy in the future. The intelligent department recommendation task can also be abstracted as a multilabel classification task for texts. Accordingly, multiple department categories can be recommended for patients’ questions covering multiple departments, etc. to further improve the accuracy of the proposed recommendation model, expecting to apply it to more online medical consultation platforms.

Data Availability

The data were collected with help from the administrator of the WeiYi platform. Due to third-party rights, patient privacy, and commercial confidentiality, data is not open source.

Ethical Approval

The data in this paper is divided into two parts. One part is the information crawled from the platform, such as patient

comments and doctor profiles. This kind of information is open to the public and everyone can use computer technology to obtain it on the platform. The other part is the patient's age, gender, geographical location, and other information provided by the microdoctor. The WEI-Yi platform is one of the hundreds of online medical platforms in China, with tens of thousands of registered hospitals, registered doctors, and hundreds of thousands of patients using the platform. The platform itself has a sound risk control system, and we have also signed a confidentiality agreement with the platform to define the scope of data use.

Disclosure

The paper was published in a reduced version at the IEEE 6th International Conference on Big Data Analysis (ICBDA) in 2021.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

SZ and CJ refined the topics and methods at the initial stage of paper writing. Then, SZ conducted the statistical analysis and wrote the paper under the guidance of CJ. Both authors reviewed, revised, and approved the final draft.

Acknowledgments

This project was funded by grants from the National Natural Science Foundation of China: Research on Consumer Credit Value Measurement Integrating Online Social Relationships in eCommerce (71571162). The data were collected with help from the administrator of the WeiYi platform. The data were collected with help from the administrator of the WeiYi platform.

References

- [1] Y. Balarajan, S. Selvaraj, and S. Subramanian, "Health care and equity in India[J]Health care and equity in India," *The Lancet*, vol. 377, article 9764, pp. 505–515, 2011.
- [2] J. M. Goh, G. Gao, and R. Agarwal, "The creation of social value: Can an online health community reduce rural-urban health disparities?," *MIS Quarterly*, vol. 40, no. 1, pp. 247–263, 2016.
- [3] J. Pan and D. Shallcross, "Geographic distribution of hospital beds throughout China: a county-level econometric analysis," *International Journal for Equity in Health*, vol. 15, no. 1, p. 179, 2016.
- [4] <https://data.iimedia.cn/page-category.jsp?nodeid=13210377>.
- [5] H. Bo, *Design and realization of AISCOP guiding system built in knowledge base*, SUZHOU:SoochowUniversity, 2006.
- [6] H. Ru, *The design and implementation of the guidance system based on the reasoning algorithm*, H. E. FEI, Ed., Anhui University, 2016.
- [7] L. Huang, *Research on the intelligent medical guide system based on multi-words TF-IDF algorithm*, Zhengzhou University, Zhengzhou, 2015, <https://d.wanfangdata.com.cn/thesis/D646806>.
- [8] C. Ju and S. Zhang, "Research on doctor recommendation model for Pre-Diagnosis online based on Big data Mining," in *2021 IEEE 6th International Conference on Big Data Analysis (ICBDA 2021)*, 2021.
- [9] C. Chuan-Peng and W. Zhi-Gang, "A method of sentence similarity computing based on Hownet," *Computer Engineering and Science*, vol. 34, no. 2, pp. 172–175, 2012.
- [10] X. Yifeng, L. Lijun, H. Qingsong, and F. Tiewei, "Research on TF-IDF weight improvement algorithm in intelligent guidance system," *Computer Engineering and Applications*, vol. 53, no. 4, pp. 238–243, 2017.
- [11] X. Hai-Ling, W. Xiao, L. Xiao-Dong, and B.-P. Yan, "Comparison study of internet recommendation system," *Journal of Software*, vol. 20, no. 2, pp. 350–362, 2009.
- [12] C.-G. Huang, J. Yin, J. Wang, Y.-B. Liu, and J.-H. Wang, "Uncertain Neighbors' Collaborative filtering recommendation algorithm," *Chinese Journal of Computers*, vol. 33, no. 8, pp. 1369–1377, 2010.
- [13] Z. Liang and Z. Na, "Improved collaborative filtering algorithm," *Computer Systems & Applications*, vol. 25, no. 7, pp. 147–150, 2016.
- [14] J. Mingming, *Incorporate Topic Model into Collaborative Filtering*, Beijing Insititute of Technology, Beijing, 2016.
- [15] Y. Wu, T. Rui, and L. Ling, "News recommendation method by fusion of content-based recommendation and collaborative filtering," *Journal of Computer Applications*, vol. 36, no. 2, pp. 414–418, 2016.
- [16] M. López-Nores, Y. Blanco-Fernández, J. J. Pazos-Arias, and A. Gil-Solla, "Property-based collaborative filtering for health-aware recommender systems," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7451–7457, 2012.
- [17] Y. Surong, F. Xiaoqing, and L. Yixing, "Matrix factorization based social recommender model," *Journal of Tsinghua University(Science and Technology)*, vol. 56, no. 7, pp. 793–800, 2016.
- [18] F. Bing and N. Xiaoting, "Tag-based matrix factorization recommendation algorithm," *Application Research of Computers*, vol. 34, no. 4, pp. 1021–1025, 2017.
- [19] Z. X. Huang, X. D. Lu, H. L. Duan, and C. Zhao, "Collaboration-based medical knowledge recommendation," *Artificial Intelligence in Medicine*, vol. 55, no. 1, pp. 13–24, 2012.
- [20] J. H. Kim, D. S. Lee, and K. Y. Chung, "Item recommendation based on context-aware model for personalized u-healthcare service," *Multimedia Tools and Applications*, vol. 71, no. 2, pp. 855–872, 2014.
- [21] M. Deshpande and G. Karypis, "Item-based top-N recommendation algorithms," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 143–177, 2014.
- [22] Z. U. O. Chun-tong, Y. U. Ben-gong, and J. I. A. N. G. Shu, "Bayesian networks for knowledge discovery in large medical data Set," *Microelectronics & Computer*, vol. 25, no. 6, pp. 113–115, 2008, https://www.wanfangdata.com.cn/details/detail.do?_type=perio&id=wdzxyjsj200806030.
- [23] B. S. Hu, D. Feng, W. C. Cao, F. LQ, and G. JH, "Mobile intelligent disease diagnosis system based on Bayesian analysis," *Journal of Computer Applications*, vol. 28, no. 6, pp. 15–17, 2008.
- [24] X. Shoukun and W. Weiwei, "Balance recommendation algorithm for medical resources based on semantic," *Computer Engineering*, vol. 41, no. 9, pp. 74–79, 2015.

- [25] Z. Yan, L. Shiyao, and Z. Can, "An improved recommendation algorithm for mobile health care system," *Journal of University of Chinese Academy of Sciences*, vol. 34, no. 1, pp. 112–118, 2016.
- [26] M. M. Jiang, D. G. Song, L. J. Liao, and F. Zhu, "A Bayesian recommender model for user rating and review profiling," *Tsinghua Science and Technology*, vol. 20, no. 6, pp. 634–643, 2015.
- [27] H. Long-Sheng and Z. Ri-Quan, "0-1 Distributed Bayesian test and application in the medical examination," *Application of Statistics and Management*, vol. 28, no. 6, pp. 1052–1058, 2009, https://www.wanfangdata.com.cn/details/detail.do?_type=perio&id=sltjygl200906013.
- [28] M. Xiang-Wu, L. Shu-Dong, Z. Yu-Jie, and H. Xun, "Research on social recommender systems," *Journal of Software*, vol. 26, no. 6, pp. 1356–1372, 2015.
- [29] A. Sheng-sheng, Z. Hai-yan, C. Qing-Kui, and C. Jian, "Probabilistic matrix factorization recommendation algorithm based on social tag and social trust," *Journal of Chinese Computer Systems*, vol. 37, no. 5, pp. 921–926, 2016, https://www.wanfangdata.com.cn/details/detail.do?_type=perio&id=xxwxjsjxt201605007.
- [30] W. Yang, Z. Yong, L. Zhendong, and Y. Guanci, "Rating prediction algorithm based on semantic similarity and matrix factorization," *Journal of Computer Applications*, vol. 37, Supplement 1, pp. 287–291, 2017.
- [31] F. Xiaoyu, D. Yongxiang, Z. Pengwei, and Z. Xiao, "Study for the construction method of scientist profile with multi source data fusion," *Library and Information Service*, vol. 62, no. 15, pp. 31–40, 2018.
- [32] X. Qiu, Q. Zhang, and X. Huang, "Fudan NLP:a toolkit for Chinese natural language processing," in *Proceedings of the meeting of the Association for Computational Linguistics: system demonstrations*, pp. 49–54, Sofia: the Association for Computational Linguistics, 2013.
- [33] L. Zhou and D. Zhang, "NLPIR: A theoretical framework for applying natural language processing to information retrieval," *Journal of the American Society for Information Science and Technology*, vol. 54, no. 2, pp. 115–123, 2003.
- [34] L. Ting, C. Wanxiang, and L. Zhenghua, "Language technology platform," *Journal of Chinese Information Processing*, vol. 25, no. 6, pp. 53–62, 2011.
- [35] K. Wang, *User information extraction and analysis big data environment*, Beijing University of Posts and Telecommunications, Beijing, 2018.

Research Article

Whole Volume Apparent Diffusion Coefficient (ADC) Histogram as a Quantitative Imaging Biomarker to Differentiate Breast Lesions: Correlation with the Ki-67 Proliferation Index

Yuan Guo,^{1,2} Qing-cong Kong,³ Li-qi Li,⁴ Wen-jie Tang,² Wan-li Zhang,² Guan-yuan Ning,² Jun Xue,⁵ Qian-wei Zhou,⁶ Ying-ying Liang ², Mei Wu ² and Xin-qing Jiang ^{1,2}

¹The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, 510630, China

²Department of Radiology, Guangzhou First People's Hospital, South China University of Technology, Guangzhou 510180, China

³Department of Radiology, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou 510630, China

⁴Department of Radiology, Guangzhou Red Cross Hospital, Medical College, Jinan University, Guangzhou, Guangdong Province, 510220, China

⁵University of Hospital, Shanxi University of Finance and Economics, Taiyuan 030600, China

⁶College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Correspondence should be addressed to Ying-ying Liang; 1368970168@qq.com, Mei Wu; may9@sina.com, and Xin-qing Jiang; eyjiangxq@scut.edu.cn

Received 8 May 2021; Accepted 9 June 2021; Published 24 June 2021

Academic Editor: Lei Zhang

Copyright © 2021 Yuan Guo 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.

Objectives. To evaluate the value of the whole volume apparent diffusion coefficient (ADC) histogram in distinguishing between benign and malignant breast lesions and differentiating different molecular subtypes of breast cancers and to assess the correlation between ADC histogram parameters and Ki-67 expression in breast cancers. **Methods.** The institutional review board approved this retrospective study. Between September 2016 and February 2019, 189 patients with 84 benign lesions and 105 breast cancers underwent magnetic resonance imaging (MRI). Volumetric ADC histograms were created by placing regions of interest (ROIs) on the whole lesion. The relationships between the ADC parameters and Ki-67 were analysed using Spearman's correlation analysis. **Results.** Of the 189 breast lesions included, there were significant differences in patient age ($P < 0.001$) and lesion size ($P = 0.006$) between the benign and malignant lesions. The results also demonstrated significant differences in all ADC histogram parameters between benign and malignant lesions (all $P < 0.001$). The median and mean ADC histogram parameters performed better than the other ADC histogram parameters (AUCs were 0.943 and 0.930, respectively). The receiver operating characteristic (ROC) analysis revealed that the 10th percentile ADC value and entropy could determine the human epidermal growth factor receptor 2 (HER-2) status (both $P = 0.001$) and estrogen receptor (ER)/progesterone receptor (PR) status ($P = 0.020$ and $P = 0.041$, respectively). Among all breast cancer lesions, 35 tumours in the low-proliferation group (Ki-67 < 14%) and 70 tumours in the high-proliferation group (Ki-67 ≥ 14) were analysed with ROC curves and correlation analyses. The ROC analysis revealed that entropy and skewness could determine the Ki-67 status ($P = 0.007$ and $P < 0.001$, respectively), and there were weak correlations between ADC entropy ($r = 0.383$) and skewness ($r = 0.209$) and the Ki-67 index. **Conclusion.** The volumetric ADC histogram could serve as an imaging marker to determine breast lesion characteristics and may be a supplemental method in predicting tumour proliferation in breast cancer.

1. Introduction

Breast cancer is a heterogeneous disease, and the different subtypes can be defined by the immunohistochemical

(IHC) approach based on estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER-2) and Ki-67 expression levels [1–3]. The accurate preoperative diagnosis of breast lesions and further

classification of breast cancer are very important for the selection of an appropriate treatment strategy and prognostic evaluation [4].

The value of diffusion-weighted imaging (DWI) in breast cancer detection and differentiation has already been investigated in a number of previous reports [4–6]. However, the procedure for ADC measurements in breast lesions has not been standardized, and size and positioning of the region of interest (ROI) affect both the ADC levels and reproducibility of the measurements [7]. The ADC measurements in the previous study were mostly based on traditional 2D regions of interest (ROIs) manually drawn from a single representative slice of the breast lesion, which might limit these ADC measurements in their ability to reflect whole tumour characteristics [4, 8–11]. Assessments with whole volume histogram analyses of the ADC might provide more reliable results to reflect the biological characteristics of the heterogeneous breast lesions [3, 8–12].

To the best of our knowledge, the Ki-67 index is considered to represent tumour proliferation status, and a high Ki-67 is associated with an adverse clinical outcome [13]. Ki-67 is helpful for identifying women with early and advanced stages of the disease [14–17], and the change in Ki-67 levels through neoadjuvant therapy has been used as a marker of treatment response recently [13]. Therefore, it makes sense to find a noninvasive imaging biomarker to predict the Ki-67 index.

For all imaging biomarkers, DWI maps and ADC values correlated with tumour cell density, and a low ADC value indicated high cell density or less extracellular space in the histologic analysis. Therefore, the possibility of applying ADC values to predict the Ki-67 index as a prognostic factor has received close attention. In addition, the whole volume ADC histogram could supply more information and predict Ki-67 more accurately than a single ADC value. Some studies analysed the associations between the ADC value and the expression of Ki-67 in breast cancer [7, 18, 19] with different ROI placements; however, data about the relationships between the Ki-67 index and ADC value were inconsistent. Mori et al. reported that there was a moderately significant correlation between the whole tumour ADC histogram (ADC-mean) and Ki-67 [13]. However, Surov et al. found only a weak negative correlation between these two parameters [18]. Some studies found that there were no statistically significant correlations between the ADC value and Ki-67 [7, 19]. Overall, the possibility of using ADC as an imaging marker for proliferation activity in breast cancer is uncertain in clinical practice.

The purpose of the present study was to certify the value of whole volume ADC histograms in differentiating between benign and malignant breast lesions and molecular subtypes of breast cancer and to test the correlation between the ADC histogram parameters and expression of Ki-67 in breast cancer.

2. Methods and Materials

2.1. Patients. The retrospective, single-centre study was approved by our institutional review board. Between September

2016 and February 2019, 259 patients with suspicious findings on mammography or ultrasound underwent breast MRI at our institution. A total of 189 patients who fulfilled the following inclusion and exclusion criteria were retrospectively evaluated. The inclusion criteria were as follows: (1) patients with pathologically diagnosed breast lesions after surgery or biopsy; (2) all patients who underwent standard breast magnetic resonance imaging, including axial T1-weighted images, fat-suppressed T2-weighted images, axial fat-saturated T1-weighted images pre- and postenhancement, and DWI sequences; and (3) all patients who had complete relevant clinical data; if the patients had breast cancer, immunohistochemistry data and Ki-67 data were needed. The exclusion criteria were as follows: (1) breast-related clinical treatment before MRI and (2) poor image quality due to patient motion, eddy current-induced distortions, or inadequate fat suppression. The patient selection process is demonstrated in Figure 1.

2.2. MR Examination Protocol. A total of 189 patients underwent breast MR imaging examinations using a 1.5 T system (uMR 560 1.5 T scanner (United Imaging, Shanghai, China)) with the use of a dedicated four-channel SENSE breast coil. The patients were placed in the prone position with the breasts immobilized. The MRI acquisition protocols were standardized as follows. First, transverse T1-weighted and fat suppressed T2-weighted images were obtained. Second, transverse DWI was performed using a single-shot spin-echo echo-planar imaging sequence with the following parameters: repetition time/echo time (TR/TE), 3800/78 msec; field of view, $350 \times 200 \text{ mm}^2$; matrix, 156×156 ; slice thickness, 4 mm; 27 slices with 0.8 mm gap; voxel size, $2.0 \times 2.0 \times 4.0 \text{ mm}^3$; b value, 50 and 800 sec/mm^2 ; number of averages, 1; and acquisition time, 103 seconds. Third, the gadolinium-based agent Gd-DTPA (gadopentetate dimeglumine, Magnevist; Bayer Healthcare, Berlin, Germany) was intravenously injected at a dose of 0.2 ml/kg of body weight at a rate of 1.5 ml/s, followed by a 20 ml saline flush performed with a high-pressure injector. Axial 3D fat-saturated T1WI were obtained immediately before contrast administration and at six consecutive time points following the administration of the Gd-DTPA contrast agent, with the following parameters: TR/TE, 5.1/2.1 msec; flip angle, 10; field of view, $320 \times 320 \text{ mm}^2$; matrix, 400×70 ; and slice thickness, 2.4 mm. ADC maps were generated with a mono-exponential fit for the diffusion data with b values of 50 and 800 sec/mm^2 using the following formula: $\text{ADC} = [\ln S_0 - \ln S(b)]/b$ (where S_0 and $S(b)$ represent the DWI signal intensity at $b = 50$ and 800 sec/mm^2 , respectively [20, 21]). EPI (fat-suppressed single-shot spin-echo echo-planar imaging) was used for fat suppression.

2.3. Imaging Analysis. All DWI scans were retrospectively reviewed by radiologist G.Y. (with 12 years of experience in breast MRI); if the result is questionable or uncertain, the case was discussed with a second senior radiologist to determine by consensus. The radiologist was blinded to the histopathological results. Axial T2-weighted MRI images, dynamic contrast-enhanced images, DWI scans,

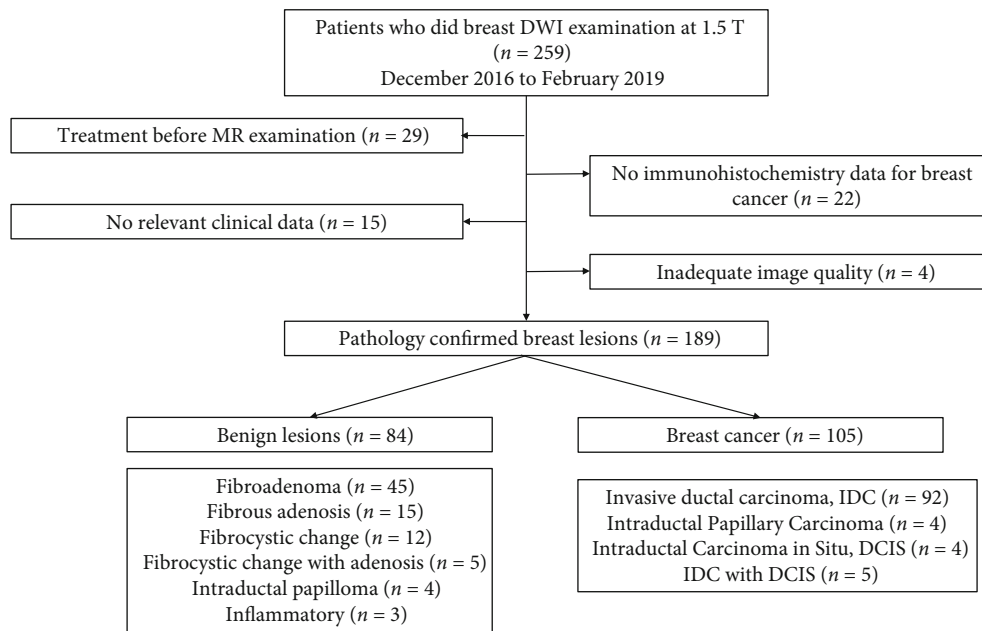


FIGURE 1: Flowchart of the patient selection process used in this study.

TABLE 1: Summary of clinical and pathological features of study subjects and tumour characteristics.

	Benign (n = 84)	Malignant (n = 105)	Statistical value	P value
Age (years)	46.97 ± 12.83	55.37 ± 10.61	4.927	<0.001
Position			-0.032	0.974
Left	43 (51.2%)	54 (51.4%)		
Right	41 (48.8%)	51 (48.6%)		
Menstrual status			-0.676	0.499
Premenopausal	28 (33.3%)	40 (38.1%)		
Postmenopausal	56 (66.7%)	65 (61.9%)		
Lesion size			-2.732	0.006
≤20	64 (76.2%)	40 (38.1%)		
>20	20 (23.8%)	65 (61.9%)		
Lesion type			-1.371	0.170
Mass	82 (97.6%)	98 (93.3%)		
Nonmass	2 (2.4%)	7 (6.7%)		

and ADC maps were transmitted from the workstation to a personal computer for the histogram analysis. The reference for tumour detection was the dynamic contrast-enhanced images and axial T2-weighted images; the largest lesion was chosen for analysis in cases of multifocal or multicentric cancer.

Whole volume ROI placement approaches were applied by each observer (ROI-w): multiple large 2D ROIs were manually drawn on each slice containing the whole lesion of interest and were then combined to create a 3D ROI using the ITK-SNAP tool (<http://www.itksnap.org/pmwiki/pmwiki.php>). ITK-SNAP is a software application used to segment structures in 3D medical images, and it is free, open-source, and multiplatform. The ROI-w, including any cystic or necrotic portions and haemorrhagic components, was evaluated to assess the heterogeneity of the tumour. The analysis was performed with python

software. The ROI containing the whole tumour generated an entire tumour volume reconstruction and displayed the calculated results in the form of a histogram with the Matplotlib package in python. Various ADC histogram parameters were calculated: 10th percentile, mean, 50th percentile (median), 90th percentile, skewness (a measure of asymmetry of the histogram about its mean), kurtosis (a measure of the peakedness of the histogram), and entropy (measure of the variation in the histogram distribution). We followed the methods of Tang et al. [12].

2.4. Histopathological Analysis. All patients underwent mastectomy and lumpectomy, and histopathologic evaluations were performed on the resected specimens. All immunohistochemical materials were reassessed in the breast cancer cases, and the findings were confirmed by a dedicated breast pathologist (W.W., with 13 years of experience).

TABLE 2: Comparison of different parameters of whole volumetric ADC histogram ROC curves in differentiation of benign and malignant breast lesions.

Parameter	Benign	Malignant	AUC	95% CI	Sensitivity	Specificity	Cut-off	<i>P</i> value
10%	1.259	0.807	0.922	0.874~0.956	91.4	88.1	1.022	<0.001
50%	1.591	1.077	0.943	0.900~0.972	86.7	90.5	1.288	<0.001
90%	1.904	1.457	0.843	0.783~0.891	68.6	84.5	1.615	<0.001
Mean	1.585	1.106	0.930	0.884~0.962	87.6	85.7	1.331	<0.001
Skewness	-0.165	0.601	0.808	0.744~0.861	78.1	71.4	0.160	<0.001
Kurtosis	0.785	1.659	0.705	0.635~0.769	77.1	65.5	0.540	<0.001
Entropy	5.691	6.724	0.768	0.701~0.826	80.0	71.4	5.740	<0.001

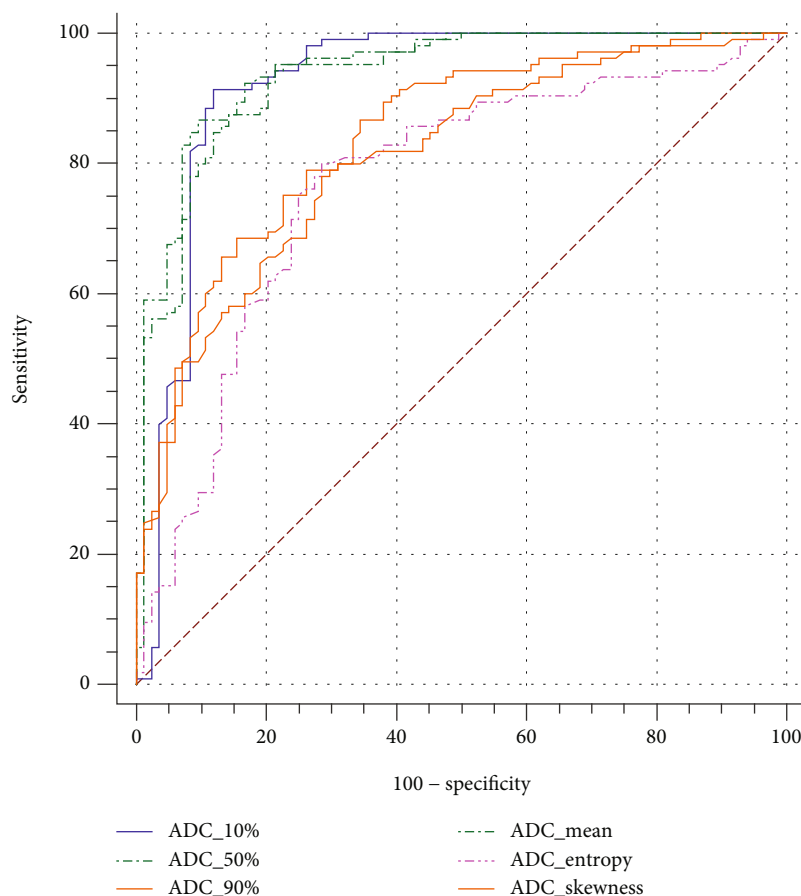


FIGURE 2: Receiver operating characteristic curve (ROC) for differentiation of benign versus malignant lesions using whole volumetric ADC histogram.

The evaluated pathological data included ER, PR, and HER-2 expression and the Ki-67 index. All cases were divided into luminal (luminal A and luminal B) and nonluminal subtypes (HER-2 overexpressed and triple-negative breast cancer).

2.5. Statistical Analysis. Statistical analysis was performed using SPSS 21.0 (IBM Corp., Armonk, NY, USA), MedCalc 8 (MedCalc Software, Ostend, Belgium). Levene's test was used to determine whether the continuous variables of the histogram parameters were normally distributed. Continu-

ous variables were compared with Student's *t*-test or Mann-Whitney *U* test if the variables were not normally distributed. Categorical variables were compared using Pearson's chi-squared test or Fisher's exact test. ROC analysis was performed to compare the diagnostic performance of each parameter in distinguishing between benign and malignant breast lesions and different subtypes of breast cancer. Corresponding areas under the ROC curve (AUCs), and the 95% confidence intervals (95% CIs), cut-off value, sensitivity, and specificity are listed. A *P* value ≤ 0.05 was considered statistically significant.

TABLE 3: Comparison of different parameters of whole volumetric ADC histogram ROC curves in differentiation of subtypes of breast cancer.

(a)

Parameter	Her-2(-) (n = 75)	Her-2(+) (n = 30)	Her-2 status				Cut-off	P value
			AUC	95% CI	Sensitivity (%)	Specificity (%)		
ADC-10%	0.779	0.872	0.679	0.581~0.767	80.00	49.33	0.771	0.001
ADC-50%	1.093	1.142	0.585	0.485~0.681	46.67	73.33	1.153	0.165
ADC-90%	1.532	1.571	0.590	0.490~0.685	53.33	70.67	1.587	0.135
Skewness	0.554	0.584	0.505	0.406~0.604	16.67	93.33	1.310	0.934
Kurtosis	1.347	1.791	0.570	0.470~0.667	40.00	78.67	2.170	0.277
Entropy	6.672	7.516	0.680	0.582~0.768	90.00	42.67	6.140	0.001

(b)

Parameter	ER/PR status		AUC	95% CI	Sensitivity (%)	Specificity (%)	Cut-off	P value
	Luminal (n = 79)	Nonluminal (n = 26)						
ADC-10%	0.785	0.869	0.642	0.543~0.734	100	24.05	0.665	0.020
ADC-50%	1.090	1.159	0.595	0.495~0.689	42.31	78.48	1.202	0.163
ADC-90%	1.521	1.610	0.574	0.530~0.722	42.30	70.89	1.555	0.258
Skewness	0.539	0.636	0.541	0.441~0.639	19.23	92.41	1.260	0.538
Kurtosis	1.412	1.663	0.532	0.432~0.630	26.92	87.34	3.130	0.643
Entropy	6.738	7.445	0.626	0.526~0.718	92.31	39.24	6.090	0.041

3. Results

3.1. Patient Characteristics. Of the 189 breast lesions included, 84 (44.4%) were diagnosed as benign and 105 (55.6%) were malignant. The benign and malignant lesion characteristics are shown in Table 1. There was a significant difference in patient age ($P < 0.001$) and lesion size ($P = 0.006$), but there were no significant differences in lesion position ($P = 0.974$), lesion type ($P = 0.170$), and menopausal status ($P = 0.499$) between the benign and malignant breast lesions.

3.2. Performance Efficiency of the Whole Volume ADC Histogram in Differentiating between Benign and Malignant Breast Lesions. The results demonstrated significant differences in all ADC histogram parameters (including mean, 10th percentile, 50th percentile, 90th percentile, skewness, kurtosis, and entropy) between the benign and malignant lesions (all $P < 0.001$, Table 2). The median and mean ADC histogram parameters performed better than the other ADC parameters (AUC were 0.943 and 0.930, respectively), as shown in Figure 2.

3.3. Performance Efficiency of the Whole Volume ADC Histogram in Differentiating between Different Molecular Subtypes of Breast Cancer. The receiver operating characteristic analysis revealed that the 10th percentile whole volume ADC volume and entropy could determine the Her-2 status ($P = 0.001$ and $P = 0.001$, respectively) and ER/PR status ($P = 0.020$ and $P = 0.041$, respectively) (Table 3).

3.4. Correlation between the ADC Histogram Parameters and Ki-67 Index. For the 105 breast cancer lesions, pathologic evaluation of the Ki-67 ranged from 1 to 86 (median, 45); 35 lesions had a Ki-67 of less than 14 and were categorized as the low-proliferation group (Figure 3), and 70 had a Ki-67 of 14 or greater and were categorized as the high-proliferation group (Figure 4). Receiver operating characteristic analysis revealed that the whole volume ADC entropy and skewness could reflect the Ki-67 status ($P = 0.007$ and $P < 0.001$, respectively) (Table 4).

Spearman's rank correlation maps showed weak correlations between ADC entropy ($r = 0.383$) and skewness ($r = 0.209$) and Ki-67 index (Figure 5), and the 10th, 50th, and 90th percentages had no correlation with Ki-67.

4. Discussion

We examined whether the ADC histogram analysis of the whole lesion was reliable and helpful in determining the breast lesion characteristics and whether the ADC histogram parameters were correlated with the Ki-67 index in breast cancer. In this work, the whole volume ADC histogram was used for three purposes: (1) to discriminate between benign and malignant lesions, (2) to assess the molecular subtypes of cancers, and (3) to correlate the ADC parameters with the Ki-67 expression in breast cancer.

The results indicated that the whole lesion ADC histogram exhibited a higher diagnostic performance in distinguishing between benign and malignant breast lesions than between different subtypes of breast cancer, and the ADC histogram showed a relatively higher diagnostic accuracy,

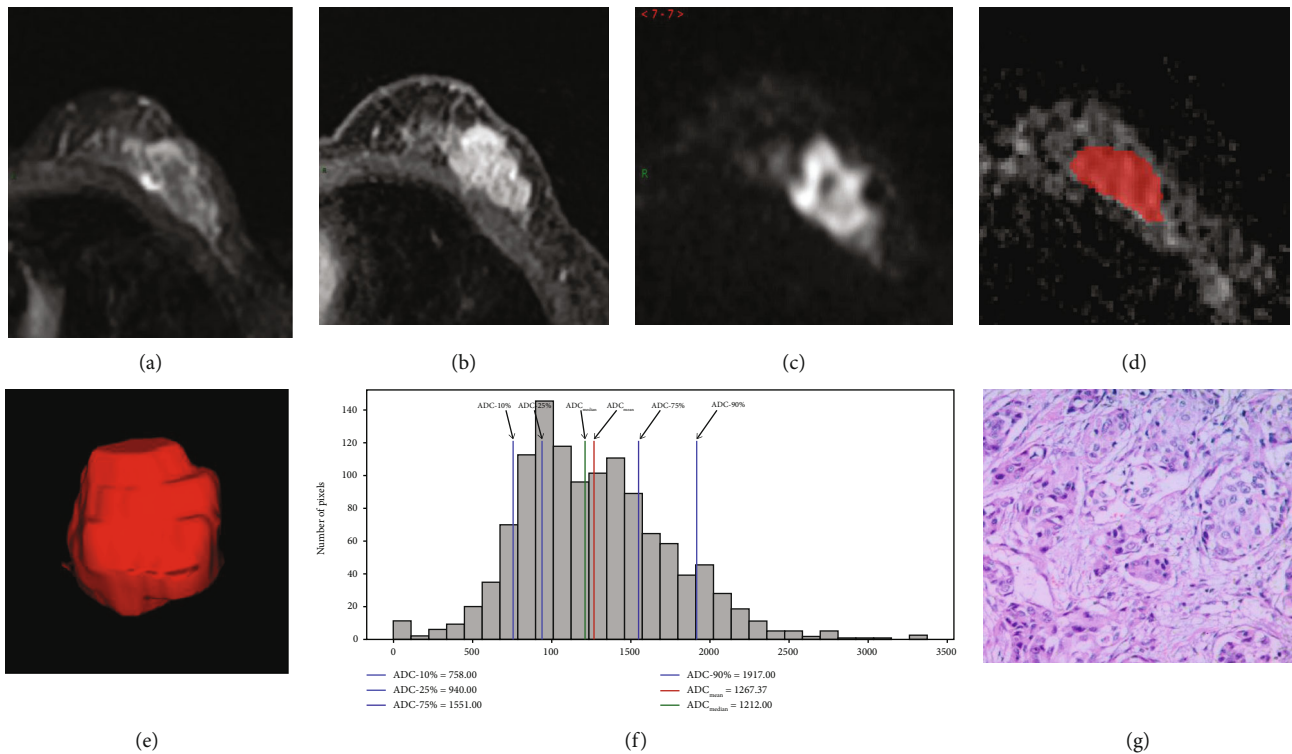


FIGURE 3: A 64-year-old woman with left breast invasive carcinoma (triple negative breast cancer). T2-weighted imaging (a) shows an irregular left breast mass, 23 mm \times 12 mm, with heterogeneous signal. Fat-suppressed contrast-enhanced T1-weighted imaging (b) shows a significant enhancement mass. Diffusion weighted imaging (c) shows high signal mass. (d, e) show the measurement process of whole lesion region of interest (ROI) measurement: manually drawn large 2D-ROIs on each slice (d), then combined multiple 2D-ROI slices to create a 3D-ROI (e). (f) is the whole lesion ADC histogram: $ADC_{mean} = 1.267$; $ADC_{10\%} = 0.758$; $ADC_{50\%} = 1.212$; $ADC_{90\%} = 1.917$ (unit: $10^{-3} \text{mm}^2/\text{s}$); skewness: 1.1; kurtosis: 3.59; entropy: 6.02. (g) HE staining shows left breast invasive carcinoma (HE staining: $\times 100$).

such as the 10th percentile ADC value and ADC entropy. In addition, the results showed weak correlations among ADC entropy, skewness, and Ki-67. These results suggest the potential clinical advantage of ADC histograms as imaging markers in diagnosing breast lesions and that ADC histograms may be a supplement for predicting tumour proliferation in breast cancer.

Initially, there were statistically significant differences in the patient age and lesion size between benign and malignant lesions because in our study, most patients were young and had benign lesions, and the masses were relatively small (≤ 20 mm), which is associated with the sample capacity. Moreover, the value of ADC was emphasized, especially in the histogram-based assessment, which has been used to improve the performance of ADC values in a quantitative manner. Our study showed that the mean, 10th percentile, 50th percentile, and 90th percentile ADC values and skewness, kurtosis, and entropy derived from the whole lesion ADC histogram were able to differentiate between benign and malignant lesions with statistical significance. In malignant lesions, the mean, mode, and percentile ADC values tended to be lower, while the skewness, kurtosis, and entropy values were higher than in benign lesions [10]. The current results were consistent with several previous studies, in which the usefulness of ADC values for providing a differential diagnosis between benign and malignant lesions has been reported, either with 1.5 T or 3.0 T MRI [2]. Therefore,

the first aim of our study was to distinguish between benign and malignant breast lesions; then, we paid more attention to differentiating between the molecular subtypes of breast cancer.

In our study, among the whole lesion ADC histograms, the parameters with best discriminative power to differentiate between different molecular subtypes of breast cancer based on ER/PR and HER-2 status were the 10th percentile ADC value and entropy. It is well known that ADC is inversely correlated with tissue cellularity. We assume that a low percentile ADC value based on a whole lesion histogram analysis may accurately define invasive and high cellularity [8]. The region showing the 10th percentile ADC value may reflect the area with the highest cellularity within the tumour, which is highly representative of tumour grade and aggressiveness.

Entropy is a texture-based statistical measure of the variation in the histogram distribution of a given metric and represents the predictability of the intensity of the metric within the tissue. Malignant pathologies tend to affect a tissue heterogeneously and are expected to result in less predictable intensity characteristics within the tissue and thus higher entropy than benign pathologies [9]. In our study, the entropy of Her-2 overexpression and nonluminal breast cancer were higher than those of Her-2-negative and luminal breast cancer, which means that the former had more heterogeneous features.

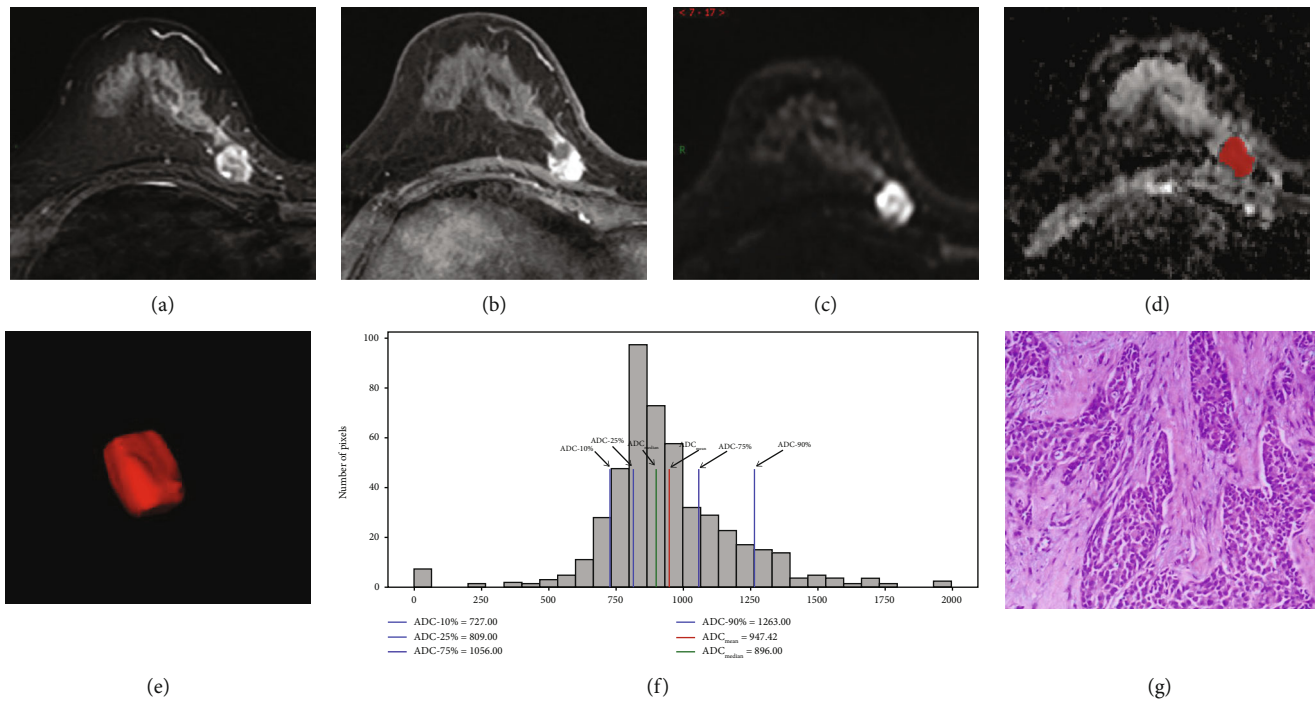


FIGURE 4: A 62-year-old woman with left breast invasive carcinoma (luminal A). T2-weighted imaging (a) shows a round right breast mass, 18 mm × 12 mm, with heterogeneous signals. Fat-suppressed contrast-enhanced T1-weighted imaging (b) shows a significant heterogeneous enhancement mass. Diffusion weighted imaging (c) shows high signal mass. (d, e) show the measurement process of whole lesion region of interest (ROI) measurement: manually drawn large 2D-ROIs on each slice (d), then combined multiple 2D-ROI slices to create a 3D-ROI (e). (f) is the whole tumour ADC histogram: ADC_{mean} : 0.947; ADC-10%: 0.727; ADC-50%: 0.896; ADC-90%: 1.263 (unit: $10^{-3} \text{mm}^2/\text{s}$); skewness: 1.28; kurtosis: 2.61; entropy: 5.89. (g) HE staining shows left breast invasive carcinoma (HE staining: $\times 100$).

TABLE 4: Comparison of different parameters of whole volumetric ADC histogram ROC curves in differentiation of low or high Ki-67 of breast cancer.

Parameter	<14% (n = 35)	$\geq 14\%$ (n = 70)	AUC	Ki-67 status		Sensitivity (%)	Specificity (%)	Cut-off	P value
				95% CI					
ADC-10%	0.809	0.803	0.509	0.409~0.608		88.60	28.60	0.643	0.889
ADC-50%	1.114	1.103	0.532	0.433~0.630		44.29	74.29	1.029	0.591
ADC-90%	1.496	1.471	0.599	0.499~0.693		71.43	45.71	1.422	0.089
Skewness	0.397	0.646	0.647	0.547~0.737		70.00	60.00	0.390	0.007
Kurtosis	1.183	1.620	0.573	0.473~0.669		40.00	80.00	1.730	0.211
Entropy	6.103	7.319	0.746	0.651~0.826		75.71	65.71	6.290	<0.001

Ki-67 has been used as a prognostic biomarker for cell proliferation in breast cancer. However, due to the invasive nature of the examination, it could be meaningful in clinical practice to predict the expression of Ki-67 with some noninvasive imaging parameters. The Ki-67 level is evaluated immunohistochemically in the most proliferative area of the tumour and is expected to correlate best with the minimum ADC value and thus is associated with smaller ADC values [13]. In our study, however, none of the 10th to 90th percentile ADC values showed correlations with the Ki-67 index; only skewness and entropy showed a weak correlation with the Ki-67 index. We know that there were different results regarding the correlation between the ADC value and the Ki-67 index. These differences can be explained by several reasons: (1) Mori et al. showed that the 25th, 50th, and 75th percentile and mean values showed similar negative

correlations with Ki-67 in invasive breast cancer. We presume that the cause might be the differences in patient cohorts. Only luminal breast cancer was studied by Mori et al., which, compared to nonluminal breast cancer, has less heterogeneous histologic components, with little or no necrotic or degenerative components [13]. In our study, 79 luminal and 26 nonluminal breast cancers were studied, and the heterogeneity of the latter was greater than that of the former. (2) Surov et al. suggested that the ADC value could not be used as a surrogate marker for proliferation activity in breast cancer [18]. However, the threshold of the Ki-67 value was 25% to discriminate between tumours with low Ki-67 expression (<25%) and those with high Ki-67 expression ($\geq 25\%$). In addition, the traditional manual ROI measurement was assessed. However, this measurement could not reflect the heterogeneity of the whole tumour. In

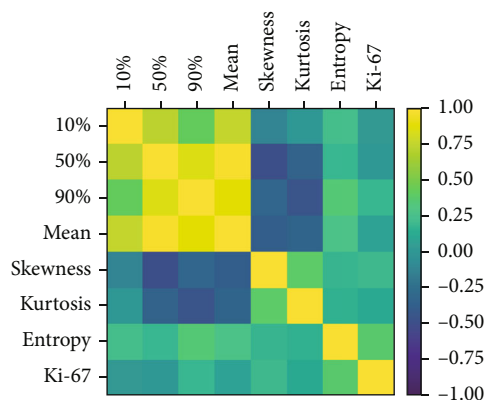


FIGURE 5: Spearman's rank correlation maps between the parameters of ADC histogram and the Ki-67. The correlation coefficient is indicated with color maps ranging from purple to yellow; purple and yellow indicate the negative and positive strongest correlations, respectively. It shows a weak correlation between ADC entropy and Ki-67.

our study, the threshold of the Ki-67 value was 14% for discriminating between tumours, which is the most common threshold. (3) Arponen et al. showed that there was no association between the whole lesion ADC values and the Ki-67 proliferation index [7], and this result was similar to ours. However, only the mean ADC value was calculated in that study, and there were no other parameters. In our study, the 10th to 90th percentile ADC values showed no correlations with the Ki-67 index, but skewness and entropy showed weak correlations with the Ki-67 index. This finding may be explained by the heterogeneous nature of breast cancer, which is correlated with entropy.

We acknowledge several limitations. First, this was a retrospective study. Further prospective and multicentre studies are required to validate our results. Second, our study includes a relatively small number of benign breast lesions and different molecular subtypes of breast cancer. Therefore, it is necessary to expand the database in the future, especially to complement more data on different molecular subtypes of breast cancer, to verify our results. Third, the procedure for ADC measurements in breast lesions should be standardized.

In conclusion, volumetric ADC histograms exhibited a higher diagnostic performance in distinguishing between benign and malignant breast lesions than between different subtypes of breast cancer. The ADC histogram showed a relatively higher diagnostic accuracy than the 10th percentile ADC value and ADC entropy, whereas ADC histogram entropy was weakly correlated with Ki-67. All these results suggest the potential clinical advantage of applying the ADC histogram as an imaging marker in the diagnosis of breast lesions and that the ADC histogram may be a supplemental tool in predicting tumour proliferation in breast cancer.

Data Availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable

request. Anyone who is interested in the information should contact eyguoyuan@scut.edu.cn.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yuan Guo, Qing-cong Kong, and Li-qi Li contributed equally to this work.

Acknowledgments

Grant support was provided by the National Natural Science Foundation of China (No. 81901711); the Natural Science Foundation of Guangdong Province of China (2021A1515011350); and the Science and Technology Project of Guangzhou City, China (201904010422, 202002030268, 202102010025).

References

- [1] M. Fan, T. He, P. Zhang, J. Zhang, and L. Li, "Heterogeneity of diffusion-weighted imaging in tumours and the surrounding stroma for prediction of Ki-67 proliferation status in breast cancer," *Scientific Reports*, vol. 7, no. 1, article 2875, 2017.
- [2] S. Suo, K. Zhang, M. Cao et al., "Characterization of breast masses as benign or malignant at 3.0T MRI with whole-lesion histogram analysis of the apparent diffusion coefficient," *Journal of Magnetic Resonance Imaging*, vol. 43, no. 4, pp. 894–902, 2016.
- [3] T. Xie, Q. Zhao, C. Fu et al., "Differentiation of triple-negative breast cancer from other subtypes through whole-tumor histogram analysis on multiparametric MR imaging," *European radiology*, vol. 29, no. 5, article 5804, pp. 2535–2544, 2019.
- [4] E. J. Kim, S. H. Kim, G. E. Park et al., "Histogram analysis of apparent diffusion coefficient at 3.0T: correlation with prognostic factors and subtypes of invasive ductal carcinoma," *Journal of Magnetic Resonance Imaging*, vol. 42, no. 6, pp. 1666–1678, 2015.
- [5] C. Spick, H. Bickel, K. Pinker et al., "Diffusion-weighted MRI of breast lesions: a prospective clinical investigation of the quantitative imaging biomarker characteristics of reproducibility, repeatability, and diagnostic accuracy," *NMR in Biomedicine*, vol. 29, no. 10, pp. 1445–1453, 2016.
- [6] H. L. Liu, M. Zong, H. Wei et al., "Differentiation between malignant and benign breast masses: combination of semi-quantitative analysis on DCE-MRI and histogram analysis of ADC maps," *Clinical Radiology*, vol. 73, no. 5, pp. 460–466, 2018.
- [7] O. Arponen, M. Sudah, A. Masarwah et al., "Diffusion-weighted imaging in 3.0 Tesla breast MRI: diagnostic performance and tumor characterization using small subregions vs. whole tumor regions of interest," *PLoS One*, vol. 10, no. 10, article e0138702, 2015.
- [8] J. Y. Kim, J. J. Kim, J. W. Lee et al., "Risk stratification of ductal carcinoma in situ using whole-lesion histogram analysis of the apparent diffusion coefficient," *European Radiology*, vol. 29, no. 2, pp. 485–493, 2019.

- [9] H. Bougias, A. Ghiatas, D. Priovolos, K. Veliou, and A. Christou, "Whole-lesion apparent diffusion coefficient (ADC) metrics as a marker of breast tumour characterization-comparison between ADC value and ADC entropy," *The British Journal of Radiology*, vol. 89, no. 1068, 2016.
- [10] H. Bougias, A. Ghiatas, D. Priovolos, K. Veliou, and A. Christou, "Whole-lesion histogram analysis metrics of the apparent diffusion coefficient as a marker of breast lesions characterization at 1.5 T," *Radiography*, vol. 23, no. 2, pp. e41–e46, 2017.
- [11] S. J. Ahn, M. Park, S. Bang et al., "Apparent diffusion coefficient histogram in breast cancer brain metastases may predict their biological subtype and progression," *Scientific Reports*, vol. 8, no. 1, article 28315, p. 9947, 2018.
- [12] W. J. Tang, Z. Jin, Y. L. Zhang et al., "Whole-lesion histogram analysis of the apparent diffusion coefficient as a quantitative imaging biomarker for assessing the level of tumor-infiltrating lymphocytes: value in molecular subtypes of breast cancer," *Frontiers in oncology*, vol. 10, article 611571, 2021.
- [13] N. O. H. Mori, H. Ota, S. Mugikura et al., "Luminal-type breast cancer: correlation of apparent diffusion coefficients with the Ki-67 labeling index," *Radiology*, vol. 274, no. 1, pp. 66–73, 2015.
- [14] P. A. Fasching, P. Gass, L. Häberle et al., "Prognostic effect of Ki-67 in common clinical subgroups of patients with HER2-negative, hormone receptor-positive early breast cancer," *Breast Cancer Research and Treatment*, vol. 175, no. 3, article 5198, pp. 617–625, 2019.
- [15] J. Cuzick, M. Dowsett, S. Pineda et al., "Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer," *Journal of Clinical Oncology*, vol. 29, no. 32, pp. 4273–4278, 2011.
- [16] P. A. Fasching, K. Heusinger, L. Häberle et al., "Ki67, chemotherapy response, and prognosis in breast cancer patients receiving neoadjuvant treatment," *BMC Cancer*, vol. 11, no. 1, p. 486, 2011.
- [17] J. K. Shin and J. Y. Kim, "Dynamic contrast-enhanced and diffusion-weighted MRI of estrogen receptor-positive invasive breast cancers: associations between quantitative MR parameters and Ki-67 proliferation status," *Journal of Magnetic Resonance Imaging*, vol. 45, no. 1, pp. 94–102, 2017.
- [18] A. Surov, P. Clauser, Y. W. Chang et al., "Can diffusion-weighted imaging predict tumor grade and expression of Ki-67 in breast cancer? A multicenter analysis," *Breast Cancer Research*, vol. 20, no. 1, p. 58, 2018.
- [19] Y. Akin, M. U. Ugurlu, H. Kaya, and E. Aribal, "Diagnostic value of diffusion-weighted imaging and apparent diffusion coefficient values in the differentiation of breast lesions, histopathologic subgroups and correlation with prognostic factors using 3.0 Tesla MR," *The journal of breast health*, vol. 12, no. 3, pp. 123–132, 2016.
- [20] L. Zhang, A. A. Mohamed, R. Chai, Y. Guo, B. Zheng, and S. Wu, "Automated deep learning method for whole-breast segmentation in diffusion-weighted breast MRI," *Journal of Magnetic Resonance Imaging*, vol. 51, no. 2, pp. 635–643, 2020.
- [21] Y. N. Jin, Y. Zhang, J. L. Cheng, D. D. Zheng, and Y. Hu, "Monoexponential, biexponential, and stretched-exponential models using diffusion-weighted imaging: a quantitative differentiation of breast lesions at 3.0T," *Journal of Magnetic Resonance Imaging*, vol. 50, no. 5, pp. 1461–1467, 2019.