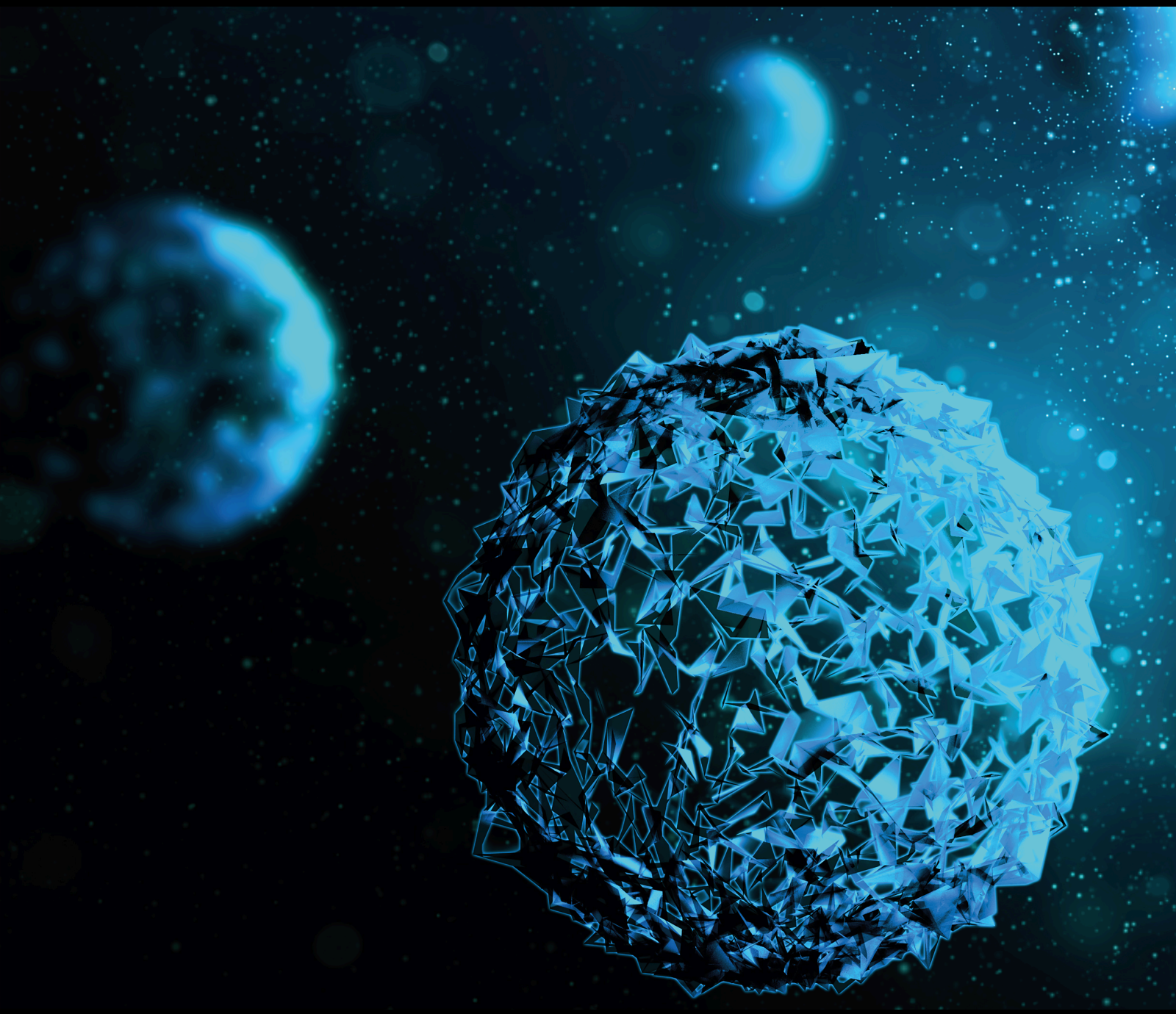
BioMed Research International

Artificial Intelligence for Medical Image Analysis

Lead Guest Editor: Yong Xia
Guest Editors: Changming Sun, Lin Gu, and Huiyu Zhou



Artificial Intelligence for Medical Image Analysis

BioMed Research International

Artificial Intelligence for Medical Image Analysis

Lead Guest Editor: Yong Xia
Guest Editors: Changming Sun, Lin Gu, and Huiyu Zhou

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-Rodriguez, Spain
Momiao Xiong, USA
Hui Zhang, China

Academic Editors

Computational Biology

Contents

Contents

Hindawi
BioMed Research International
Volume 2021, Article ID 6654247, 15 pages
https://doi.org/10.1155/2021/6654247

Research Article

Modeling Respiratory Signals by Deformable Image Registration on 4DCT Lung Images

Pham The Bao,[1] **Hoang Thi Kieu Trang**,[2,3] **Tran Anh Tuan**,[3,4] **Tran Thien Thanh**,[2,3] **and Vo Hong Hai**[2,3]

[1]*Computer Science Department, Information Science Faculty, Sai Gon University, Ho Chi Minh City, Vietnam*
[2]*Department of Nuclear Physics, University of Science, Ho Chi Minh City, Vietnam*
[3]*Vietnam National University, Ho Chi Minh City, Vietnam*
[4]*Department of Computer Science, University of Science, Ho Chi Minh City, Vietnam*

Correspondence should be addressed to Pham The Bao; ptbao@sgu.edu.vn and Hoang Thi Kieu Trang; htktrang@hcmus.edu.vn

Received 30 November 2020; Revised 20 July 2021; Accepted 7 September 2021; Published 30 October 2021

Academic Editor: Fu-Ming Tsai

The lung organ of human anatomy captured by a medical device reveals inhalation and exhalation information for treatment and monitoring. Given a large number of slices covering an area of the lung, we have a set of three-dimensional lung data. And then, by combining additionally with breath-hold measurements, we have a dataset of multigroup CT images (called 4DCT image set) that could show the lung motion and deformation over time. Up to now, it has still been a challenging problem to model a respiratory signal representing patients' breathing motion as well as simulating inhalation and exhalation process from 4DCT lung images because of its complexity. In this paper, we propose a promising hybrid approach incorporating the local binary pattern (LBP) histogram with entropy comparison to register the lung images. The segmentation process of the left and right lung is completely overcome by the minimum variance quantization and within class variance techniques which help the registration stage. The experiments are conducted on the 4DCT deformable image registration (DIR) public database giving us the overall evaluation on each stage: segmentation, registration, and modeling, to validate the effectiveness of the approach.

1. Introduction

Nowadays, diseases of the respiratory system have been increasing because of more and more pollution in many cities. Besides, smoking cigarettes or aging also affects the respiratory tract of people. An approach to model or visualize a respiratory cycling process from 4DCT images for diagnosis is highly encouraged but still has a lot of challenges. Although researchers in this field try to investigate and solve the problem, the results are still rather limited and unsatisfied. To develop a treatment plan by the way of modeling lung movements, registration methods must be taken into account carefully.

There are many conventional approaches in 4DCT lung images, but generally, we can classify them into three types: segmentation, registration, and modeling. In lung segmentation, in 2019, Pang et al. suggested a novel automatic segmentation model using a combination of handcrafted features (gray-level cooccurrence matrix) and deep features (U-Net) [1]. In the paper [2], in 2020, Peng et al. applied two processes to extract coarse lung contours first and then refine the segmentation depending on the basis of the principal curve model. This approach is rather complicated and requires a model initialization for the process. For registration, some researchers use deep learning approaches based on the displacement field to obtain the optimal parameters [3], which must be trained with big data until reaching the optimization. Some other approaches require a landmark tracking process [4], which must be determined by specialists.

In respiratory modeling, a question in regard to performing the registration using only the lung images still needs more researches. There are only a few papers mentioned about lung modeling in computer vision fields such

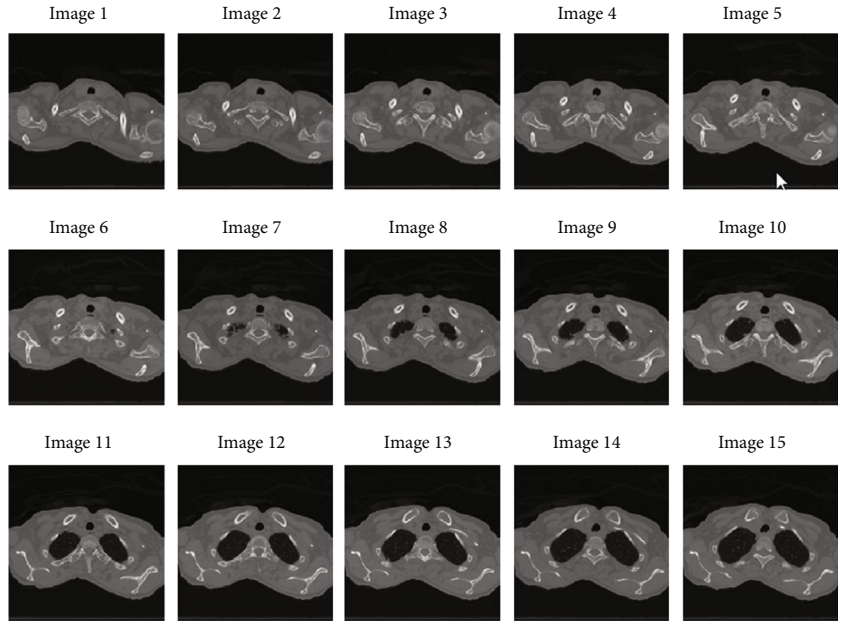


Figure 1: The representation of 4DCT DIR database from slides 1 to 15 in phase T00.

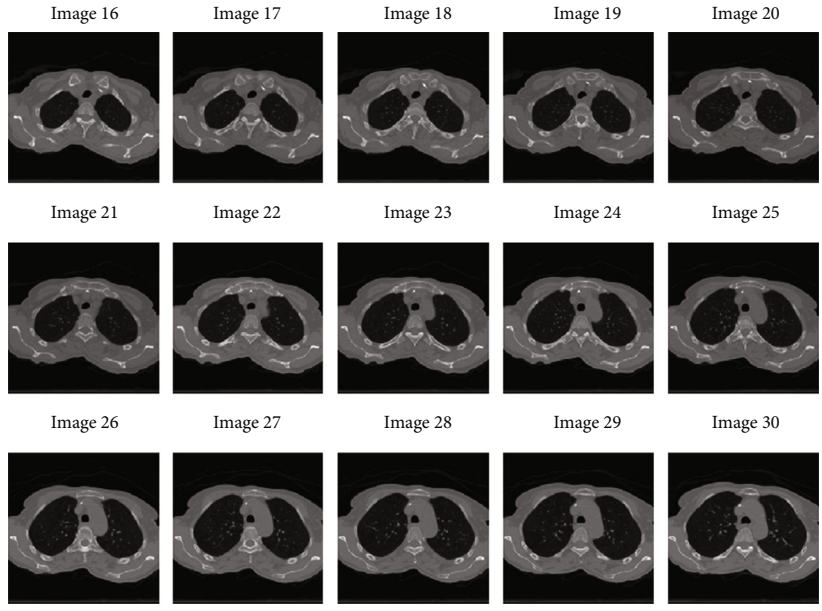


Figure 2: The representation of 4DCT DIR database from slides 16 to 30 in phase T00.

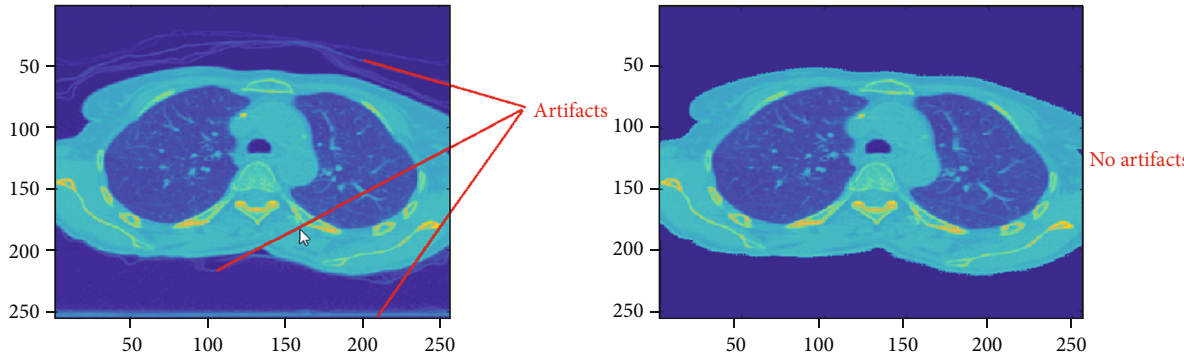


Figure 3: The before and after resultant of lung image in artifact removal.

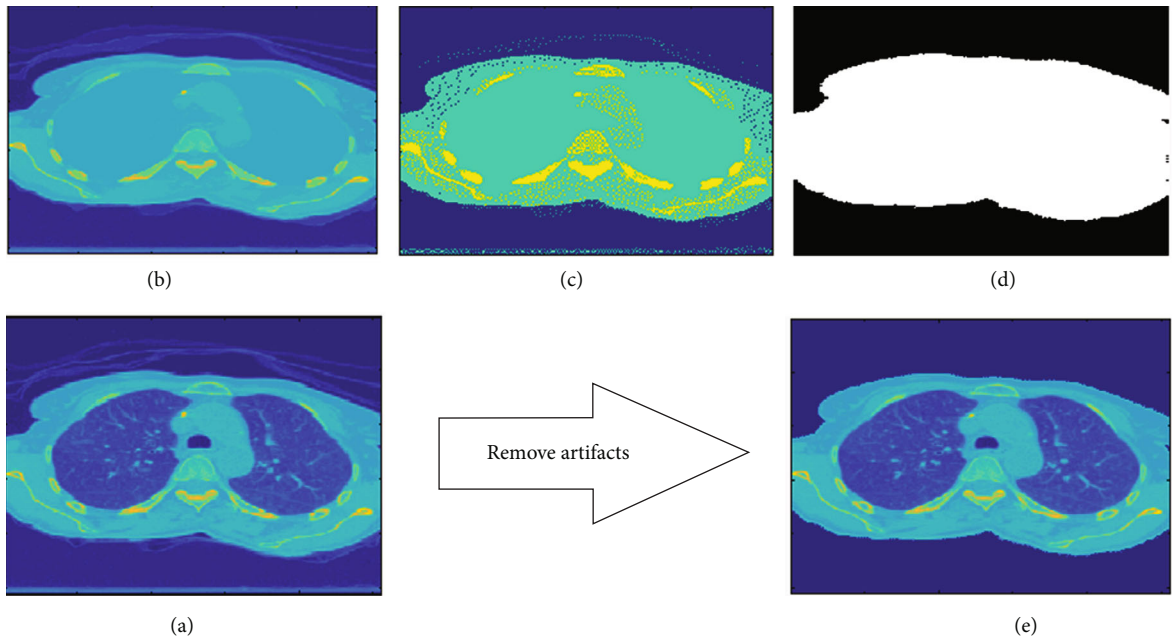


Figure 4: (a) Original image with artifacts. (b) Image after filling holes. (c) Image with three indexes after applying minimum variance quantization. (d) Select the maximum index region, and fill holes. (e) Segmented image with only the index region from the previous step.

1. Input the Matrix SliceIM (One lung slice image in each phase)
2. Find the Dark and Light Area in an image
   LightArea = find(image, 'light');
   DarkArea = find(image, 'dark');
3. Fill the DarkArea within a LightArea
   SliceIM = floodfill(SliceIM, LightArea, DarkArea);
4. Quantitate SliceIM into three indexed images using the Minimum Variance Quantization
   IndexIM = Quantization(SliceIM, 3)
5. Select the index partition in which it is the largest area
   MaxPartitionIM = MaxArea(IndexIM)
6. Fill the holes in MaxPartionIM to get the whole lung partition without artifacts
   OutputIM = FillHoles(MaxPartitionIM)
7. End
8. Result in the Matrix OutputIM (The lung slice image without artifacts, have only lung and body area)

Appendix

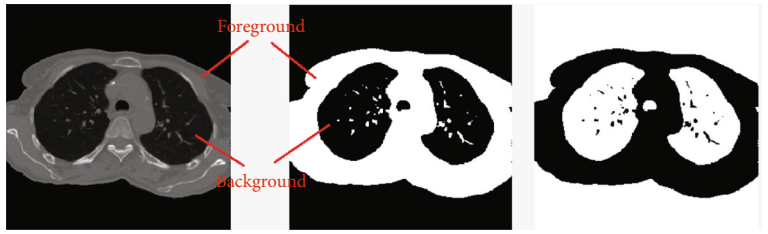
FillHoles method

1. Find the Dark and Light Area in an image
   LightArea = find(image, 'light');
   DarkArea = find(image, 'dark');
2. Fill the DarkArea within a LightArea by floodfill
   image = floodfill(image, LightArea, DarkArea);

Algorithm 1: Remove artifacts.

as the paper of Yang et al. [5] proposed using optical flow to model the motion of the lung. The registration is applied using a multigrid approach and a feature-preserving image downsampling max filter to achieve higher computational speed and registration accuracy. Ehrhardt et al. [6] suggested using the statistical modeling which gives good model result but still depends on landmarks.

In this paper, we suggest an approach using local binary pattern (LBP) and entropy error evaluation (EEE) for registration and modeling 4DCT images into a breathing signal without using any landmark. LBP descriptor is a grayscale and rotation-invariant operator. It is not affected by rotation and variation of the images. The 4DCT lung images have dark and light areas that look like grayscale images. LBP can run faster than other descriptors and extract relevant features for the lung [7]. Then, we make a visualization of the output signal. It will help a doctor easily track or monitor a patient respiratory process for an accurate treatment plan.



(a) (b) (c)

Figure 5: (a) Original image. (b) Foreground and background separation by Otsu threshold. (c) The complement of the (b) result.

(a) (b) (c)

Figure 6: (a) The complemented binary image. (b) The body binary image. (c) The result after multiplication of two binary images (a) and (b).

(a) (b) (c)

Figure 7: (a) Center line based on the center point of the body. (b) The left lung is the largest region on the left. (c) The right lung is the largest region on the right.

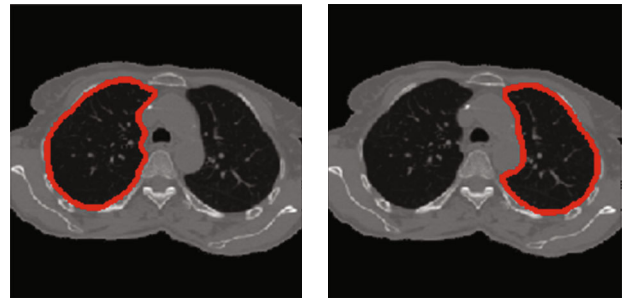
(a) (b) (c)

Figure 8: (a) Original image. (b) The binary image of two lungs. (c) The segmented two lungs.

The segmentation, registration, and modeling stages will be described in detail. Firstly, the 4DCT images include inhale and exhale states. The images for the exhale state are segmented and served as the reference model. Secondly, images that belong to different respiratory phases from a given anatomical position are aligned with each other. The accuracy of lung segmentation is very important for registration and modeling. Minimum variance quantization (MVQ) and within class variance (WCV) methods are applied for segmentation effectively and precisely.

2. 4DCT Data Structure Exploratory

Generally, in a single scan, a 4DCT dataset includes about 700 to 1500 computer tomography (CT) images. Each image has two dimensions corresponding to the width and height



(a) (b)

Figure 9: Left and right lung segmentation and highlight.

1. Input the SliceIM (One lung slice image in a phase without artifacts)
2. Within Class Variance approach to binarize an image
   BinaryIM = WithinClassVariance(SliceIM)
3. Complement of the BinaryIM to change the area of interest into 1 and background into 0 with BinaryIM =1 – BinaryIM;
4. Fill all holes in BinaryIM and keep these large areas (in this case the large area is greater than 30)
   BinaryIM = FillHoles(BinaryIM)
   BinaryIM = LargeArea(BinaryIM, 30)
5. Store the result as a Candidate Lung partition
   CandidateIM = BinaryIM
6. End
7. Result in CandidateIM (The candidate lung partitions)

Appendix

WithinClassVariance method

1. Compute histograms and probabilities of each intensity level
2. Set up initial $\omega_i(0)$ and $\mu_i(0)$
3. Step through all possible thresholds t =1, .. maximum intensity
   a. Update $\omega_i(0)$ and $\mu_i(0)$
   b. Compute $\sigma^2{}_b(t)$
4. Desired threshold corresponds to the maximum $\sigma^2{}_b(t)$

Algorithm 2: Candidate lung partition segmentation.

of the image. The third dimension is the order number of slices, which is scanned at a certain defined interval along the patient's body. The last dimension is phases of scanning time.

Deformable image registration (DIR) is an emerging technology with diagnostic and therapeutic medical applications. DIR algorithms were first developed in computer vision research to estimate motion by warping a source image onto a target, producing an estimated image that visually appeared like the target image.

In this research, the 4DCT dataset was acquired as a part of the standard planning process for the treatment of thoracic malignancies at The University of Texas M. D. Anderson Cancer Center in Houston and offered by DIR-LAB [3]. In 4DCT imaging, thoracic movements are monitored by a Varian Real-time Position Management (RPM) system during the CT scan. The RPM system divides the complete respiratory cycle into ten phases, from 0% (phase T00) to 90% (T90) at 10% intervals, where 0% corresponds to the end inspiration [8]. Then, the reconstructed CT images are sorted into the ten phases based on the temporal correlation between the RPM respiration data and the CT data acquisition time of each image. The dataset has the following structure:

(i) First and second dimensions: 256×256 images

(ii) Third dimension: 92 slices from the top to the bottom of a lung with 2.5 mm slice spacing

(iii) Fourth dimension: phases of time from T00 to T90

Figures 1 and 2 demonstrate a part of the dataset along the third dimension from slice 1 to slice 30 in phase T00.

3. Lung Segmentation and Artifact Removal

The process of segmentation has two steps. The first one is artifact removal, and the second one is lung segmentation.

Step 1 (artifact removal): because the outside area of the lung and body region contains some artifacts that might affect segmentation result, the body and the lung area from the image should be extracted. To enhance the virtualization

1. Input the LungIM (the binary image containing all candidate lung partitions)
2. Find the vertical line in the image
    Height = Height(LungIM);
    CenterX = CenterPoint(LungIM,'X');
    VerticalLine = (CenterX, 1) to (CenterX, Height)
3. Get a List of Candidate Partitions on the left and right side of VerticalLine
    CandidateList = GetListPartition(LungIM);
    LeftCandidateList = CandidateList ? CandidateList on left of VerticalLine : null;
    RightCandidateList = CandidateList ? CandidateList on right of VerticalLine : null;
4. Find the largest Candidates on the left and the right
    LeftLungIM = LeftCandidateList (Index(LeftCandidateList,'largest'))
    RightLungIM = RightCandidateList (Index(RightCandidateList,'largest'))
5. End
6. Result in LeftLungIM (the left lung partition) and RightLungIM (the right lung partition)

Appendix

GetListPartition method

1. Set label to each unconnected partition from 1 … number of partitions
2. Initiate a list
3. Step through all possible idx =1, .. number of partitions
    a. Extract the partition in index = idx
    b. Store it to the list
4. Finish

Index method

1. Sort the list from the largest area to the smallest one
2. Take the first element in the list if the input is 'largest'
3. Take the last element in the list if the input is 'smallest'

Algorithm 3: Left and right lung segmentation.

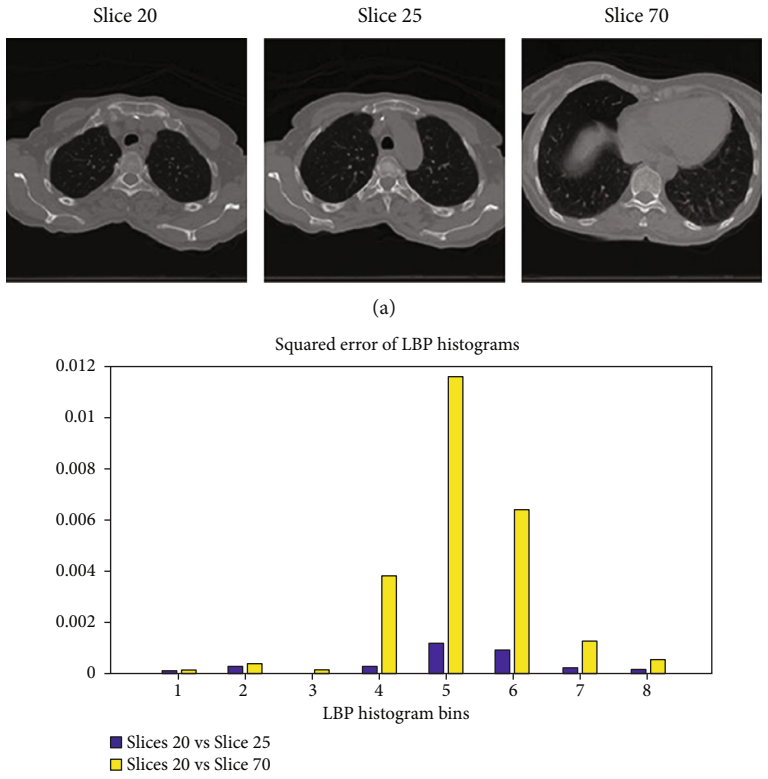


Figure 10: (a) The lung presentation in slices 20th, 25th, and 70th. (b) The squared error of LBP (formula in Appendix in Algorithm 4) between slices. Slices 20th and 70th have lower squared error than slices 20th and 25th in all bins. This means that the slices 20th and 25th are in the same respiratory stage, while the slices 20th and 70th are not in the same stage.

1. Input the source slice: SrcSliceIM (contains only the left and right lungs) and the target slice: TarSliceIM (contains only the left and right lungs)
2. Extract the LBP features of the source and target slices
    SrcLBPFeatures = ExtractLBPFeatures(SrcSliceIM)
    TarLBPFeatures = ExtractLBPFeatures(TarSliceIM)
3. Gauge the similarity between the LBP features by computing the squared error between them
    Similarity = square(TarLBPFeatures – SrcLBPFeatures)
4. Local Binary Pattern Error Rate
    LBPErrorRate = sum(Similarity)
5. End
6. Result in LBPErrorRate

Appendix

ExtractLBPFeatures method

1. The texture T as the joint distribution of the gray levels of P +1 image pixel

$$T = t(g_c, g_0 \cdots g_{p-1}),$$

where g_c is the gray level value of the center pixel, surrounded by P equally spaces pixels of gray levels g_p, located on a circle of radius R.

2. Define the Local Binary Pattern (LBP), a grayscale invariant and rotation invariant operator:

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{i=0}^{P-1} \sigma(g_p - g_c) & \text{if } U(LBP_{P,R}) \leq 2 \\ P+1 & \text{otherwise} \end{cases}$$

Where

$$U(LBP_{P,R}) = |\sigma(g_{p-1} - g_c) - \sigma(g_0 - g_c)| + \sum_{i=1}^{P-1} |\sigma(g_i - g_c) - \sigma(g_{i-1} - g_c)|$$

and $\sigma(.)$is the sign function. The uniformity function $U(LBP_{p,R})$ corresponds to the number of spatial transitions in the neighborhood: the larger it is, the more likely a spatial transition occurs in the local pattern.

Algorithm 4: Local binary pattern error rate.

1. Input the source slice: SrcSliceIM (contains only the left and right lungs) and the target slice: TarSliceIM (contains only the left and right lungs)
2. Compute the entropy of each source and target slices
    SrcEntropyFeatures = ExtractEntropyFeatures(SrcSliceIM)
    TarEntropyFeatures = ExtractEntropyFeatures(TarSliceIM)
3. Entropy Error Rate
    EntropyErrorRate = abs(TarEntropyFeatures - SrcEntropyFeatures)
4. End
5. Result in EntropyErrorRate

Appendix

ExtractEntropyFeatures method

1. Calculated p contains the normalized histogram counts returned from the image
2. Entropy is defined as -sum(p.∗log2(p))

Algorithm 5: Entropy error rate.

of the artifacts, the original 4DCT images are converted from grayscale to color as shown in Figure 3. Then, the minimum variance quantization method [9] is applied to cluster image pixels.

Minimum variance quantization associates pixels into groups based on the variance between their pixel values. For example, a set of blue pixels might be grouped together because they have a small variance from the mean pixel value of the group. In the lung image, the region of interest is the group of pixels at the center of the image, which contains those representing the lung. By focusing on this group, all artifacts outside the body part could be removed. These artifacts come from the lighting of background objects outside the patient's body in a scan.

In general, minimum variance quantization can be replaced by other clustering methods such as K-means, K-nearest-neighbor (KNN), and expectation maximization (EM). By comparing their results, we decide to use the minimum variance quantization for artifact removal. The details of this step are demonstrated in Figure 4 and described in Algorithm 1.

Step 2 (lung segmentation): after removing artifacts, we need to segment two lungs from the image. The segmented result allows a comparison between phases and

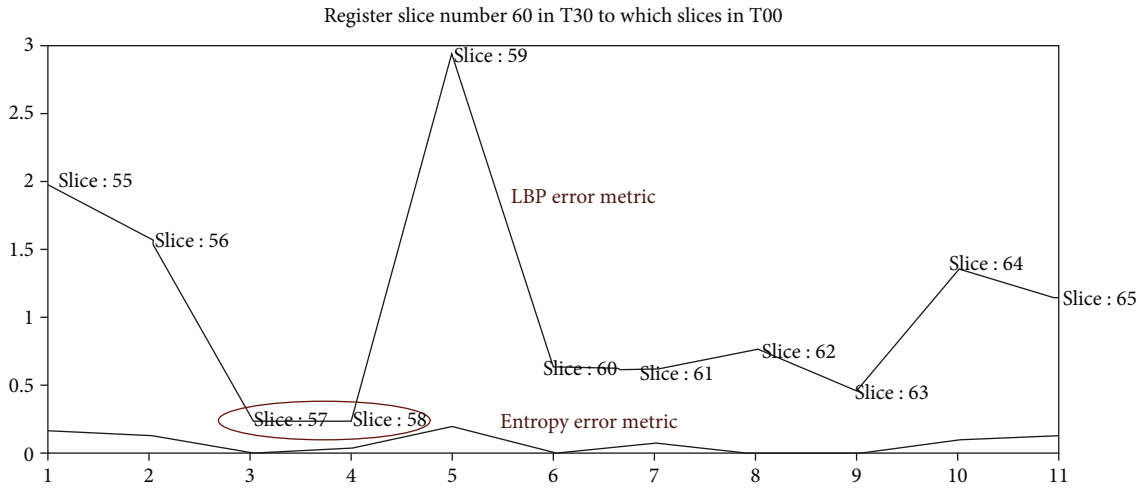


Figure 11: The registration from one image with the whole phase using LBP and entropy error metric.

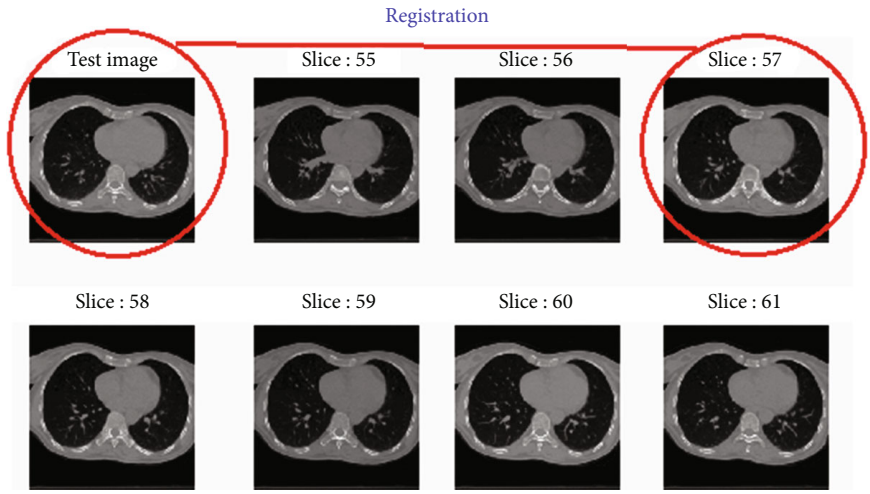


Figure 12: The registration from one image with the whole phase in visualization.

determining the inhalation and exhalation phases in a breathing cycle.

Within class variance method by Otsu [10] was applied to separate the foreground and background regions from the input image. Otsu's thresholding method involves iterating through all the possible threshold values and calculating a measure of spread for pixel values from each side of the threshold. The aim is to find the threshold value where the sum of foreground and background spreads is at the minimum.

We perform the following steps to segment left and right lung areas. First, we make the complement of the binary image. Two lungs will be represented in the complemented image (in Figure 5). Second, the body region (i.e., the outline of the patient's body) is multiplied with the complemented image to obtain the regions of two lungs (in Figure 6). Note that there remain some unexpected regions besides lung areas. Third, the center point of the body image is used to segment the two lungs exactly. The left and right lungs are now on the opposite sides of the center point and represented by the largest white regions in the multiplied image (Figures 7–9). Detailed calculation steps are described in Algorithms 2 and 3.

4. Deformable Image Registration

In this step, we need to locate the position of a slice belonging to one phase to match with another slice in a different phase. By matching the slice of two phases, we can register these slices and reconstruct the exhalation and inhalation phases.

4.1. Texture Matching by Local Binary Pattern. Before applying local binary pattern (LBP) [11] to the lung image, a LBP descriptor should be determined. First, we convert the input color image to grayscales, since LBP works only on grayscale images. For each pixel, we calculate the LBP value using its neighborhood. After calculating the LBP value of the pixel, we update the corresponding pixel location in the LBP mask,

1. Input the phase T00, T10, T90 are the checking phases. All SliceIM(i,j) is the SliceIM in the phase i and in the index j. And SliceIM must contain only the left and right lung partitions. Slices with valid lung segmentation are selected for modeling.
2. Registration
    a) Step through all possible phases = T10, .. T90
    b) Step through all possible slices =1, .. number of slice in the considered phase
    c) Calculate LBPErrorRate(i,j,phase) = CalcLBPErrorRate(Slice(slice,phase), Slice(slice-5:slice+5,T00)
    d) Calculate EntropyErrorRate(i,j,phase) = CalcEntropyErrorRate(Slice(slice,phase), Slice(slice-5:slice+5,T00)
    e) Find the index of slice with minimum LBP and Entropy Error Rate RegisterIdx = Index(LBPErrorRate(i,j,m), 2, 'smallest') UNION Index(EntropyErrorRate(i,j,m), 2, 'smallest')
    f) Store EntropyErrorRate and LBPErrorRate LBPErrorRateRegistrationResult(slice,phase) = LBPErrorRate(i, RegisterIdx, phase) and EntropyErrorRateRegistrationResult(slice,phase) = EntropyErrorRate(i, RegisterIdx, phase)
3. Signal Modeling
    a) Step through all possible phases = T10, .. T90
    b) Calculate the standard deviation of LBP Error Rate and Entropy Error Rate for each phase from slices STD_LBPErrorRate(phase) = StandardDeviation(LBPErrorRateRegistrationResult(:,phase)) and STD_EntropyErrorRate(phase) = StandardDeviation(EntropyErrorRateRegistrationResult(:,phase))
    c) Take the sum of error rates on each phase in registration to phase T00 ErrorRate(phase) = STD_LBPErrorRate(phase) + STD_EntropyErrorRate(phase)
    d) Signal Model by plotting the variation of error rates from phases T10, …, T90
    e) Evaluation If the signal increases, it represents the inhalation process and If the signal decreases, it represents the exhalation process
4. End
5. Result in Respiratory signal

Appendix

StandardDeviation method

1. Calculate Mean
    FOR i =0 to N
    sum = sum + X[i]
    next i
    ENDLOOP
    M = sum / N // Divides the sum by the total number, N, to get Mean
2. Calculate Variance
    FOR j =0 to N
    sumOfSquares = sumOfSquares + ((X[j] - M)^2) // etc...
    next j
    ENDLOOP
3. Standard Deviation
    stdDev = sqrt(sumOfSquares / (N -1))

Algorithm 6: Respiratory signal modeling.

which has the same matrix dimension as the input image, with the calculated LBP value.

Around each pixel, there are 8 neighboring ones. If the central pixel value is greater or equal to the value of a given neighboring pixel, the corresponding value in the binary array is set to 1, otherwise is set to 0. After calculating the LBP mask, we construct the LBP histogram. The LBP mask values range from 0 to 255, giving the LBP descriptor size of 1×256. Then, the LBP histograms are normalized for comparison. Figure 10 illustrates the application of LBP in comparing two contexts from two images.

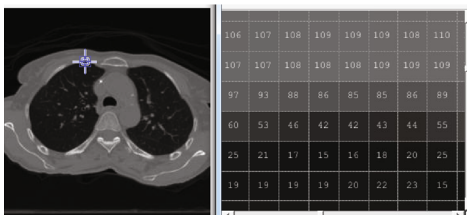
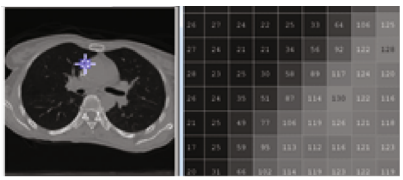
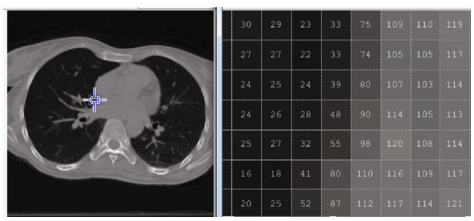
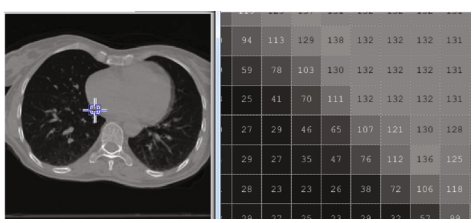
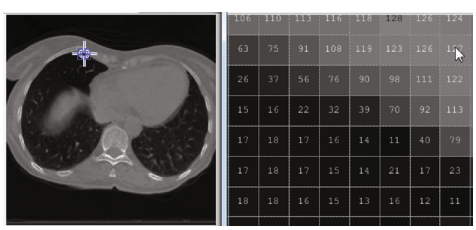
Next, we apply the LBP to create a metric for a comparison of slices in different phases. LBP will return a pair of slices with the most similarity in texture. By subtracting the LBP of two slices, we can extract the LBP error metric for the registration process. Algorithm 4 describes the method of calculating local binary pattern error rate.

4.2. Registration Decision by Entropy Error Measurement. Entropy is a measure of the disorder level of a system [12]. The more the disorder, the higher the entropy of the system. Two slices with the same entropy will have a high probability to be in the same registered position. Although the LBP helps us to make the texture matching between slices, in some specific cases, we can get wrong results or are unable to decide which slice in two or three slices having the similar LBP metrics. Therefore, entropy can support our decision in registration. By subtracting the entropy of two slices, we can get the entropy error metric for our registration process. Algorithm 5 shows the steps to calculate the entropy error rate of the images.

5. Modeling Respiratory Signals of Inhalation and Exhalation

In the process of modeling the respiratory of signals of inhalation and exhalation, we apply LBP and entropy methods in

Table 1: The ground truth lung segmentation from phase T20.

Pixel boundary specification	Threshold	Ground truth segmentation
	Slice 30 Phase T20 42	
	Slice 40 Phase T20 51	
	Slice 50 Phase T20 48	
	Slice 60 Phase T20 65	
	Slice 70 Phase T20 76	

Section 4. The following is an example of registering phase T30 to phase T00 and decide if the checking image belongs to inhaling or exhaling stages.

For example, for slide 60 in phase T30, we need to find the most similar slice in phase T00. Three steps are performed as follows (Figures 11 and 12 and Algorithm 6):

(i) Considering only the slices from 55 to 65 in T00 (the margin is 5 slices)

(ii) Comparing the LBP error metrics, there are two slices 57 and 58 with minimum LBP error metrics. We need to determine which one could be registered for slice 60 in phase T30

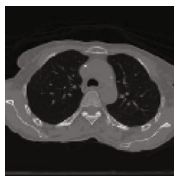
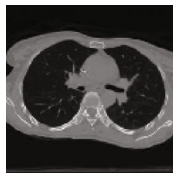
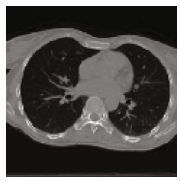
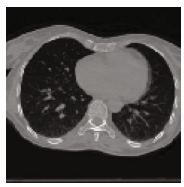
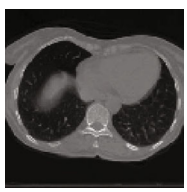
(iii) Comparing the entropy error metrics of the slices 57 and 58, we see that the slice 60 in T30 can be registered to the slice 57 in T00 because the entropy error metric of slice 57 is less than that of slice 58

(iv) After registration, we notice that the process from phases T00 to T30 is the inhalation stage of the breathing process of a patient

6. Evaluation of Experimental Results

6.1. Ground Truth Lung Segmentation Determination. The DIR database provides 4DCT lung image datasets from the phase indexes T00 to T90. In each phase, a 4DCT dataset contains 94 images which are scanned from the top to the bottom of a patient lung. However, there is no ground truth lung segmentation that is specified by a specialist. To solve this problem, the ground truth segmentation is determined

Table 2: The DSC measurement to evaluate the lung segmentation on ground truth in phase T20.

Slide	Original slice	Ground truth	Segmentation	DSC
Slice 30, phase T20				98.08%
Slice 40, phase T20				98.54%
Slice 50, phase T20				96.13%
Slice 60, phase T20				99.61%
Slice 70, phase T20				97.68%

based on grayscale pixel values on the boundary between the lung and body partitions.

Table 1 shows the ground truth segmentation method of some slices in phase T20. The ground truth lung segmentation has an important role in the next processes of registration and modeling. If the segmentation is not close to the real lung partition, all following calculated comparison metrics in the registration will give unexpected results.

6.2. Lung Segmentation Evaluation. To evaluate the quality of lung segmentation for the left and right partitions, we use Dice's similarity coefficient (DSC), which measures the volume overlap percentage. The DSC is described as

$$\mathrm{DSC} = \frac{V_s \cap V_t}{(V_s + V_t)/2}.100, \tag{1}$$

where V_s is the volume of the left (or right) experimental segmentation and V_t is the volume of the corresponding ground truth segmentation. The closer to 100% DSC is, the better confident and efficient the segmentation is. Table 2 shows that the result of DSC is from 96% to 99%. The slices from 30 to 70, which are the major slices in a phase dataset, have especially high DSC values.

6.3. Deformable Lung Registration Evaluation. For evaluation of registration, we apply the coefficient of variation (CVar) to compare with the registration for other datasets. The formula for CVar is

$$\mathrm{CVar} = \frac{s}{\bar{X}}.100, \tag{2}$$

where s and $\bar{X}$ are the standard deviation and the mean of all registration results, respectively.

In a 4DCT dataset, not all slices contribute to the registration or modeling the respiratory phase. In general, only the slices from 30 to 70 are significant in the comparison because they have clear lung segmentation information. Table 3 shows that lung information is trivial or undeterminable for images outside that range.

Table 4 demonstrates the coefficient of variation from the phases T10 to T90. In this experiment, the source phase is T00 and the target phase is T10 to T90. The

Table 3: The slices from 1 to 30 and 70 to 94 are unused and insignificant for registration because of little lung information.

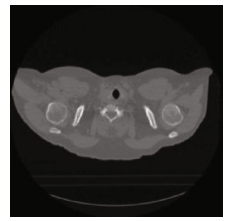
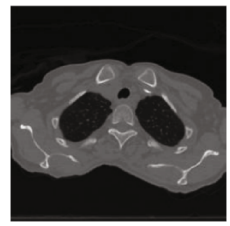
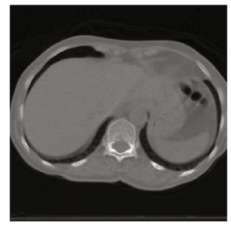
Slice index	Slice lung information	Remark
01		No lung information
15		Too little lung information
80		The undetermined shape of the lung
90		No left lung information

Table 4: The CVar measurement of the registration process from target phases T10 to T90 in registration to the phase T00.

Registration of the target phase to the source phase T00	Slice index range in calculation	CVar measurements	Remark
T10	Index from 30 to 70	1.6117	Confident
T20	Index from 30 to 70	1.9134	Acceptable
T30	Index from 30 to 70	2.0173	Relatively high
T40	Index from 30 to 70	2.1167	Relatively high
T50	Index from 30 to 70	1.9572	Acceptable
T60	Index from 30 to 70	1.9207	Acceptable
T70	Index from 30 to 70	1.9559	Acceptable
T80	Index from 30 to 70	1.8909	Acceptable
T90	Index from 30 to 70	1.2455	Confident

CVars are small enough to indicate high confidence. The registration of phases T10 and T90 is more confident. The registration of T20, T60, T70, and T80 has acceptable CVars. The CVars for T30 and T40 are high but are still controllable.

6.4. Respiratory Signal Evaluation. Inhalation (exhalation) is a process of inbreathing (breathing). The lung becomes small (large) in the inhalation (exhalation) stage. If phase T00 is the starting of the inhalation process, the error rate of LBP and entropy will be small in the registration for T10 and T00. On the contrary, if any phase in registration to T00 has a high error rate LBP and entropy, that phase is in the exhalation stage.

In Table 5, the sum of standard deviations of LBP and entropy error rate on each slice from 30 to 70 in each phase is calculated. LBP and entropy error rate are the appropriate metrics to represent the inhalation and exhalation of a lung. Starting from phase T00, the error rate summation increases in phase T50 and decreases in phase T90. Figure 13, which illustrates the values in Table 5, shows that registration and modeling are successful.

Figure 14 describes the overall framework for this registration step. In this workflow, the artifact removal and lung segmentation are applied for testing images and reference images. The registration process with LBP and entropy measurements is the key for checking the best candidate before giving the final decision in choosing one of the two states,

Table 5: The sum of LBP and entropy error rate represents the respiratory signal for slides from 30 to 70.

Registration of target phase to the source phase T00	Slice index	Sum of standard deviation of LBP and entropy error rate	Remark
T10	Index from 30 to 70	0.1370	Start inhalation
T20	Index from 30 to 70	0.3220	Inhalation
T30	Index from 30 to 70	0.4321	Inhalation
T40	Index from 30 to 70	0.5655	Inhalation
T50	Index from 30 to 70	0.7124	Start exhalation
T60	Index from 30 to 70	0.6361	Exhalation
T70	Index from 30 to 70	0.4863	Exhalation
T80	Index from 30 to 70	0.4721	Exhalation
T90	Index from 30 to 70	0.1552	Finish

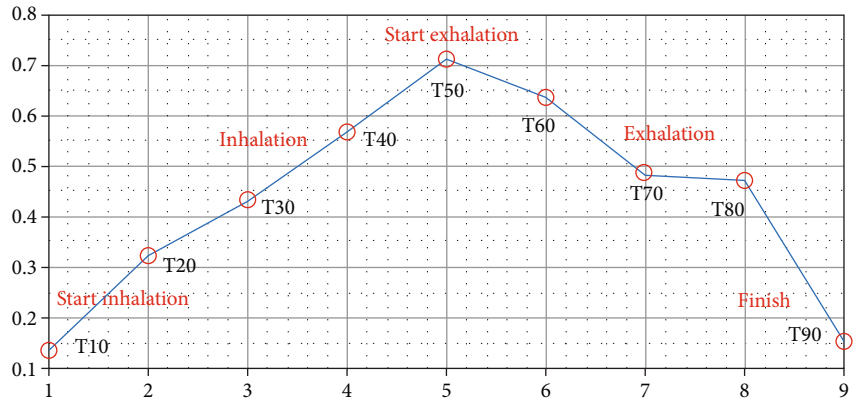


Figure 13: The demonstration of respiratory signal from phases T10 to T90 in registration to phase T00.

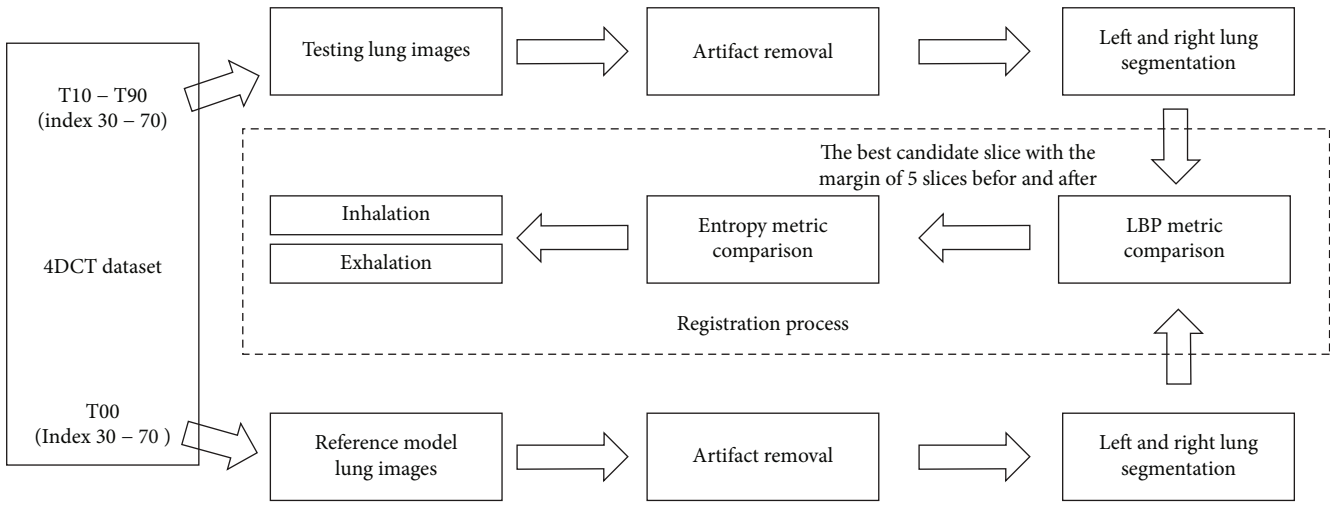


Figure 14: The overall framework of the proposed registration process with LBP and entropy measurements.

inhalation or exhalation. The reference model lung images are used from the source phase T00, and the testing lung images are used from the target phases T10-T90. In each phase, the images with indexes from 30 to 70 are used because there are available lung segments in this index range.

In comparison with the learning-based approach in registration problem VoxelMorph (Balakrishnan et al.) [13, 14], diffeomorphic (Mok and Chung) [15], and DeepFLASH (Wang and Zhang) [16] registration, we have the following conclusion about advantages and disadvantages.

These approaches are applied mainly to the MRI brain images, which are more complicated with at least five segments: background, skull, white matter, gray matter, and cerebrospinal fluid. Moreover, the movements of these segments are also too difficult to track. Therefore, the authors Balakrishnan et al. and Mok and Chung propose the learning-based approach using a feature map from U-Net to minimize the loss function. Their approaches require large amounts of data for feeding and tracking in the training process.

Because we only want to model the respiratory signals, using U-Net is more complicated than necessary in the lung registration step. This is the main point of using a hybrid LBP descriptor with entropy registration in our approach. We do experiments for VoxelMorph and diffeomorphic methods with our data. The Dice measurements of VoxelMorph and diffeomorphic methods are 90% and 97%, respectively, in comparison with 96% of our proposed method. If we feed more training data, the result of VoxelMorph and DeepFLASH would be higher. Another comparison is with the DeepFLASH method, which applies the duel net with frequency spectrum domain. The Dice measurement for the DeepFLASH method is 91%. Similarly, if we continue training, the result might be improved. The advantage of our approach is that it is a fast and effective method for modeling respiratory. This method does not require more data for feeding training. We only need the reference model lung images to control the modeling.

7. Conclusion

There are two stages in the process of registration and respiratory modeling for the 4DCT image. The first stage, which is essential to the whole process, is lung segmentation, and the second stage is registration and modeling. If the artifacts are not removed completely, the subsequent metrics used in the registration and modeling give incorrect results. The more accurate segmentation is performed, the more accurate registration is obtained. Therefore, the minimum variance quantization and within class variance are combined for a good segmentation.

After segmentation, the LBP and entropy are applied in sequence to perform the registration. LBP can be used to find near context information between two images in different phases. Then, the entropy verifies and decides the correct registered image. If LBP and entropy are applied independently, the result becomes incorrect. Because all images in neighbor slices are similar in visualization, our method enhances efficiency of the automatic process in registration and respiratory modeling for the 4DCT datasets.

In summary, our proposed approach in modeling respiratory signals by deformable image registration on 4DCT lung images has some discriminant and promising features in comparison to conventional and deep learning approaches as follows:

(i) We construct a complete process from segmentation, registration, and modeling with careful selections from the minimum variance quantization method, LBP feature descriptor to entropy measurement to minimize the complexity of the process

(ii) We still ensure the high accuracy in segmentation via DSC measurement and in registration via CVar measurement, as well as in modeling via LBP and entropy error rate

(iii) We do not need too many images like other deep learning approaches for training data

(iv) We can have a comparative and robust result in comparison to other traditional computer vision approaches

(v) The results of DSC, CVar, and entropy in segmentation, registration, and modeling can be applied as parameters for constructing loss function in deep learning approaches

Besides the above advantages, the only limitation is that our approach cannot work well if the background illumination is quite different between the reference and test images.

Data Availability

This research uses public data offered by DIR-LAB.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is funded by Vietnam National University Ho Chi Minh City under grant number C2017-18-08.

References

[1] T. Pang, S. Guo, X. Zhang, and L. Zhao, "Automatic lung segmentation based on texture and deep features of HRCT images with interstitial lung disease," *BioMed Research International*, vol. 2019, Article ID 2045432, 8 pages, 2019.

[2] T. Peng, T. C. Xu, Y. Wang et al., "Hybrid automatic lung segmentation on chest CT scans," *IEEE Access*, vol. 8, pp. 73293–73306, 2020.

[3] X. Hu, J. Yang, and J. Yang, "A CNN-based approach for lung 3D-CT registration," *IEEE Access*, vol. 8, pp. 192835–192843, 2020.

[4] J. Wu, B. Hu, and X. Yang, "Deformable registration of 4D-CT lung image using landmark tracking," *Biomedical Research*, vol. 27, no. 3, 2016.

[5] D. Yang, W. Lu, D. A. Low, J. O. Deasy, A. J. Hope, and I. el Naqa, "4D-CT motion estimation using deformable image registration and 5D respiratory motion modeling," *Medical Physics*, vol. 35, no. 10, pp. 4577–4590, 2008.

[6] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, and H. Handels, "Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 251–265, 2011.

[7] M. George and R. Zwiggelaar, "Comparative study on local binary patterns for mammographic density and risk scoring," *Journal of Imaging*, vol. 5, no. 2, p. 24, 2019.

[8] J. Ehrhardt and C. Lorenz, Eds., "4D modeling and estimation of respiratory motion for radiation therapy," in *Biological and Medical Physics, Biomedical Engineering*, vol. 10, pp. 27–32, Springer, Berlin, 2013.

[9] P. Rambabu and C. Nagaraju, "The optimal thresholding technique for image segmentation using fuzzy ostu method," *International Journal of Applied Engineering Research*, vol. 10, no. 13, pp. 33842–33846, 2015.

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

[11] B. Xiao, K. Wang, X. Bi, W. Li, and J. Han, "2D-LBP: an enhanced local binary feature for texture image classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 9, pp. 2796–2808, 2019.

[12] M.-H. Lim and P. C. Yuen, "Entropy measurement for biometric verification systems," *IEEE transactions on cybernetics*, vol. 46, no. 5, pp. 1065–1077, 2016.

[13] Y. Sugawara, H. Tachibana, S. Moriya, and A. Sawant, "SU-E-J-266: a pitfall of a deformable image registration in lung cancer," *Medical Physics*, vol. 41, no. 6Part10, 2014.

[14] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: a learning Framework for deformable medical image registration," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[15] T. C. W. Mok and A. C. S. Chung, "Fast symmetric diffeomorphic image registration with convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4644–4653, Seattle, WA, USA, 2020.

[16] J. Wang and M. Zhang, "DeepFLASH: an efficient network for learning-based medical image registration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4444–4452, Seattle, WA, USA, 2020.

Hindawi
BioMed Research International
Volume 2021, Article ID 5566885, 9 pages
https://doi.org/10.1155/2021/5566885

Research Article

Assessment of MRI-Based Radiomics in Preoperative T Staging of Rectal Cancer: Comparison between Minimum and Maximum Delineation Methods

Haidi Lu,[1] Yuan Yuan,[1] Zhen Zhou,[1] Xiaolu Ma,[1] Fu Shen ,[1] Yuwei Xia,[2] and Jianping Lu[1]

[1]*Department of Radiology, Changhai Hospital, No. 168 Changhai Road, Shanghai, China*
[2]*Huiying Medical Technology Co., Ltd., B2, Dongsheng Science and Technology Park, HaiDian District, Beijing, China*

Correspondence should be addressed to Fu Shen; ssff_53@163.com

Received 6 January 2021; Revised 24 May 2021; Accepted 2 July 2021; Published 12 July 2021

Academic Editor: Lin Gu

The manual delineation of the lesion is mainly used as a conventional segmentation method, but it is subjective and has poor stability and repeatability. The purpose of this study is to validate the effect of a radiomics model based on MRI derived from two delineation methods in the preoperative T staging of patients with rectal cancer (RC). A total of 454 consecutive patients with pathologically confirmed RC who underwent preoperative MRI between January 2018 and December 2019 were retrospectively analyzed. RC patients were grouped according to whether the muscularis propria was penetrated. Two radiologists segmented lesions, respectively, by minimum delineation (Method 1) and maximum delineation (Method 2), after which radiomics features were extracted. Inter- and intraclass correlation coefficient (ICC) of all features was evaluated. After feature reduction, the support vector machine (SVM) was trained to build a prediction model. The diagnostic performances of models were determined by receiver operating characteristic (ROC) curves. Then, the areas under the curve (AUCs) were compared by the DeLong test. Decision curve analysis (DCA) was performed to evaluate clinical benefit. Finally, 317 patients were assessed, including 152 cases in the training set and 165 cases in the validation set. Moreover, 1288/1409 (91.4%) features of Method 1 and 1273/1409 (90.3%) features of Method 2 had good robustness ($P < 0.05$). The AUCs of Model 1 and Model 2 were 0.808 and 0.903 in the validation set, respectively ($P = 0.035$). DCA showed that the maximum delineation yielded more net benefit. MRI-based radiomics models derived from two segmentation methods demonstrated good performance in the preoperative T staging of RC. The minimum delineation had better stability in feature selection, while the maximum delineation method was more clinically beneficial.

1. Introduction

Rectal cancer (RC) is one of the most frequently diagnosed malignancies worldwide [1]. Accurate preoperative assessment of T staging of rectal cancer is a critical step in clinical treatment strategy, where a total mesorectal excision (TME) is considered as an optimal treatment approach for early staged RC (T1–2 and N-), while the treatment strategy for a locally advanced stage of RC (T3–4 and/or N+) is neoadjuvant chemotherapy (CRT) before TME [2, 3].

Currently, magnetic resonance imaging (MRI) is the common first-line modality for accurate pretreatment assessment of patients with RC. Moreover, rectal high-resolution T2-weighted images (T2WIs) have a vital role in the preoperative T staging of RC [3–5]. However, when there is an invasion of muscular layers by vessels, exudative changes around the lesion, and desmoplastic reaction, it is often hard to distinguish them from tumor infiltration outside the intestinal wall, which often leads to common mistakes in the staging of T2 and early T3 [4, 5].

Radiomics, a novel noninvasive tool, has shown multiple gratifying advantages in the preoperative assessment, prediction of treatment outcome, and distant metastasis of RC [6–10], thereby providing important details of tissue features, including the preoperative T staging. Among the factors that affect radiomics analysis, segmentation is vital as the first step

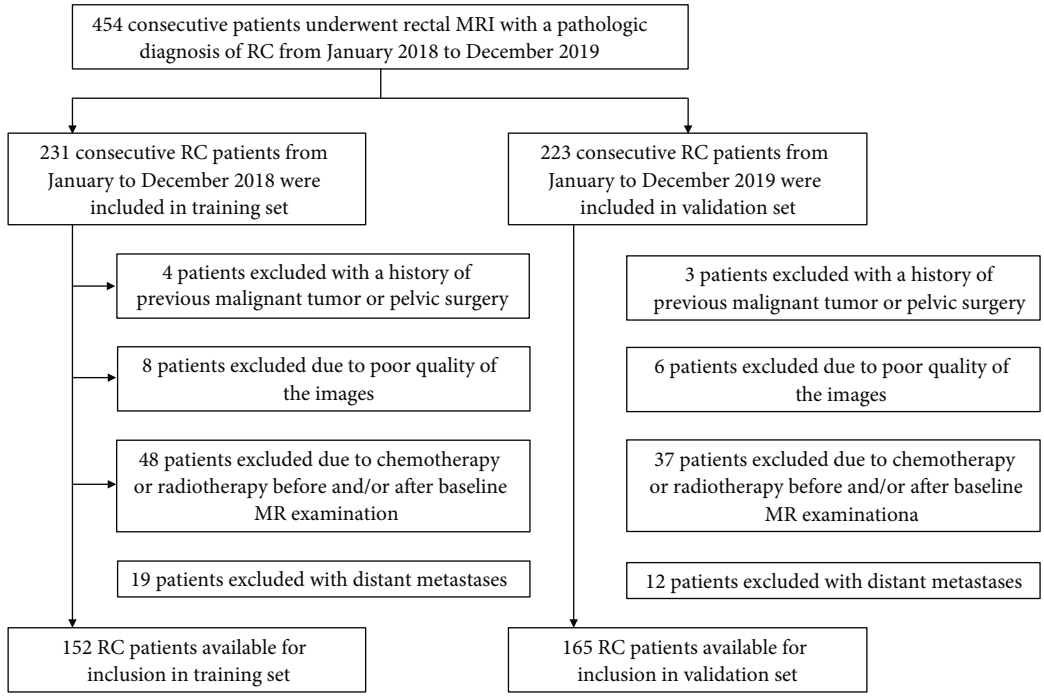


Figure 1: Diagram for the inclusion of patients into the study. RC: rectal cancer.

of the imaging process. Still, recent publications have demonstrated that manual delineation of lesions is mainly used as a conventional segmentation method, but it is subjective and has poor stability and repeatability [9–11]. Zhang et al. [12] showed that delineation discrepancy in volumes of interest (VOIs) might affect predicting the performance of nasopharyngeal carcinoma and breast cancer radiomics models.

Some studies have reported on manual delineation based on MR images in RC patients. Most methodologies advocate using the volume of the whole primary tumor, which is manually drawn along the border of the tumor on each axial slice to cover the lesion [7, 8, 13–18]. Yet, most studies have no precise definition of the outer edge of the tumor. The type of manual segmentation method that can yield higher clinical benefit in patients with RC has been less discussed and requires further quantitative assessment. Therefore, the aim of our study was to validate and compare different radiomics tumor delineation models in evaluating the repeatability of feature extraction and exploring the preoperative T staging of RC based on high-resolution T2WI.

2. Materials and Methods

2.1. Participants. 454 consecutive patients with RC who underwent 3.0 T rectal MRI before surgical resection at Changhai hospital between January 2018 and December 2019 were retrospectively assessed. Inclusion criteria were (1) pathologically confirmed RC with baseline MRI data, (2) baseline MRI within 14 days before surgical resection, and (3) single focus. Exclusion criteria were (1) a history of previous malignant tumor or pelvic surgery ($n = 7$), (2) poor quality of the images ($n = 14$), (3) received any treatment before and/or after baseline MR examination ($n = 85$), and (4) distant metastases ($n = 31$).

Based on the National Comprehensive Cancer Network (NCCN) and American Joint Committee on Cancer (AJCC) staging system [19], the patients were grouped according to different pathological T stages: T1–2 as a group without the penetrated muscularis propria and T3–4 as the group with penetration.

The training dataset and validation dataset were chronologically divided: 152 consecutive RC patients between January and December 2018 were included in the training set, while 165 consecutive RC patients between January 2019 and December 2019 were enrolled in the validation set (Figure 1).

The present study received approval from the local Institutional Review Board (Committee on Ethics of Biomedicine, Changhai Hospital). Informed consent was waived for this retrospective study.

2.2. Imaging Acquisition. Rectal MRI was scanned on two 3.0 T MR systems (Siemens Skyra 3.0T and GE Discovery 750w 3.0T) using a phased array coil. Before scanning, intestinal cleaning was performed by enema administration with 20 ml of glycerin. Oblique-axial high-resolution T2WI was perpendicular to the long axis of the rectum comprising the lesion. Routine sequences including sagittal T2WI, axial diffusion-weighted images (DWI, b-value: 0, 1000 s/mm^2), axial T1-weighted images (T1WI), and gadolinium contrast-enhanced T1WI of the pelvis were obtained in the sagittal, coronal, and axial planes. Details on parameters applied for high-resolution T2WI, which were used for radiomics models, are shown in Supplemental Table 1.

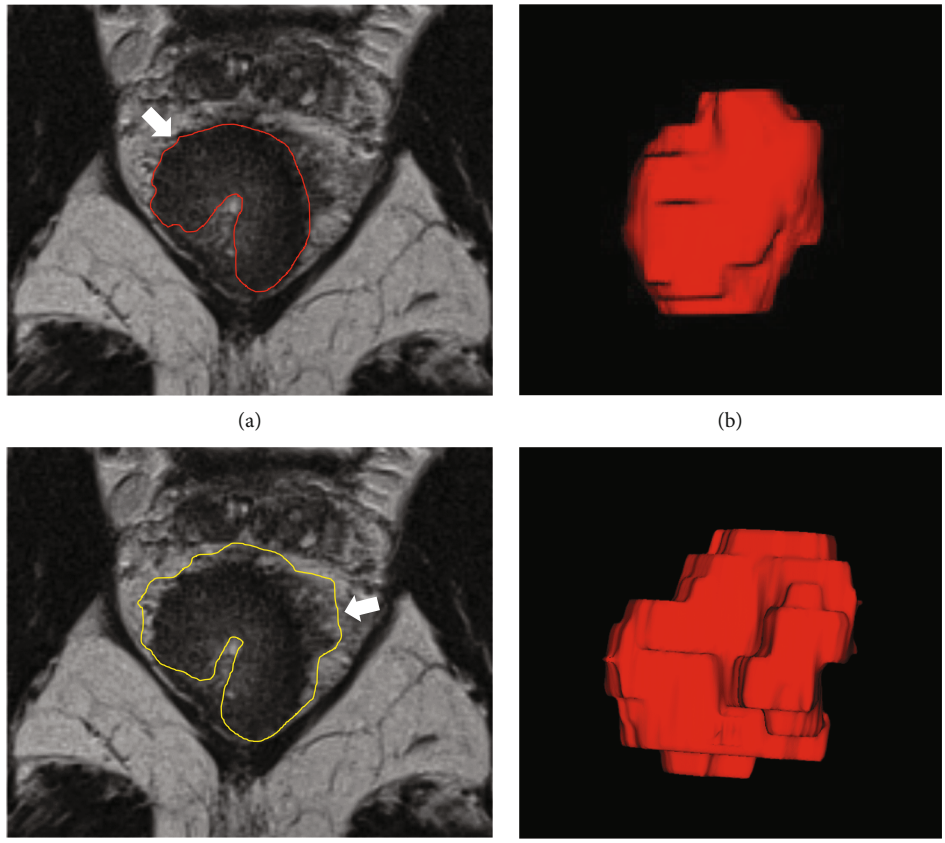


Figure 2: Representative images for lesion delineation. (a, b) Minimum delineation of ROI on oblique-axial T2-weighted MR images (arrow) and volume renderings of VOIs (Method 1). (c, d) Maximum delineation of ROI on oblique-axial T2-weighted MR images (arrow) and volume renderings of VOIs (Method 2).

2.3. Image Segmentation. All original high-resolution T2WI DICOM data were uploaded to the Huiying Medical Radcloud radiomics platform (http://radcloud.cn/). As the T2W images were required from two different MR systems in our study, image normalization was essential for all data to achieve homogeneity. Each image intensity was normalized to minimize the MRI signal variations using the following formula:

$$f(x) = \frac{s(x - \mu_x)}{\sigma_x}, \quad (1)$$

where $f(x)$ indicates the normalized intensity, x indicates the original intensity, μ refers to the mean value, σ indicates the variance, and s is an optional scaling, which is by default set to 1. While reserving the diagnostic intensity discrepancy, the signal discrepancy in MR parameters was decreased for subsequent radiomics analysis.

The region of interest (ROI) of each lesion was manually delineated slice-by-slice on high-resolution T2W images. We used two kinds of manual segmentations for ROI: Method 1—minimum delineation and the smallest and clearest solid border that best fit the tumor region, excluding the blurry region of the margin; Method 2—maximum delineation, while the maximum margin of the lesion, including the entire region of perirectal tissues, was used to define the ROI (Figure 2). Then, the volume of interest (VOI) was reconstructed through the ROIs.

2.4. Feature Extraction and Reduction. Two radiologists with 8 (H.L.) and 5 years (Z.Z.) working experience in abdominal imaging independently reviewed all these images, who were blinded to the patient information. Next, all delineations were checked by one senior radiologist (Y.Y., who had 10 years of working experience in rectal MRI). Two radiologists (H.L. and Z.Z.) performed image processing of all cases on the platform, comprising Method 1 and Method 2, respectively. One radiologist (H.L.) repeated the segmentations of all cases one week later for final feature selection.

1409 radiomics features were extracted from each method of segmentation with the above platform. All features were grouped into four categories: (1) first-order features, which quantitatively delineated the distribution of voxel intensities of MR image by basic indexes; (2) shape-based features, including the shape and size of the VOI (e.g., the volume of segmentation); (3) texture features and quantification of the region heterogeneity differences; (4)

Table 1: Pathological characteristics of the patients.

Variables		Training set ($n = 152$)	Validation set ($n = 165$)	P value
Gender	Male	94 (61.8%)	109 (66.1%)	0.434
	Female	58 (38.2%)	56 (33.9%)	
Age (years)		58.9 ± 8.3	57.5 ± 8.8	0.147
BMI (kg/m^2)		23.8 ± 3.2	23.5 ± 3.1	0.397
Tumor location	Upper	36 (23.7%)	32 (19.4%)	0.648
	Middle	92 (60.5%)	105 (63.6%)	
	Lower	24 (15.8%)	28 (17.0%)	
Histological type	Adenocarcinoma	131 (86.2%)	146 (88.5%)	0.325
	Mucinous adenocarcinoma	15 (9.9%)	17 (10.3%)	
	Signet ring cell carcinoma	6 (3.9%)	2 (1.2%)	
Differentiation	High	20 (13.2%)	17 (10.3%)	0.713
	Moderate	112 (73.7%)	127 (77.0%)	
	Poor	20 (13.2%)	21 (12.7%)	
T stage	T1	22 (14.5%)	17 (10.3%)	0.320
	T2	44 (28.9%)	51 (30.9%)	
	T3	74 (48.7%)	90 (54.5%)	
	T4	12 (7.9%)	7 (4.2%)	
N stage	N0	94 (61.8%)	99 (60.0%)	0.056
	N1	37 (24.3%)	28 (17.0%)	
	N2	21 (13.8%)	38 (23.0%)	
Tumor deposit	Negative	118 (77.6%)	137 (83.0%)	0.226
	Positive	34 (22.4%)	28 (17.0%)	
Lymphovascular invasion	Negative	91 (59.9%)	100 (60.6%)	0.893
	Positive	61 (40.1%)	65 (39.4%)	
Perineural invasion	Negative	106 (69.7%)	117 (70.9%)	0.819
	Positive	46 (30.3%)	48 (29.1%)	
Tumor budding	Negative	114 (75.0%)	126 (76.4%)	0.777
	Positive	38 (25.0%)	39 (23.6%)	
CEA*	Negative	107 (70.4%)	115 (69.7%)	0.892
	Positive	45 (29.6%)	50 (30.3%)	
CA19-9*	Negative	126 (82.9%)	126 (76.4%)	0.150
	Positive	26 (17.1%)	39 (23.6%)	

BMI: body mass index; CEA: carcinoembryonic antigen; CA19-9: carbohydrate antigen 19-9. *Preoperative blood samples.

higher order features, which included the transformation of first-order statistics and shape and texture characteristics, such as logarithm, exponential, gradient, square, square root, local binary patterns (LBP), and wavelet transformation [7, 8].

The inter- and intraclass correlation coefficient (ICC) was calculated to assess the reliability and reproducibility of all features. Features with both inter- and intraobserver ICCs exceeding 0.8 were applied for subsequent analysis, which suggested good robustness of features. To reduce the redundant features and select the optimal features, the variance threshold algorithm (variance threshold = 0.8) and Select-K-Best algorithm were adopted. The Select-K-Best algorithm used $P < 0.05$ to determine optimal features related to the T stage.

2.5. Machine Learning and Model Analysis. The radiomics analysis was performed in the Radcloud platform. Based on the selected features, the radiomics-based model was constructed with the support vector machine (SVM) in the training set, then verified in the validation set. For SVM, details of the parameters, kernel (linear), penalty coefficient (1), gamma (auto), class weight (balanced), decision function shape (one-to-many), and random state (NA), were used.

To assess the model's diagnostic performance, the receiver operator characteristic (ROC) curve was obtained by calculating areas under the curve (AUCs) in both datasets. The DeLong test was performed to evaluate differences between the ROC curves. The clinical benefits of radiomics models were estimated by decision curve analysis (DCA). Statistical significance was defined as $P < 0.05$.

Table 2: Selected radiomics features.

Model	No	Radiomics feature	Radiomics class	Filter
Method 1	1	Skewness	First order	Wavelet-HLL*
	2	Maximum	First order	Wavelet-HLL*
	3	High gray level zone emphasis	GLSZM	Wavelet-HLH*
	4	Gray level nonuniformity	GLSZM	Wavelet-LHL*
Method 2	1	Skewness	First order	Wavelet-HLL*
	2	High gray level zone emphasis	GLSZM	Wavelet-LHL*
	3	Skewness	First order	Wavelet-LHL*
	4	High gray level run emphasis	GLRLM	Original
	5	High gray level run emphasis	GLRLM	Logarithm
	6	High gray level run emphasis	GLRLM	Square root
	7	High gray level run emphasis	GLRLM	Wavelet-LLL*

GLSZM: gray level size zone matrix; GLRLM: gray level run length matrix. *The wavelet transform decomposes the tumor area image into low-frequency components (L) or high-frequency components (H) in the x, y, and z axes. Method 1: minimum delineation method; Method 2: maximum delineation method.

3. Statistical Analysis

The Kolmogorov-Smirnov statistical test was used to test for the normality in all continuous variables. A paired Student's t -test or Wilcoxon test was used to compare variables between the two groups. Qualitative variables were assessed by the chi-square test or Fisher's exact test. SPSS software (version 20.0, Chicago, IL, USA) and R software (version 3.4.3) were used for statistical analysis. A P value of <0.05 was considered to be a statistically significant difference.

4. Results

4.1. Participant Characteristics. A total of 317 patients were finally enrolled. There was no significant difference between the training and validation sets. The patient characteristics and pathological outcomes are summarized in Table 1. According to the T stage by postoperative pathological examination, 183 patients (57.7%) were assigned to the penetration group.

4.2. Radiomics Features. All radiomics features extracted from Method 1 and Method 2 with ICCs ranged from 0.005 to 1.000. 1288/1409 features of Method 1 (91.4%) and 1273/1409 features of Method 2 (90.3%) had good robustness and were applied for subsequent analysis (both inter- and intraobserver ICCs ≥ 0.8). There was a significant statistical difference ($Z = 18.574$, $P < 0.001$) between the two methods.

The median (quartile range) volume of the two methods was 5.981 (2.490, 13.907) cm^3 and 11.617 (5.594, 31.117) cm^3, respectively. There was a significant difference in tumor size between Method 1 and Method 2 ($Z = 3.29$, $P = 0.001$).

Finally, 4 optimal features (Method 1) and 7 optimal features (Method 2) associated with T stage were selected to build the radiomics models (Model 1 and Model 2) (Table 2 and Supplemental Figure 1).

4.3. Performance of Radiomics Model. The ROC curves of the SVM classifier showed good performance with AUCs of 0.838 and 0.928 for Model 1 and Model 2 in the training set, respectively. For estimating differences in the two models in the validation set, Model 2 had an AUC of 0.903 (95% CI: 0.807-0.999), with a sensitivity of 87.0% and specificity of 82.3%, indicating a better performance compared with Model 1 that had an AUC of 0.808 (Figure 3). The DeLong test showed a significant difference ($P = 0.035$). Details contained in the models are shown in Table 3.

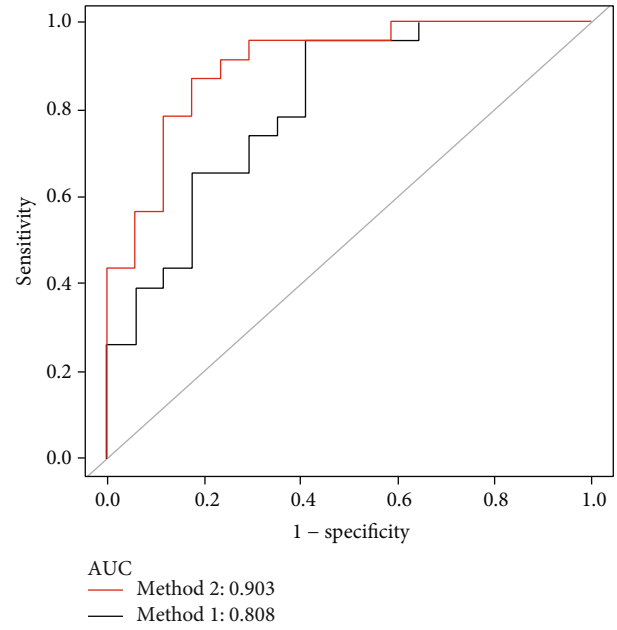


Figure 3: Receiver operator characteristic (ROC) curves in the validation set. AUC was 0.808 for the minimum delineation model (Method 1); AUC was 0.903 for the maximum delineation model (Method 2).

The decision curves demonstrated better performance of SVM models in predicting the T stage of RC than either the "all" or the "none" scheme at a threshold probability of 0.0-0.9 (Figure 4). The DCA showed that the Model 2 algorithm added more net benefit than that of Model 1.

Table 3: ROC analysis of the prediction model for the training and validation sets.

	Training set		Validation set	
	Method 1	Method 2	Method 1	Method 2
AUC	0.838	0.928	0.808	0.903
95% CI	0.764-0.912	0.864-0.992	0.669-0.947	0.807-0.999
Sensitivity	0.871	0.903	0.956	0.870
Specificity	0.805	0.866	0.588	0.823
Accuracy	0.823	0.876	0.800	0.850
PLR	4.464	6.733	2.323	4.927
NLR	0.160	0.112	0.074	0.158
PPV	0.628	0.718	0.759	0.870
NPV	0.943	0.960	0.909	0.823
P^*	0.036		0.035	

PLR: positive likelihood ratio; NLR: negative likelihood ratio; NPV: negative predictive value; PPV: positive predictive value. *Compared by DeLong test.

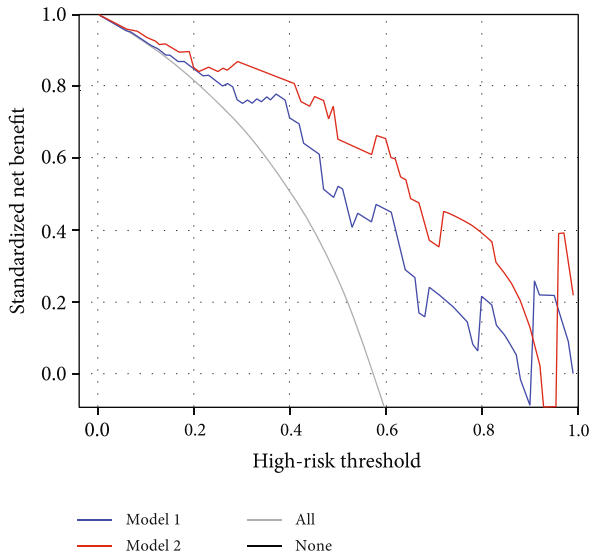


Figure 4: Decision curve analysis (DCA) of the two schemes of delineation. DCA showed that at the probability threshold of 0.0 to 0.9, the SVM model based on the maximum algorithm provided more net benefit than utilizing the minimum delineation scheme. Model 1: minimum delineation method; Model 2: maximum delineation method.

5. Discussion

Our work showed that Method 2 had a better value in differentiating T1-2 from T3-4. Although the statistical difference was found between the two manual segmentations of MRI-based radiomics in ROC, Method 1 gained more stability and repeatability.

Due to the diverse treatments and prognoses, the distinction between T1-2 and T3-4 is quite important as it can prevent undertreatment or overtreatment. Among the widely used imaging methods, high-resolution MRI is the most commonly used imaging approach for this purpose. Even though rectal high-resolution T2WI is suggested for the conventional preoperative staging of RC, differentiation between T2 and early T3 tumors is still unsatisfactory [20, 21]. One common misunderstanding is caused by penetration to the muscular propria layers by small vessels and desmoplastic reaction, which may lead to a great challenge in staging by using traditional imaging methods [4, 5].

Previous academic studies have demonstrated that radiomics have good performance in evaluating many types of tumors and can be utilized as a profitable noninvasive modality for the local staging in RC [6–12]. The workflow involves acquisition and segmentation of images and extraction and reduction of features, and when the features are selected, a statistical model is established [10]. Among the factors that affect radiomics analysis, segmenting is essential as the imaging processing step. There are three segmentation methods: manual, semiautomatic, and automatic, each of which has its advantages and disadvantages. At present, manual delineation of the ROI is most commonly used as a conventional segmentation method; however, it is subjective and has poor stability and repeatability [22].

In the present study, two different manual segmentations were utilized to explore the influence of diverse delineation on the stability of feature selection and preoperative T staging's diagnostic efficiency. The inter- and intraclass correlation coefficients of features were computed. Our results showed that features based on minimum delineation had high robustness, which suggested good reliability and reproducibility.

Meanwhile, our results also showed that the diagnostic performance of radiomics models could be affected by delineation discrepancy. The above analysis indicated that the SVM model based on maximum delineation had a higher predictive performance than the minimum delineation model ($P < 0.05$) for T stage classification, thus suggesting good diagnostic efficiency. In their nasopharyngeal carcinoma and breast cancer studies, Zhang et al. [12] built a quantitative image postprocessing algorithm that demonstrated delineation differences in segmentation affecting radiomics-based diagnostic performance. Kocak et al. [23] analyzed the effect of radiomics segmentation with margin shrinkage in the evaluation of renal carcinomas. Nevertheless, manual segmentation tends to lead to the excessive delineation of the lesion border to ensure the entire lesion is recognized in most clinical practices [24]. Our clinical decision-making curves revealed that the clinical benefits of the maximum delineation algorithm were greater than the minimum approach in the evaluation of the T stage in RC patients, which is consistent with previous research [12, 25–29] and could be explained by the dilated margin of perirectal tissues containing complex information about identifying tumor heterogeneity.

This present study has several limitations. First, VOIs were manually delineated instead of being semiautomatically/automatically segmented, thus making it difficult to avoid subjective errors and making it unsuitable for large-scale data processing [30, 31]. Studies had indicated that semiautomated/automated segmentations can provide the reproducible and accurate estimates of the tumor [31–34]. However, similar

to the previous studies, which used manual segmentation in RC patients, these studies described a semiautomated/automated delineated manner along the tumor's outer edge on each consecutive slice, with no precise definition of the border of the whole lesion. Second, this was a retrospective single-center cohort study without external validation. Therefore, a future multicenter study is required to verify our findings. Finally, we only discussed the effects of two manual segmentations of VOIs using T2WI. The effect of other routine sequences on diverse delineations, such as DWI and contrast-enhanced MRI, is still unclear and needs to be further investigated [35].

6. Conclusions

In this study, we developed two radiomics models based on different manual segmentations to assess the T stage in RC patients. The diverse delineation could cause certain differences in feature selection. Despite this discrepancy, both methods had good diagnostic performance in the preoperative T staging of RC. The minimum delineation had better stability in feature selection, while the maximum delineation was more beneficial in clinical decision-making.

Abbreviations

RC: Rectal cancer
MRI: Magnetic resonance imaging
AUCs: Areas under the curves
TME: Total mesorectal excision
ICC: Intraclass correlation coefficient
ROC: Receiver operator characteristic
DCA: Decision curve analysis
T2WI: T2-weighted imaging
ROI: Region of interest
VOI: Volume of interest
SVM: Support vector machine
TR: Repetition time
TE: Time of echo
DWI: Diffusion weighted imaging.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Additional Points

STARD Checklist. The checklist according to STARD 2015 guidelines for reporting diagnostic accuracy studies.

Ethical Approval

The present trial was approved by the local institutional review board (Committee on Ethics of Biomedicine, Changhai Hospital).

Consent

Written informed consent to publish the presented information was obtained from study participants.

Disclosure

The role of the funder is as follows: to develop the main idea and design the study.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

JL and FS conceived of the present idea. XM and ZZ acquired the data. HL analyzed and interpreted the patient data regarding the radiomics features. YX performed the statistical radiomics analysis. HL and YY were major contributors in writing the manuscript. All authors read and approved the final manuscript. Haidi Lu, Yuan Yuan, and Zhen Zhou contributed equally to this work.

Acknowledgments

The study was supported by the Youth Initiative Fund of the Naval Medical University (2018QN05).

Supplementary Materials

Supplemental Table 1: details on parameters applied for high-resolution T2WI, which were used for radiomics models. Supplemental Figure 1: the diagram of feature extraction. A and B: we used the variance threshold method to select 441 features (A: minimum delineation) and 444 features (B: maximum delineation) from 1409 features, respectively. C and D: We used select K-best methods to further select radiomics features. Finally, 4 optimal features (C: minimum delineation) and 7 optimal features (D: maximum delineation) were selected. *(Supplementary Materials)*

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] B. Oronsky, T. Reid, C. Larson, and S. J. Knox, "Locally advanced rectal cancer: the past, present, and future," *Seminars in Oncology*, vol. 47, no. 1, pp. 85–92, 2020.

[3] A. B. Benson, A. P. Venook, M. M. Al-Hawary et al., "Rectal cancer, version 2.2018, NCCN clinical practice guidelines in oncology," *Journal of the National Comprehensive Cancer Network*, vol. 16, no. 7, pp. 874–901, 2018.

[4] N. Horvat, "MRI of rectal cancer: tumor staging, imaging techniques, and management," *Radiographics*, vol. 39, no. 2, pp. 367–387, 2019.

[5] S. Nougaret, C. Reinhold, H. W. Mikhael, P. Rouanet, F. Bibeau, and G. Brown, "The use of MR imaging in treatment planning

for patients with rectal carcinoma: have you checked the "DISTANCE"?," *Radiology*, vol. 268, no. 2, pp. 330–344, 2013.

[6] P. Lambin, E. Rios-Velazquez, R. Leijenaar et al., "Radiomics: extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012.

[7] X. Ma, F. Shen, Y. Jia, Y. Xia, Q. Li, and J. Lu, "MRI-based radiomics of rectal cancer: preoperative assessment of the pathological features," *BMC Medical Imaging*, vol. 19, no. 1, p. 86, 2019.

[8] M. Liu, X. Ma, F. Shen, Y. Xia, Y. Jia, and J. Lu, "MRI-based radiomics nomogram to predict synchronous liver metastasis in primary rectal cancer patients," *Cancer Medicine*, vol. 9, no. 14, pp. 5155–5163, 2020.

[9] J. E. Bibault, L. Xing, P. Giraud et al., "Radiomics: a primer for the radiation oncologist," *Cancer Radiothérapie*, vol. 24, no. 5, pp. 403–410, 2020.

[10] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[11] P. Hu, J. Wang, H. Zhong et al., "Reproducibility with repeat CT in radiomics study for rectal cancer," *Oncotarget*, vol. 7, no. 44, pp. 71440–71446, 2016.

[12] X. Zhang, L. Zhong, B. Zhang et al., "The effects of volume of interest delineation on MRI-based radiomics analysis: evaluation with two disease groups," *Cancer Imaging*, vol. 19, no. 1, p. 89, 2019.

[13] J. T. Antunes, A. Ofshteyn, K. Bera et al., "Radiomic features of primary rectal cancers on baseline T2-weighted MRI are associated with pathologic complete response to neoadjuvant chemoradiation: a multisite study," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 5, pp. 1531–1541, 2020.

[14] J. Wang, X. Liu, B. Hu, Y. Gao, J. Chen, and J. Li, "Development and validation of an MRI-based radiomic nomogram to distinguish between good and poor responders in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiotherapy," *Abdominal Radiology*, vol. 45, no. 5, 2021.

[15] K. Nie, L. Shi, Q. Chen et al., "Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI," *Clinical Cancer Research*, vol. 22, no. 21, pp. 5256–5264, 2016.

[16] W. Zhang, Z. Huang, J. Zhao et al., "Development and validation of magnetic resonance imaging-based radiomics models for preoperative prediction of microsatellite instability in rectal cancer," *Ann Transl Med.*, vol. 9, no. 2, p. 134, 2021.

[17] F. Chen, X. Ma, S. Li et al., "MRI-based radiomics of rectal cancer: assessment of the local recurrence at the site of anastomosis," *Academic Radiology*, vol. S1076-6332, no. 20, article 30567-5, 2020.

[18] Z. Li, X. Ma, F. Shen, H. Lu, Y. Xia, and J. Lu, "Evaluating treatment response to neoadjuvant chemoradiotherapy in rectal cancer using various MRI-based radiomics models," *BMC Medical Imaging*, vol. 21, no. 1, p. 30, 2021.

[19] M. B. Amin, F. L. Greene, S. B. Edge et al., "The eighth edition AJCC Cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging," *CA: a Cancer Journal for Clinicians*, vol. 67, no. 2, pp. 93–99, 2017.

[20] E. Al-Sukhni, L. Milot, M. Fruitman et al., "Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: a systematic review and meta-analysis," *Annals of Surgical Oncology*, vol. 19, no. 7, pp. 2212–2223, 2012.

[21] G. Zhang, Y. Z. Cai, and G. H. Xu, "Diagnostic accuracy of MRI for assessment of T category and circumferential resection margin involvement in patients with rectal cancer: a meta-analysis," *Diseases of the Colon and Rectum*, vol. 59, no. 8, pp. 789–799, 2016.

[22] S. Rizzo, F. Botta, S. Raimondi et al., "Radiomics: the facts and the challenges of image analysis," *Eur Radiol Exp.*, vol. 2, no. 1, p. 36, 2018.

[23] B. Kocak, E. Ates, E. S. Durmaz, M. B. Ulusan, and O. Kilickesmez, "Influence of segmentation margin on machine learning–based high-dimensional quantitative CT texture analysis: a reproducibility study on renal clear cell carcinomas," *European Radiology*, vol. 29, no. 9, pp. 4765–4775, 2019.

[24] J. Rexilius, H. K. Hahn, M. Schlüter, H. Bourquain, and H. O. Peitgen, "Evaluation of accuracy in MS lesion volumetry using realistic lesion phantoms," *Academic Radiology*, vol. 12, no. 1, pp. 17–24.e1, 2005.

[25] P. Prasanna, J. Patel, S. Partovi, A. Madabhushi, and P. Tiwari, "Radiomic features from the peritumoral brain parenchyma on treatment-naïve multiparametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings," *European Radiology*, vol. 27, no. 10, pp. 4188–4197, 2017.

[26] N. M. Braman, M. Etesami, P. Prasanna et al., "Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI," *Breast Cancer Research*, vol. 19, no. 1, p. 57, 2017.

[27] M. L. Welch, C. McIntosh, B. Haibe-Kains et al., "Vulnerabilities of radiomic signature development: the need for safeguards," *Radiotherapy and Oncology*, vol. 130, pp. 2–9, 2019.

[28] M. Hatt, "Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 11, pp. 1662–1671, 2013.

[29] W. Lv, Q. Yuan, Q. Wang et al., "Robustness versus disease differentiation when varying parameter settings in radiomics features: application to nasopharyngeal PET/CT," *European Radiology*, vol. 28, no. 8, pp. 3245–3254, 2018.

[30] S. S. Yip and H. J. Aerts, "Applications and limitations of radiomics," *Physics in Medicine and Biology*, vol. 61, no. 13, pp. R150–R166, 2016.

[31] S. Trebeschi, J. J. M. van Griethuysen, D. M. J. Lambregts et al., "Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR," *Scientific Reports*, vol. 7, no. 1, p. 5301, 2017.

[32] C. Cai, T. Hu, J. Gong et al., "Multiparametric MRI-based radiomics signature for preoperative estimation of tumor-stroma ratio in rectal cancer," *European Radiology*, vol. 31, no. 5, 2021.

[33] J. Jian, F. Xiong, W. Xia et al., "Fully convolutional networks (FCNs)-based segmentation method for colorectal tumors on T2-weighted magnetic resonance images," *Australasian Physical & Engineering Sciences in Medicine*, vol. 41, no. 2, pp. 393–401, 2018.

[34] B. Irving, A. Cifor, B. W. Papiez et al., "Automated colorectal tumour segmentation in DCE-MRI using supervoxel neighbourhood contrast characteristics," *Med Image Comput Comput Assist Interv.*, vol. 17, Part 1, pp. 609–616, 2014.

[35] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Reviews. Cancer*, vol. 18, no. 8, pp. 500–510, 2018.

Hindawi
BioMed Research International
Volume 2021, Article ID 6671417, 19 pages
https://doi.org/10.1155/2021/6671417

Review Article

A State-of-the-Art Review for Gastric Histopathology Image Analysis Approaches and Future Development

Shiliang Ai,[1] **Chen Li**,[1] **Xiaoyan Li**,[2] **Tao Jiang**,[3] **Marcin Grzegorzek**,[4] **Changhao Sun**,[1,4,5] **Md Mamunur Rahaman**,[1] **Jinghua Zhang**,[1,4] **Yudong Yao**,[6] **and Hong Li**[1]

[1]*Microscopic Image and Medical Image Analysis Group, College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110169, China*
[2]*Cancer Hospital of China Medical University, Liaoning Cancer Hospital and Institute, Shenyang 110042, China*
[3]*Control Engineering College, Chengdu University of Information Technology, Chengdu 610103, China*
[4]*Institute of Medical Informatics, University of Luebeck, Luebeck, Germany*
[5]*Shenyang Institute of Automation, Chinese Academy of Sciences, 110169, China*
[6]*Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA*

Correspondence should be addressed to Chen Li; lichen201096@hotmail.com and Xiaoyan Li; lixiaoyan@cancerhosp-ln-cmu.com

Received 23 December 2020; Revised 9 May 2021; Accepted 25 May 2021; Published 28 June 2021

Academic Editor: Yong Xia

Gastric cancer is a common and deadly cancer in the world. The gold standard for the detection of gastric cancer is the histological examination by pathologists, where Gastric Histopathological Image Analysis (GHIA) contributes significant diagnostic information. The histopathological images of gastric cancer contain sufficient characterization information, which plays a crucial role in the diagnosis and treatment of gastric cancer. In order to improve the accuracy and objectivity of GHIA, Computer-Aided Diagnosis (CAD) has been widely used in histological image analysis of gastric cancer. In this review, the CAD technique on pathological images of gastric cancer is summarized. Firstly, the paper summarizes the image preprocessing methods, then introduces the methods of feature extraction, and then generalizes the existing segmentation and classification techniques. Finally, these techniques are systematically introduced and analyzed for the convenience of future researchers.

1. Introduction

1.1. Background. Cancer is a disease in which human cells grow out of control. In 2018, about 9.6 million people died of cancer [1]. Gastric cancer is a kind of cancer that is caused by an abnormal cell population that proliferates endlessly in the stomach and eventually forms tumors. According to the morphological characteristics of gastric tumors, gastric cancer can be classified into the following categories: adenocarcinoma, mucinous carcinoma, signet ring cell carcinoma, adenosquamous carcinoma, squamous cell carcinoma, and undifferentiated cell carcinoma. Adenocarcinoma accounts for more than 95% of all gastric malignancies. In general, gastric cancer refers to gastric adenocarcinoma. Adenocarcinoma includes tubular adenocarcinoma and papillary adenocarcinoma. Tubular adenocarcinoma has a well-defined glandular lumen.

Gastric cancer ranks second in morbidity and mortality among all cancers [2]. Gastric cancer kills about 800,000 people a year, according to the World Health Organization (WHO). China and Japan have the highest incidence of stomach cancer, accounting for 30 percent of all cancers, and the number of cancer cases each year is also increasing in the United States. In gender analysis, gastric cancer is the second

most common cancer among men, at 26 percent, and the third most common among women, at 11 percent. It is estimated that about 30,000 new cases of gastric cancer occur each year.

The diagnosis of gastric cancer is mainly through pathological biopsy, which is stained with hematoxylin and eosin. The biopsy morphology and tissue characteristics under the microscope are observed, and the detection results are determined by a synthesis of the doctor's knowledge. But each pathology doctor's diagnosis is made according to their different experiences and states, which may result in making different judgments on gastric cancer tissue pathology images. At the same time, the pathologist has to examine a large number of histopathological images every day [3]. The diagnosis process requires long periods of concentration, and long hours of work can lead doctors to misdiagnose situations. Therefore, accurate screening and diagnosis of gastric cancer by pathologists is a major problem [4]. The number of pathologists is also very scarce. In order to alleviate the shortage of pathologists and reduce the misdiagnosis rate of histopathological examination, the CAD system is introduced into the detection of pathological images of gastric cancer [5]. With the assistance of a CAD system, the area of the malignant tumor is marked, and the accurate judgment of the computer is used as a second opinion to assist the pathologist in making a judgment [6].

The CAD system began in the 1980s. Its main purpose is to assist pathologists to make judgments by using the accuracy and efficiency of computers. The CAD system can make judgments objectively, and excellent algorithms can reduce the processing time [7]. In the past few decades, the continuous progress of machine learning algorithms has enabled the rapid development of CAD technology in gastric cancer, which can more quickly and accurately identify cancer regions [8] [9].

In the CAD system of machine learning in the field of gastric cancer, there are mainly two kinds of segmentation and classification. The main step of the classification algorithm is to preprocess the image first, in order to improve the quality of the image and make the data meet the experimental requirements. Then, feature extraction is carried out to find the features of the image that are of interest to the experiment. Finally, a suitable classifier is designed to classify the features. The segmentation algorithm consists of two steps: image preprocessing and image segmentation [10]. In recent years, a new technology has emerged: deep learning technology. A deep learning algorithm can directly act on RGB images, automatically learn the features of images through convolutional neural network, search for similar features of experimental data through a lot of training, and finally achieve segmentation and classification [11].

1.2. Motivation. At present, some references summarize related works of histopathological image analysis approaches, but little is done for gastric cancer. Therefore, this work focuses on the technical analysis of gastric cancer histopathological images.

In 1979, the comprehensive survey "Computer-Aided Medical Diagnosis: Literature Review" was presented [12]. This paper summarizes all the medical diagnostic techniques involved in the development, testing, and application of CAD. In this survey, there is one work related to gastric cancer.

In 2004, the paper in [13] proposed a survey paper about "Artificial Intelligence in Medicine." In this paper, artificial intelligence is introduced that can analyze complex medical data and can diagnose, treat, and predict outcomes in the field of medicine. This paper is related to seven histopathologies, one of which is gastric histopathology.

In 2005, the paper in [14] presents a survey about "automated cancer diagnosis based on histopathological images." In this paper, from the three aspects of image preprocessing, feature extraction, and image classification, a total of 75 papers are summarized. A total of 11 cancers are covered, but only one is related to stomach cancer.

In 2007, a survey on "Neural Networks and Other Machine Learning Methods in Cancer Research" was completed [15]. This paper summarizes some machine learning methods mainly used in the field of cancer and discusses their respective advantages and disadvantages in the field of medicine. In this work, ten types of cancer tissues are mentioned, but only one paper is about gastric cancer.

In 2018, a survey on "Deep Learning and Medical Diagnosis: A Review of Literature" was carried out. This paper provides a comprehensive analysis of deep learning techniques in histopathological images [16]. There are 46 articles on deep learning in this work, but there is only one paper that focuses on GHIA. Besides the related work mentioned above, there are some other surveys for histopathology analysis [17, 18, 19, 20], but they do not involve gastric cancer.

From the works mentioned above, we can find that there are many review papers on histopathological image analysis, which summarize various cancers in the medical field. However, none of them specially focus on histopathological images in the direction of gastric cancer.In this paper, we summarize a state-of-the-art review for gastric histopathology image analysis approaches. Figure 1 shows the histopathological image literature of gastric cancer collected by Google Scholar according to the keywords of histopathological analysis of gastric cancer. A total of 364 relevant papers were downloaded, and 234 papers that were not about gastric cancer were deleted through simple reading, and then 106 papers that were not about histopathological images in gastric cancer were also deleted; finally, 24 papers were selected. As shown in Figure 2, this paper classifies and summarizes the literature according to four aspects: preprocessing, feature extraction, segmentation, and classification. As can be seen from Figure 2, papers related to GHIA were published mainly since 2012, and the number of relevant papers gradually increased after that. There are only 4 and 6 papers on image preprocessing and image segmentation, while the feature extraction method has 8 papers. There are 14 papers with the largest number of classification methods.

Figure 3 is the structure diagram of this paper, summarizing the histopathology of gastric cancer from four aspects. Figure 3(a) shows the preprocessing aspect. Data enhancement is used in the preprocessing step, augmenting the dataset and preparing for the following experiments. Figure 3(b) shows the feature extraction aspect. The histograms of oriented gradient (HOG) feature and the gray-level cooccurrence matrix (GLCM) feature, along with the other features,

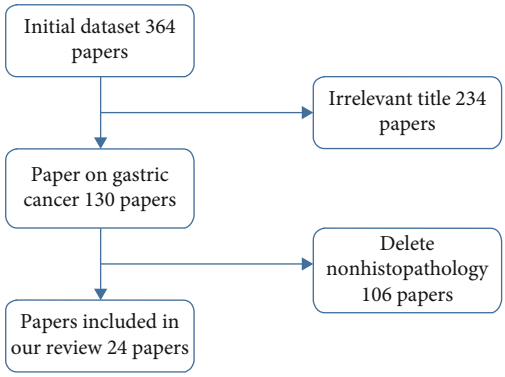


Figure 1: The systematic flowchart of paper selection for our work.

are used for extracting the color, texture, and other characteristics of the nucleus. Linear discriminant analysis (LDA) and local linear embedding (LLE) are the processing feature vectors. Figure 3(c) shows the segmentation aspect. The minimum-model method and the deep learning network are devoted to find the nuclear region. Figure 3(d) shows the classification aspect. The artificial neural network (ANN) and the convolutional neural network (CNN) are applied to screen the cells needed for the experiment from histopathological images.

2. Image Preprocessing Methods

2.1. Related Work. Image preprocessing [21] is an important part of GHIA. In experiments, the quantity of data may be small and the image quality may be poor. Therefore, in order for experiments to be more accurately classified or segmented, a mass of datasets are needed. The experimental data need to be preprocessed so that the data can meet the experimental requirements and get better results.

In [22], a method is proposed for presenting an automatic lymphocyte detection model based on the Deep Convolutional Neural Network (DCNN) in immunohistochemical images of gastric cancer. This work extracts data from a 40-fold magnification scan of full-sized micrographs of gastric cancer tissues. The experimental data consist of 3,257 images. However, these datasets are far from enough. In the study, the data expansion methods of rotation and reflection are used to increase the number of datasets, resulting in a total of 10,868 datasets.

The work in [23] comes up with an efficient learning algorithm to replace the traditional feature extraction method. This algorithm requires a large dataset for machine learning. Some standard data augmentation methods are used to generate a large number of images, and various data enhancement methods are used to process the data in this work, including 0.3x overlap, reflection, postreflection rotation, and shearing. Through these methods, 21,000 images can be produced per pathological section, and a total of 11 sections of experimental data are generated for a total of 231,000 datasets.

The work in [24] proposes a framework for automatic recognition of gastric cancer based on deep learning. In this work, the resolution of the original gastric image dataset is 2048×2048, and the deep learning network cannot directly process the gastric image, so a patch of 224×224 is intercepted from the original gastric image. The gastric dataset contains 560 gastric cancer sections and 140 normal sections. In order to increase the dataset, the patch is rotated 90°, 180°, and 270°. After rotation, the data are cut, and the data becomes 8,992 pieces of gastric cancer data and 14,000 pieces of normal data.

In [25], an image classification model is proposed that can alleviate the bad annotation training set. By fine-tuning the neural network in two stages and introducing a new intermediate dataset, the performance of the network in image classification with a poor annotation training set is improved.

2.2. Summary. It can be concluded from the above works that machine learning requires a huge number of datasets. Augmentation approaches are the main solution. Augmentation approaches mentioned in the above work include rotation, reverse, and scale transformation. In addition, there are some other data augmentation approaches, such as adding noise and color vibrance. Figure 4(a) shows the original image. Figure 4(b) shows rotation transformation, with a clockwise rotation of 90°, 180°, and 270° from the original image. Figure 4(c) shows color vibrance, with brightness enhanced at 10% or brightness reduced to 10% and 20%. In Figure 4(d), noise is added: Gaussian noise, salt and pepper noise, and Poisson noise; Figure 4(e) shows reverse transformation: horizontal rotation and vertical rotation.

3. Feature Extraction Methods

3.1. Related Work. Feature extraction [26] refers to finding representative data in the region of interest in the image. Feature extraction mainly includes three parts: the color feature, the texture feature, and the shape feature. Some of them require postprocessing of features, such as feature dimensionality reduction.

3.2. Color Feature. Features based on color intensity are very important in pathology. Due to the use of special staining, the cytoplasm, nucleus, cell wall, and other stains are different in the pathological image, which can be manifested by color features. The work in [23, 27] extracts hue, saturation, and value (HSV) histograms; gray histograms; and red, green, and blue (RGB) histograms as color features to describe the difference between the colors.

3.3. Texture Feature. Texture feature is also a commonly used histopathological image detection method. In the histopathological images of gastric cancer, the texture features can be extracted from the nucleus and cell wall. Texture features mainly include the GLCM feature, the HOG feature, and the local binary pattern (LBP) feature.

In [23], GLCM, LBP, and Gabor filter bank features are extracted, respectively, as texture feature extraction in the comparison experiment.

In [28], the HOG feature is extracted from gastric cancer histopathological images. The HOG feature is drawn on normal, benign, and malignant gastric images to obtain the HOG feature vector. The histogram of gastric cancer is drawn by HOG feature histograms. Then, the data in the HOG

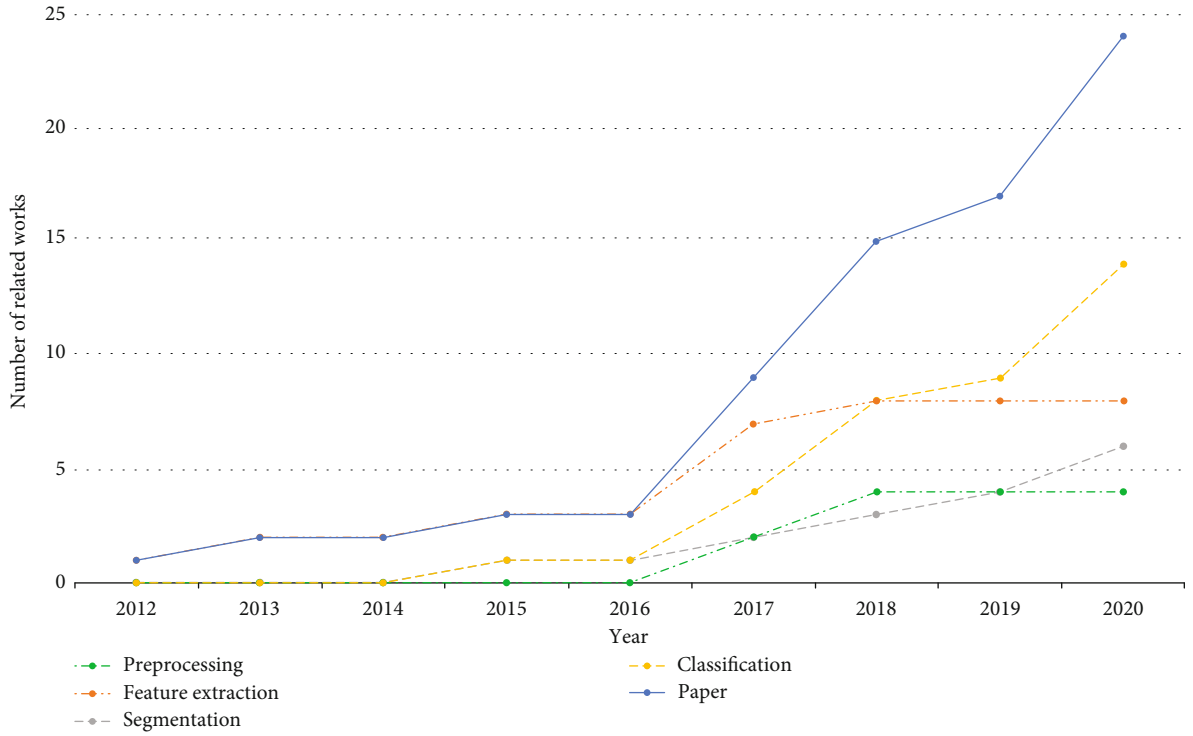


Figure 2: The related work presents line charts of time and quantity.

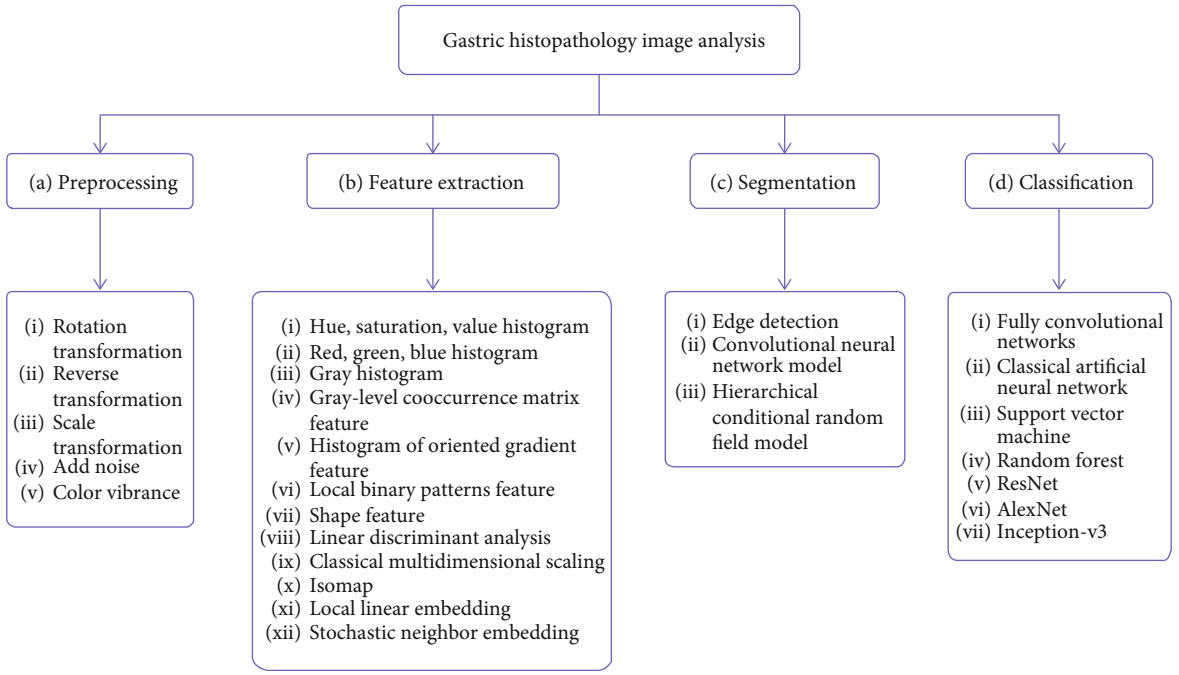


Figure 3: A framework for gastric histopathology image analysis.

histogram is directly used for classification. The accuracy rate of this work classification is 100%.

The work in [29] proposes a HOG-LDA-ANN method for gastric histopathological image classification. HOG features are compared with the GLCM and LBP features. HOG-LDA-ANN has an accuracy rate of 88.9%. There are two contrast experiments: GLCM-LPP-ANN has an accuracy rate of 85.56% and LBP-LPP-ANN has an accuracy rate of 80.12%.

The work in [30] extracts the HOG and LBP features. This work tests many feature dimension reduction methods and classifiers. Through comparison, the LBP feature is superior to the HOG feature in gastric cancer histopathological images.

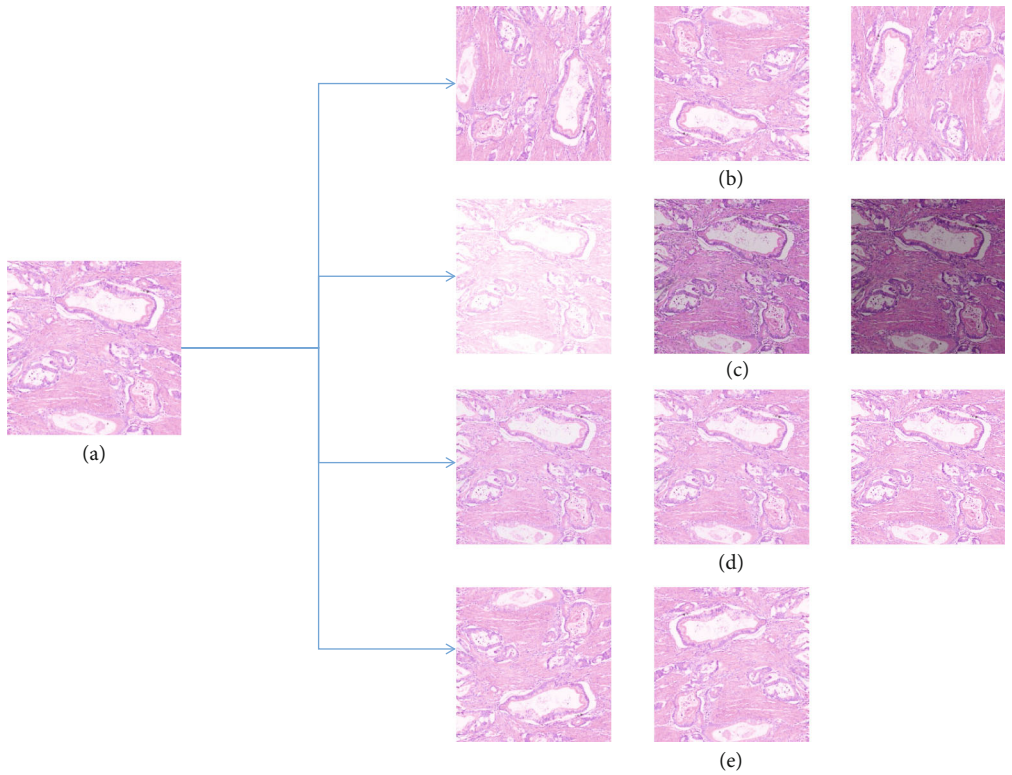


Figure 4: Some common augmentation approaches in the preprocessing stage.

3.4. Shape Feature. A shape feature is used to extract the topological structure information from the image as a feature. For example, the distance of some points in the image is extracted as features and then classified.

The work in [31] obtains the location information of the nucleus and then extracts features. First, the alignment and mean distance of the three adjacent cells are measured, and then the mean value and standard deviation of these two indicators are calculated as the first feature set. Another feature set is extracted from a circle of 40 microns around the central cell. The ratio of the number of nuclei in the circle to the empty circle section is calculated as the feature.

The work in [32] extracts four types of nuclear location information: epithelial cells, leukocytes, fibrocytes, and conglomerates, and obtains the cell nuclei attributed relational graph. Then, the mean, variance, skewness, and kurtosis features can be obtained from the vertices and edges of the nucleus.

3.5. Postprocessing Methods of Features. In [29], the HOG-LDA-ANN method for gastric cancer images is proposed. LDA is a dimension reduction technique for supervised learning, which can retain and screen effective features and eliminate inefficient features. In this work, 46,900 HOG feature vectors are obtained through feature extraction, and 90 groups of vectors are obtained through feature dimensionality reduction by the LDA algorithm.

The work in [30] compares a variety of feature reduction methods: Sammon mapping, stochastic neighbor embedding, Laplacian mapping, Isomap, classical multidimensional scaling, local linear embedding, linear discriminant analysis, and t-distribution random neighbor embedding. According to the final classification accuracy, two better classification methods are obtained: the LBP-MDS-ANN and LBP-LLE-ANN methods.

The work in [33] improves the PCA+LDA algorithm. Based on the traditional PCA+LDA, the LDA transformation is optimized. This method improves the generalization of traditional PCA+LDA to test samples and improves the classification accuracy. The optimized algorithm improves the classification accuracy by 3.43%.

3.6. Summary. This paper summarizes the feature extraction method in Table 1. In conclusion, among the three features, color features have strong limitations. It is only useful for gastric cancer histopathological images with good staining effect and high contrast. Texture features are widely used, and many articles choose the texture feature method. Shape features are used the least, and most of the extraction methods are to extract the nucleus and extract the features from the position relationship between the nuclei.

For the feature postprocessing of gastric cancer histopathology, dimensionality reduction is the main method, and features are processed by PCA, LDA, and other different algorithms. In general, LDA has a good effect on the feature dimension reduction of traditional methods. In the analysis of histopathological images of other cancers, we have also consulted relevant postprocessing methods. Although there are few relevant literatures, the main methods are PCA and LDA, such as in cervical cancer [34, 35], breast cancer [36, 37], and colorectal cancer [38].

Table 1: Summary of feature extraction.

Aim	Year	Reference	Team	Method
Feature extraction	2017	[23]	Sharma et al.	GLCM, LBP, HSV, and RGB
	2015	[27]	Sharma et al.	RGB feature, shape feature, and texture feature
	2017	[28]	Korkmaz et al.	HOG feature
	2017	[29]	Korkmaz et al.	HOG feature
	2018	[30]	Korkmaz et al.	LBP feature, HOG feature
	2013	[31]	Cosatto et al.	Feature of nuclear location relationship
	2017	[32]	Sharma et al.	Feature of nuclear location relationship
Postprocessing	2017	[29]	Korkmaz et al.	LDA
	2018	[30]	Korkmaz et al.	SNE, Isomap, MDS, LLE, LDA, and T-SNE
	2012	[33]	Gan et al.	PCA and LDA

4. Segmentation Methods

4.1. Related Work. Image segmentation [39] divides the image into several specific and unique regions. Then, the part of interest in the experiment is extracted. Image segmentation is an important step in image analysis.

In [27], the authors propose a nuclear segmentation technique, which automatically separates the nucleus by using a minimum-model method. The minimum model consists of two main steps. The first is the minimum prior information, and the second is the contour detection method independent of the image shape. This method avoids the segmentation deviation of shape features and can be accurately segmented.

In [40], the authors propose a deep learning model for the segmentation of histopathological images of gastric cancer. The workflow of this approach is shown in Figure 5. It contains two convolutional layers, three pooling layers, three multiscale module, one feature pyramid module, and one upsampling convolutional module. The convolution layer uses a3×3convolution kernel and the maximum pooling size is 2×2. In order to extract and fuse features, multiscale modules are used for the shallow layer and feature pyramids are used for the deep layer layers. The segmentation performance of the framework is 90.88%.

In [41], the authors present a gastric cancer segmentation method based on deformable convolution and multiscale embedding networks. The workflow of this approach is shown in Figure 6. This work combines atrous convolution, deformation convolution, and atrous space pyramid pool module. Then, the features of different semantic levels are extracted in the subsampling and feature fusion by using the lightweight decoder. Finally, intensive upsampling is performed. In this work, the dataset includes 500 pathological images of gastric cancer with an image size of2048×2048, and a 91.60% pixel-level accuracy and 82.65% mean intersection are achieved. This method compares with previous methods, such as FCN, VGG, U-net, and DeepLab-V3. The segmentation effect is the highest, with an accuracy rate of 91.6%.

In [42], a partial marker of the gastric tumor segmentation method based on reiterative learning is proposed. The architecture of the model is shown in Figure 7. In the absence of manual labeling, the average intersection of union coefficients obtained by training weakly labeled datasets is 0.883, and the average accuracy is 91.09%. After that, the deviation between patches is eliminated through overlapped region forecast.

In [43, 44], the authors propose a method for the segmentation of histopathological images for gastric cancer based on the hierarchical conditional random field (HCRF). The workflow of this approach is shown in Figure 8. This method can automatically locate and mark cancer regions in gastric cancer histopathological images. In this work, a total of 560 H&E-stained pathological images of gastric cancer are collected in the dataset. The segmentation accuracy is 78.91%, the recall is 65.59%, and the specificity is 81.33% through this method.

4.2. Summary. In summary, there are two main segmentation methods in gastric cancer histopathological image segmentation: the traditional machine learning segmentation method and the deep learning segmentation method. The traditional method uses edge detection for segmentation, and the deep learning method uses the FCN model for segmentation. U-net has a good effect on the segmentation of pathological images. However, there has been no work on the histopathological segmentation of gastric cancer using U-net. The advantage of depth segmentation lies in its good segmentation effect. It is a general frame structure and can adapt to various features. But the disadvantage is high time complexity and space complexity. Machine learning segmentation has the advantages of wide coverage and strong adaptability, while the disadvantages include large computation, complex model design, and high hardware requirements.

5. Classifier Design Methods

Classifier [45] design is a very important part. Choosing an appropriate classifier can make the experimental results better. In traditional machine learning, classifiers generally include SVM and RF. For deep learning networks, ResNet and U-net are commonly used networks.

5.1. Machine Learning Classifiers. In [27], a traditional machine learning method is presented for cell classification. This model uses a minimum-model method for multiresolution image segmentation and then extracts 7 intensity-based

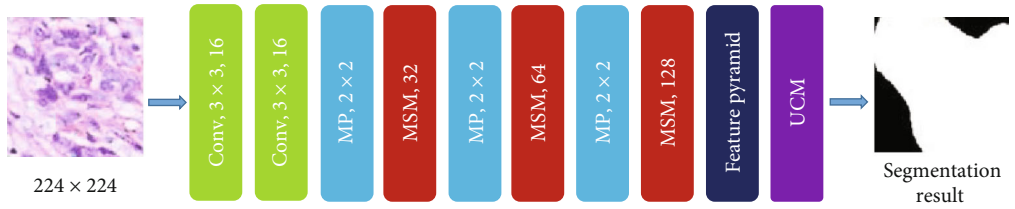


Figure 5: The framework of the proposed method in [40]. This figure corresponds to Figure 2 in the original paper.

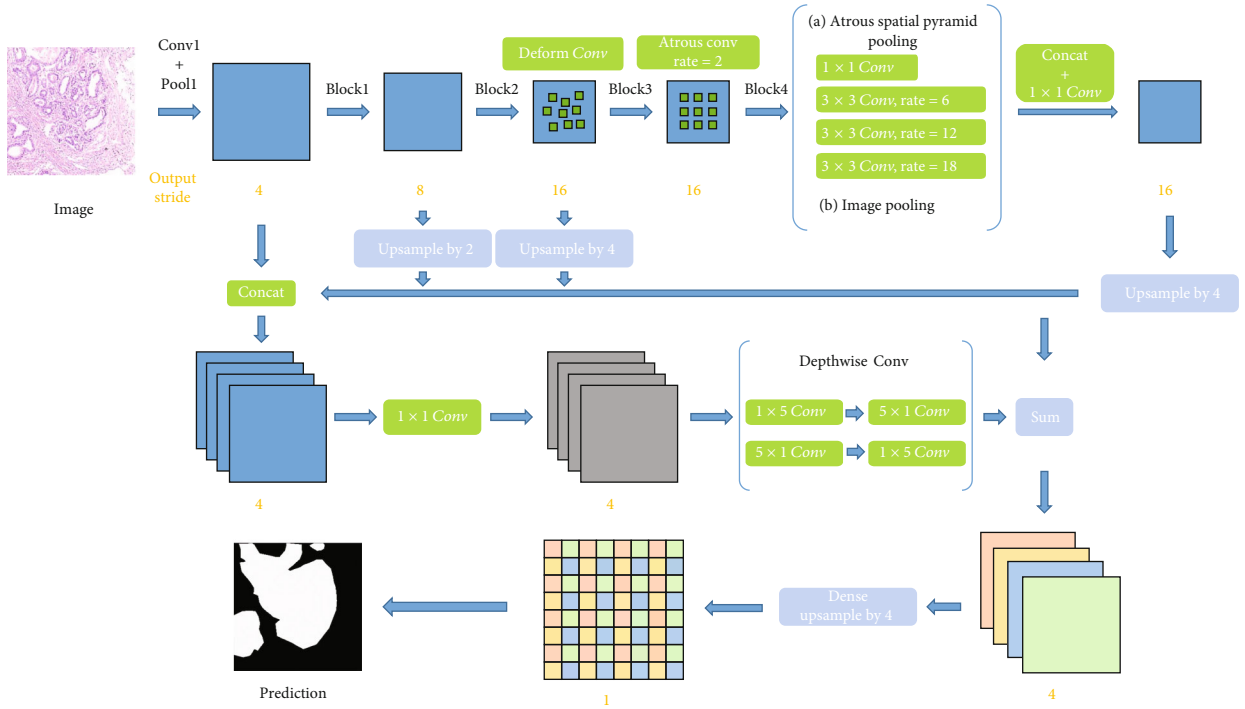


Figure 6: The framework of the proposed method in [41]. This figure corresponds to Figure 6 in the original paper.

features, 20 morphological features, and 4 texture features. Then, the cells are divided into eight categories by Adaboost. Finally, the multiresolution segmentation results are combined and evaluated.

In [30], the authors make a comparison from three perspectives in the gastric cancer histopathological image, such as feature extraction, feature dimensionality reduction, and classifier. In classifier, this work compares RF and ANN. The ANN classifier is superior to the RF classifier.

In [32], the authors propose a method to classify histopathological images of gastric cancer by nuclear attribute relation graphs. The image is preanalyzed, and the nuclei are segmented first [46], followed by selective nuclear classification. According to the classification, different types of nuclei are constructed into cell relationship maps, and each cell relationship map is extracted for features. A total of 332 feature vectors are extracted according to the characteristics of the map, including the mean, variance, skewness, and kurtosis. Finally, random forests are used for classification.

In [47], the authors propose three deep learning classification algorithms for gastric cancer histopathology. The workflow of this approach is shown in Figure 9. The first group of experiments is classified by the CNN method. The first layer is the input layer, and the input image size is 512 × 512 × 3. Each layer from the second to the fourth is convolution and pooling; the convolution kernel is 3 × 3, the step length is 1, and the maximum pooling size is 2 × 2. The fifth layer is a 64 × 1 × 1 convolution feature. The accuracy of the classification results is 86.4%. In the second group of experiments, the features are extracted by CNN and then classified by an RBF kernel support vector machine. The accuracy of the classification results is 89.2%. In the third group of experiments, K-SVD is used to learn the features extracted from CNN to obtain an overcomplete dictionary, and then, sparse decomposition is carried out. Using linear kernel SVM for classification, the classification accuracy is 95%.

5.2. Deep Learning Classifiers. In [22], a nine-layer DCNN is proposed, which is made up of three convolutional layers, three max-pooling layers, two fully connected layers, and one output layer. The workflow of this approach is shown in Figure 10. The 3 × 3 convolution kernel is convolved with

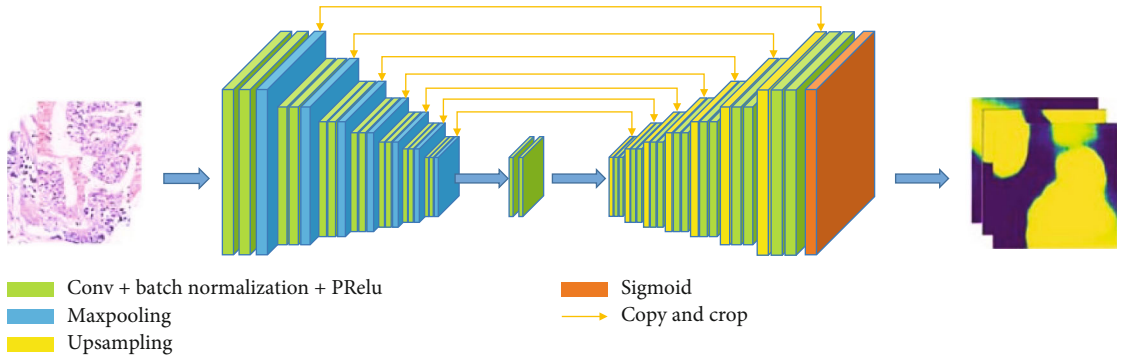


Figure 7: The framework of the proposed method in [42]. This figure corresponds to Figure 3 in the original paper.

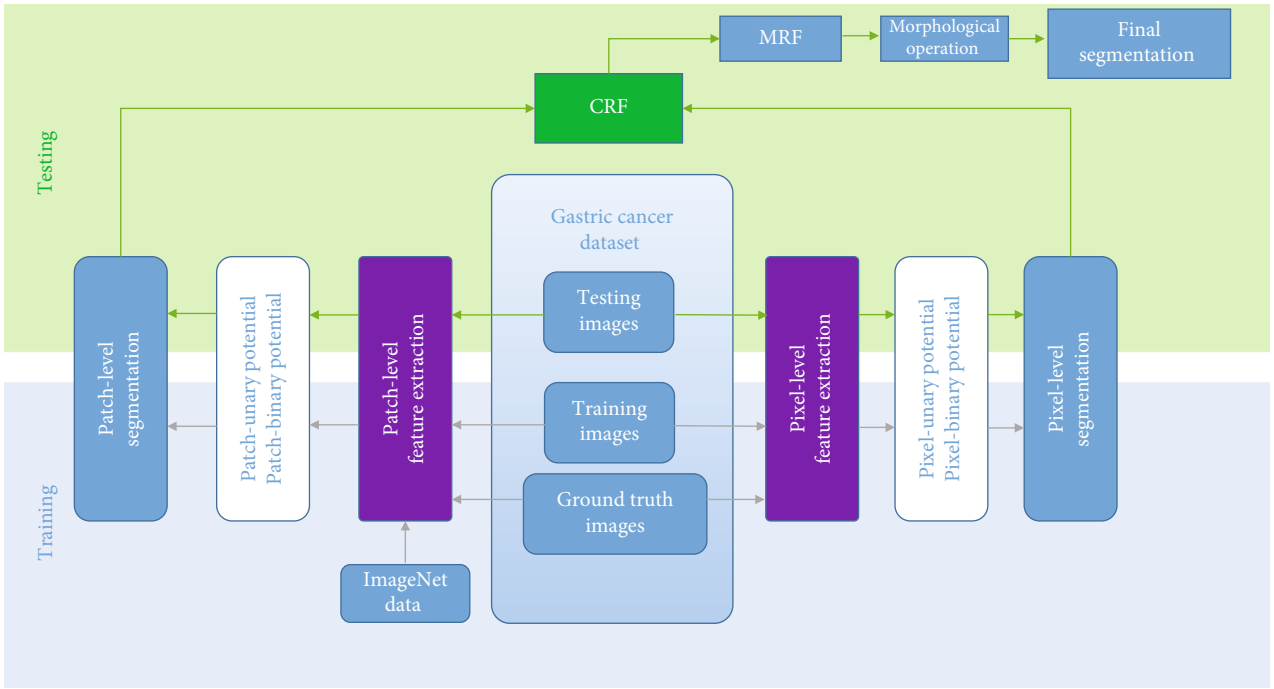


Figure 8: The framework of the proposed method in [43]. This figure corresponds to Figure 1 in the original paper.

a step length of 1. The convolution output is pooled to a maximum size of 2×2. The number of output features after three convolutional pooling is 64, 128, and 256. Finally, the pooling results are put into two full connection layers to obtain 2,048 feature vectors. This work produces an accuracy of 96.88%.

In [23], the authors design a pure supervised feedforward CNN model. As shown in Figure 11, the model consists of 9 layers, 3 convolution layers and pooling layers, and finally 3 fully connected layers. The size of the convolution kernel of the three convolutional layers is 7×7, 5×5, and 3×3. The max pooling is the pooling layer with the size of 2×2. Finally, the eigenvector enters the full connection layer. The accuracy of the network in classifying tumor and necrotic areas is 69.9% and 81.4%. In addition, several comparative experiments are also done in this work. AlexNet is used in deep learning, color and texture features are used in machine learning, and RF is used for classification.

In [24], the authors propose a new deep learning network-based classification model of gastric cancer histopathological images. In order to extract deep features, the deep learning network proposed in this work has different structures, namely, the shallow multiscale module and the deep network module. The workflow of this approach is shown in Figure 12. The green rectangle represents the convolutional layer, and it has a convolution kernel size of 3×3. The blue rectangle represents the maximum pooling layer, and the kernel size is 2×2. The orange rectangle represents the average pooling layer, and the kernel size is 7×7. The purple rectangle represents the fully connected layer. Several comparative experiments are performed, such as AlexNet, VGG-16, ResNet-50, ResNet-101, Inception-V4, and DenseNet-121. After comparison, the network of this work has achieved good results. For the patch level, the classification accuracy of the model is 97.93%. For the slice level, the classification accuracy of the model is 100%.

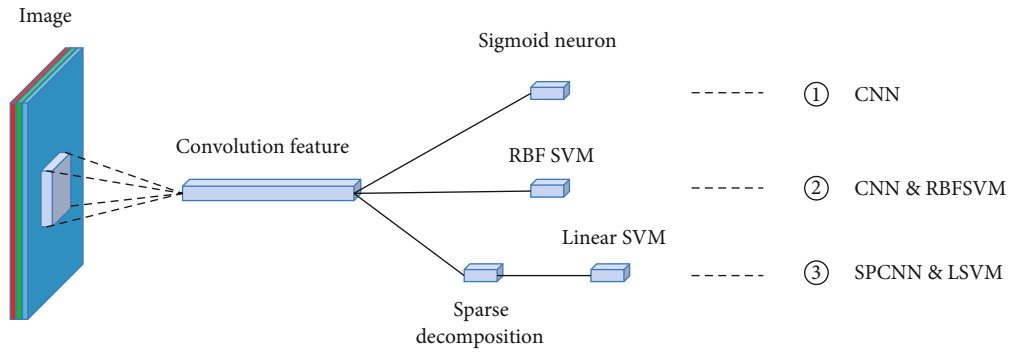


Figure 9: The framework of the proposed method in [47]. This figure corresponds to Figure 5 in the original paper.

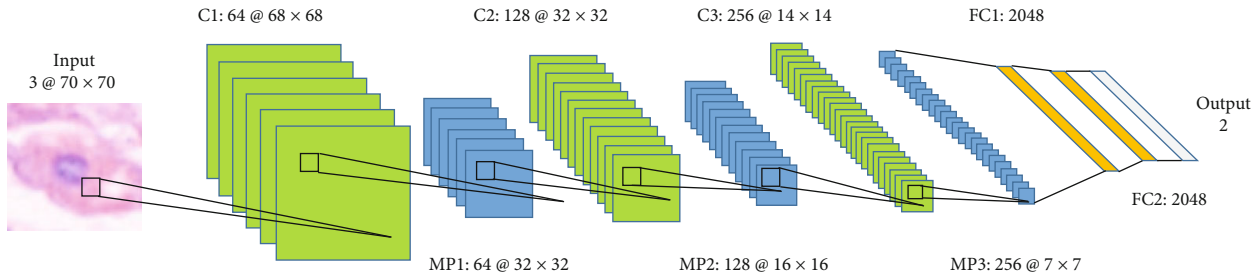


Figure 10: The framework of the proposed method in [22]. This figure corresponds to Figure 2 in the original paper.

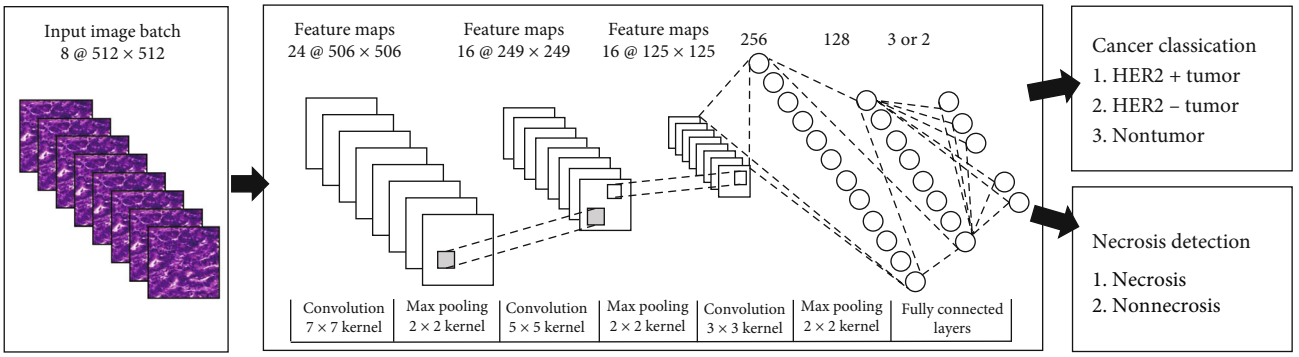


Figure 11: The framework of the proposed method in [23]. This figure corresponds to Figure 4 in the original paper.

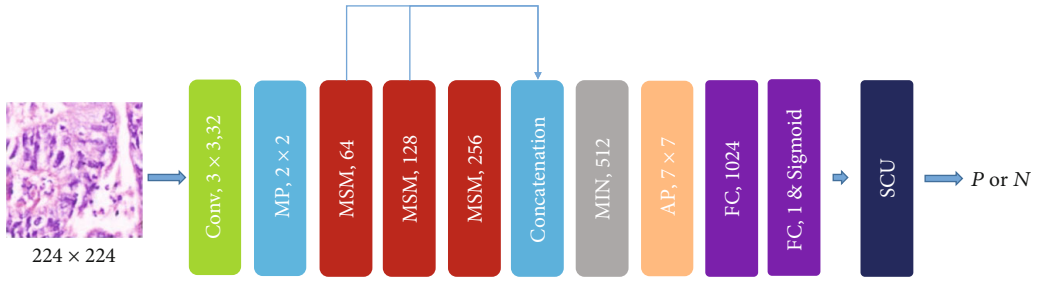


Figure 12: The framework of the proposed method in [24]. This figure corresponds to Figure 1 in the original paper.

In [48], the authors built a 50-layer residual network model. In this model, multisize convolution kernels are used to extract features, which are 7×7 and 3×3 convolution kernels with 2 steps. Then, ReLU or Sigmoid function is used to activate the features nonlinearly. After a lot of training, the network achieves an output F-score of 95.5%. On top of that, the model is optimized to increase the F-score to 96%.

In [49], a feature balanced module (FBM) is proposed that can distinguish slight differences in an image. The

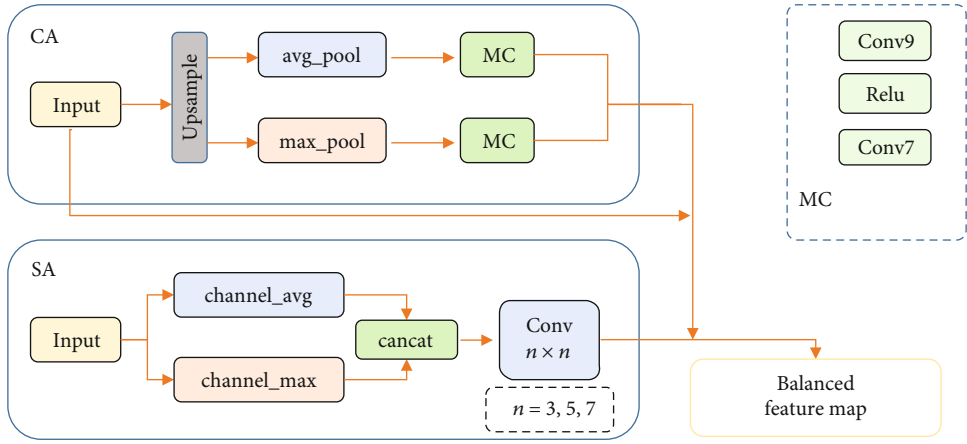


Figure 13: The framework of the proposed method in [49]. This figure corresponds to Figure 2 in the original paper.

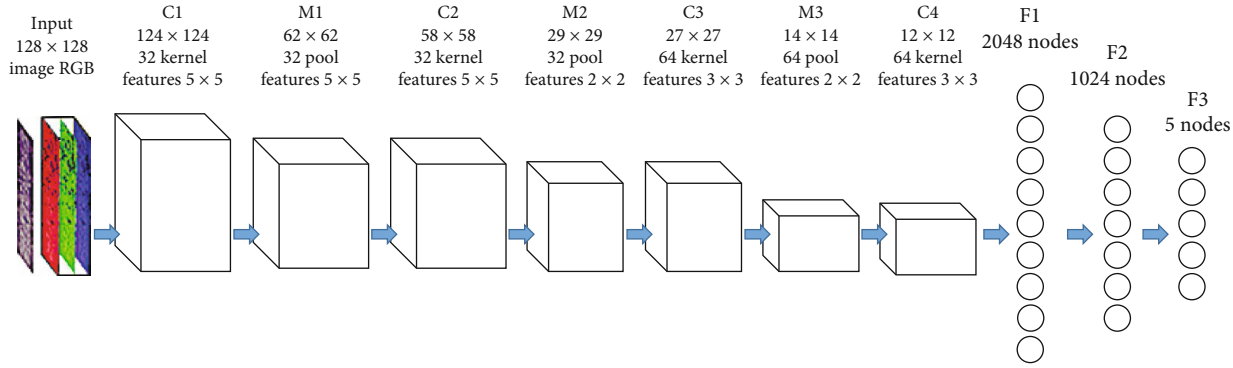


Figure 14: The framework of the proposed method in [50]. This figure corresponds to Figure 1 in the original paper.

flowchart of FBM is shown in Figure 13. The balance module has two types of channels: the first is channel concern (CA) modules, and the second is space concern (SA) modules. In CA, the input features are upsampled, and then, max pooling and average pooling are performed. Convolution uses a9 × 9 convolution kernel, ReLU function is used for activation, and convolution uses a7 × 7convolution kernel. Finally, add the two outputs as output features. In SA, the convolution kernels 3, 5, and 7 are used for convolution, and the features are compressed.

In [50], the authors propose a ten-layer convolutional neural network, in which three convolutional layers extract features, four pooling layers reduce image size, and three full connection layers output feature values. The workflow of this approach is shown in Figure 14.

In [51], the authors propose a gastric cancer histopathological image classification method based on recalibrated multi-instance deep learning. In this method, two convolution layers and one pooling layer are added to transform the ResNet-v2 network into a complete network model. The pooling layer is average pooling, and there are two convolution layers: one for feature extraction and the other for classification. This network is shown in Figure 15. The network is mainly composed of three modules, the first is local-global feature fusion, the second is the recalibration module, and the third is multi-instance pooling. The classification accuracy of the model is 86.5%.

In [52], the authors fuse the two networks of DeepLab-V3 and ResNet-50, introduce the structure of ResNet-50 network into DeepLab-V3, and build a new convolutional neural network based on DeepLab-V3. In this work, 2,166 whole slices are selected as the training set and 300 slices as the test set. After a lot of training, the final accuracy of the model is 87.3%, the sensitivity is 99.6%, and the specificity is 84.3%.

In [53], the authors propose a multiscale deep learning network, in which images of different magnification levels are selected from the whole WSI image, patches of the same size are extracted from images of different magnification rates, and then these patches are put into the deep HIPO. Then, the network can learn images at multiple scales. The network is shown in Figure 16.

In [54], the authors use a standard Inception-V3 network framework. By changing the depth multiplier, the parameters are reduced. In order to increase the robustness of the image, data enhancement methods are used such as mirror and rotation. Adam optimization algorithm is used to optimize the network. After a lot of training, the network model with the smallest verification error is selected.

In [55], the authors propose three classical convolutional neural networks for image classification: AlexNet, ResNet-50, and Inception-V3. In data selection, tenfold cross validation is used to test the performance of classification. The data is divided into ten parts, and train : validation : test = 8 : 1 : 1. For each combination, the classification results of the three

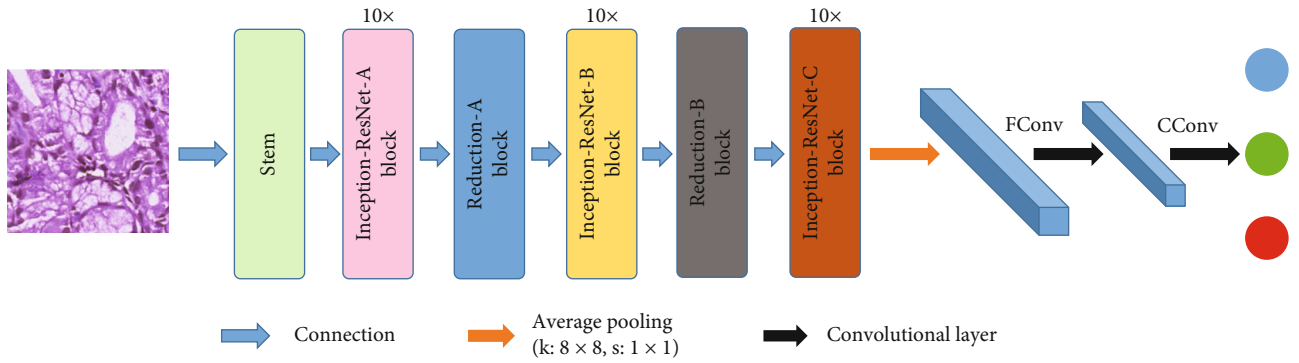


Figure 15: The framework of the proposed method in [51]. This figure corresponds to Figure 3 in the original paper.

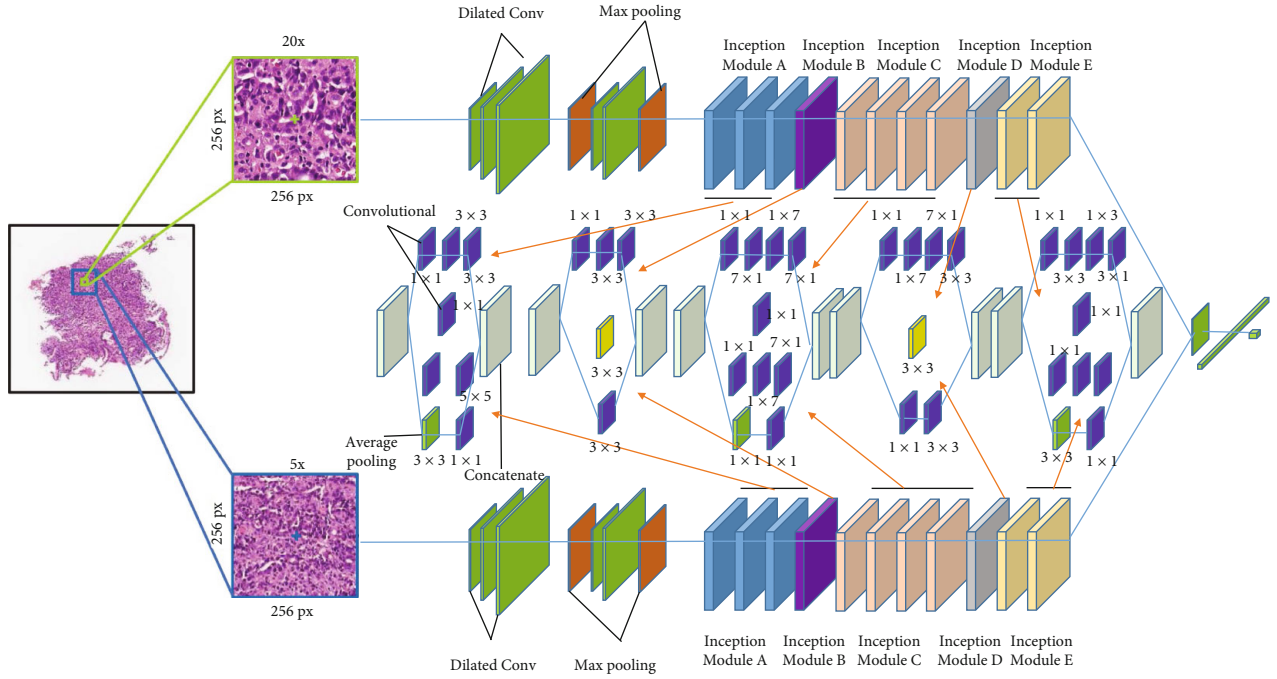


Figure 16: The framework of the proposed method in [53]. This figure corresponds to Figure 3 in the original paper.

classic networks are obtained. Finally, the ten results are the output in a series to calculate the accuracy, sensitivity, and specificity. The network is shown in Figure 17.

5.3. Summary. In the classification design of gastric cancer histopathology images, there are few articles using the traditional machine learning classification method, among which the techniques used include SVM, RF, ANN, and Adaboost. There are many classified articles using deep learning direction. Convolutional neural networks mainly include ResNet, Inception-V3, and some of their own proposed networks. In Table 2, this paper summarizes the classification methods of all gastric cancer histopathological images.

6. Method Analysis

6.1. The Gastric Histopathology Method Analysis. This paper summarizes the works on gastric cancer histopathology images from the perspectives of preprocessing, feature extraction, segmentation methods, and classifier design methods. Below, this paper briefly introduces the methods used in each step.

6.1.1. Image Preprocessing Methods. By summarizing the paper, image enhancement technology is the most commonly used method in image preprocessing. Experimental training requires a large amount of data, but the data obtained from experiments are often insufficient. At this time, the experiment needs to enhance the data image.

Figure 18 shows the preprocessing method of gastric cancer histopathological images. The work in [22, 23, 24] expanded the experimental data through data enhancement methods such as rotation and mirroring. This reduces the overfitting situation and provides a powerful help to the experiment.

6.1.2. Feature Extraction Methods. Histopathological feature extraction of gastric cancer includes feature extraction and postprocessing of features. Feature extraction methods mainly include color, texture, and shape features. Feature postprocessing is mainly about feature dimensionality

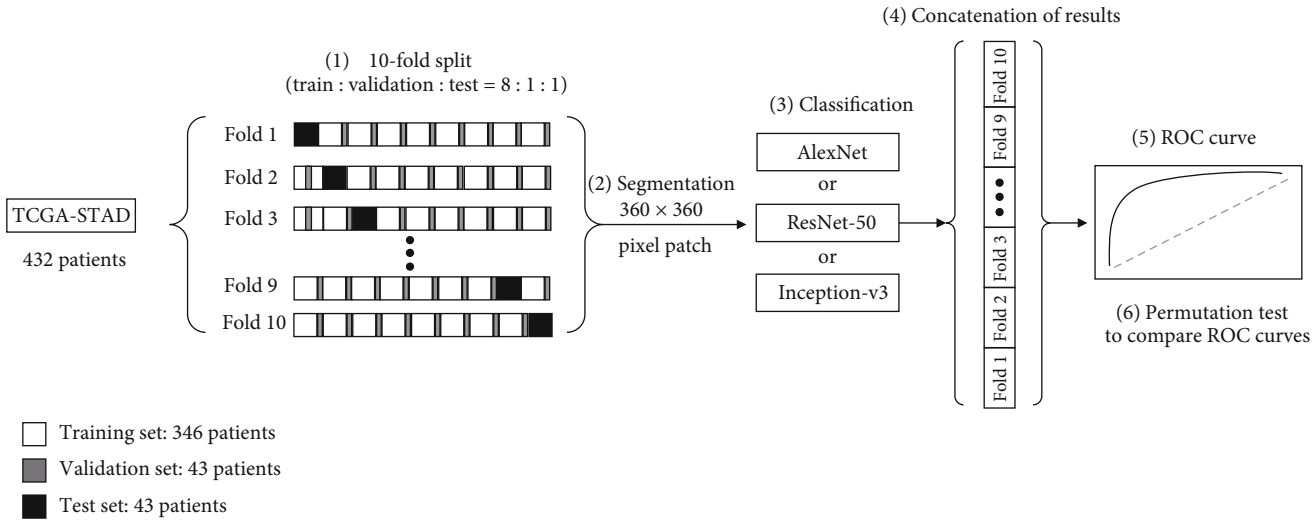


Figure 17: The framework of the proposed method in [55]. This figure corresponds to Figure 1 in the original paper.

Table 2: Classification-based gastric cancer image analysis.

Aim	Year	Reference	Team	Dataset	Method	Evaluation
Machine learning	2015	[27]	Sharma et al.	5,541 30x and 3,730 40x images	Adaboost	ACC = 59.15%
	2017	[32]	Sharma et al.	795 images	Random forest	ACC = 73.78%
	2018	[47]	Liu et al.	560 cancer and 140 noncancer	CNN and SVM	ACC = 95%
Deep learning	2017	[22]	Garcia et al.	3,275 images	DCNN	ACC = 96.88%
	2017	[23]	Sharma et al.	21,000 images	Proposed CNN	Accuracy of cancer classification is 69.9%; accuracy of necrosis detection is 81.4%
	2018	[24]	Li et al.	560 cancer and 140 noncancer	Proposed CNN	Accuracy of patch level classification is 97.93%; accuracy of slice level classification is 100%
	2018	[48]	Liu et al.	1.2 million images	ResNet	F – score = 96%
	2019	[51]	Wang et al.	608 whole slide images	ResNet	ACC = 86.5%
	2020	[52]	Song et al.	2,123 digital slides	DeepLab-V3 and ResNet	Acc = 87.3%, Sn = 99.6%, Sp = 84.3%

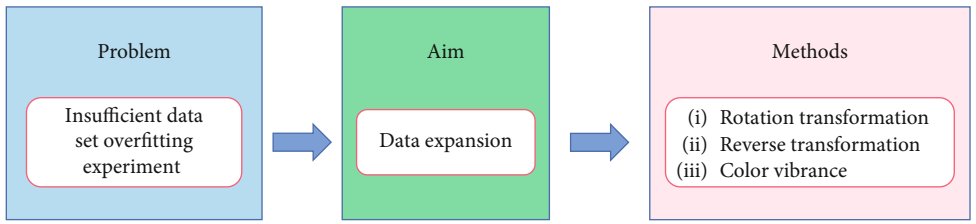


Figure 18: Image preprocessing method.

reduction. Figure 19 shows the feature extraction method for gastric cancer histopathological images.

(1) Traditional Feature Extraction Methods. The main methods of color feature are RGB histograms and HSV histograms. RGB histograms reflect the composition and distribution of colors in the image, that is, which colors appear and the probability of the appearance of various colors. The work in [56] firstly proposed the method of using color histogram as the representation of image color features. HSV histogram

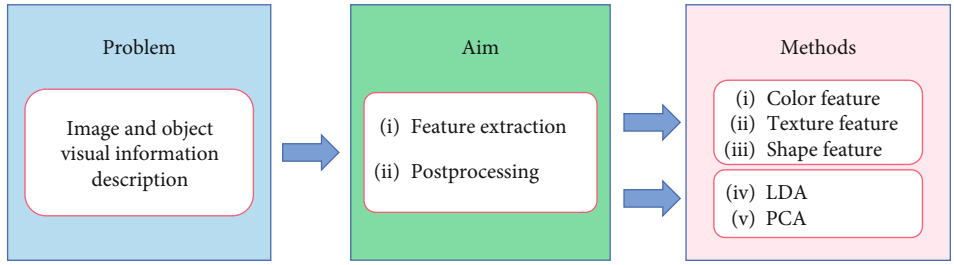


Figure 19: Feature extraction method.

[57, 58] is a further expression form of the RGB histogram. In RGB color space, only color information can be obtained, and the distance between colors cannot be used to represent the similarity. Therefore, features more representative of color information are extracted on the basis of RGB color space, for example, hue (H), saturation (S), and value (V). The histopathological images of gastric cancer are obtained mainly through H&E staining of gastric tissue sections. The colors of the tissues in the images obtained are in sharp contrast with each other and have relatively obvious color characteristics. The work in [23, 27] used HSV and RGB histograms to extract color features from gastric cancer histopathological images and achieved good results.

Texture feature is a kind of image global feature, which can describe the attributes of an image, mainly including HOG and GLCM. The HOG feature is a feature of statistical gradient direction change among pixels, which has good geometric invariance and optical invariance, and is excellent in human body detection. The work in [59] shows that HOG can be more effectively used in human detection than the existing feature. GLCM proposed by [60] is a matrix that describes the grayscale relationship between a certain pixel in a local area of an image and adjacent pixels or pixels within a certain distance. GLCM is a feature that describes the relationship between pixels. A gray-level cooccurrence matrix is constructed on the original image, and then, the statistical attributes in the matrix are extracted as feature vectors. Compared with normal gastric tissue cells, the texture of the cancerous stomach tissue cells changed, the nucleus became larger, and the shape of the nucleus became irregular. Extracting these texture features can distinguish normal gastric tissue from cancerous gastric tissue. In [28] [29], the HOG feature is drawn on normal, benign, and malignant gastric images to obtain the HOG feature vector.

Shape features are described by using an algorithm to get shape parameters. The cells in cancerous gastric tissues become dense, and the density between the cells can be expressed by describing the position relationship between the nuclei; thus, effective shape features can be extracted. The work [31] obtains shape features by extracting the location and structure of the nucleus. In [32], the authors extract the cell nuclei attributed relational graph of four types of cells as shape features.

(2) Postprocessing Methods. The dimension of feature extraction is very large in GHIA. The purpose of feature postprocessing is to reduce the dimension of features, reduce the running memory of the computer, and improve the working efficiency. The main methods include LDA and PCA. Both LDA and PCA use the idea of matrix decomposition to achieve dimensionality reduction of data.

LDA [61] is a method to realize the feature classification of two or more objects. It can be used for data dimensionality reduction or classification. It is supervised learning; LDA projects the data with a higher dimension into the vector space that can make the best discrimination. In the new vector space, the maximum interclass distance and the minimum intraclass distance of samples can be obtained. In this way, classification information and feature dimension reduction can be extracted better. The work in [29] uses LDA dimension reduction processing to get more efficient feature data.

PCA [62] is a commonly used data analysis method, which converts the original data into a set of linearly independent representations through linear transformation. It is a kind of unsupervised learning that can be used to extract the main features of the data and is often used to reduce the dimension. The work in [33] uses the dimension reduction method of PCA, and the redundant features in the original data are removed to make the variance of the projection on each dimension as large as possible.

6.1.3. Segmentation Methods. In terms of gastric cancer histopathological images, the existing image segmentation methods mainly include machine learning and deep learning. In terms of machine learning, the edge detection segmentation method is adopted, while the U-net network is used in deep learning. Figure 20 shows the segmentation methods of gastric cancer histopathological images.

Edge detection is to find out the point where the gray level of the image changes greatly, which can reduce the amount of data, find the place with the obvious boundary, and retain the important structural attributes of the image.

U-net [63] is a semantic segmentation network based on FCN, which is originally applied to the segmentation of medical cell microscopic images. In the encoder part, the input is downsampling and downsampling through maximum pooling. In the decoder part, upsampling is carried out for the output of encoder to restore the resolution, and upsampling is implemented by deconvolution. Skip-connect is used for feature fusion. The work in [41] uses U-net to segment gastric histopathology images and to compare with other neural network structures. U-net has a better segmentation effect.

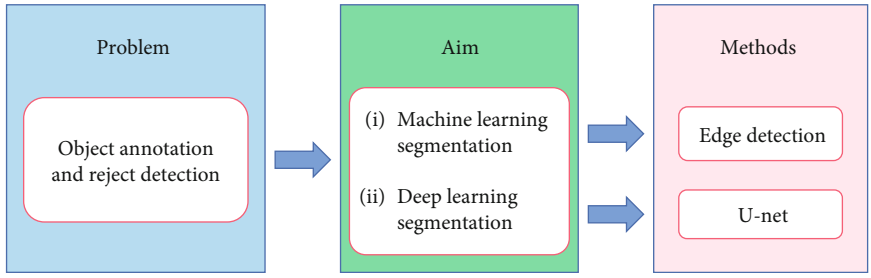


Figure 20: Segmentation method.

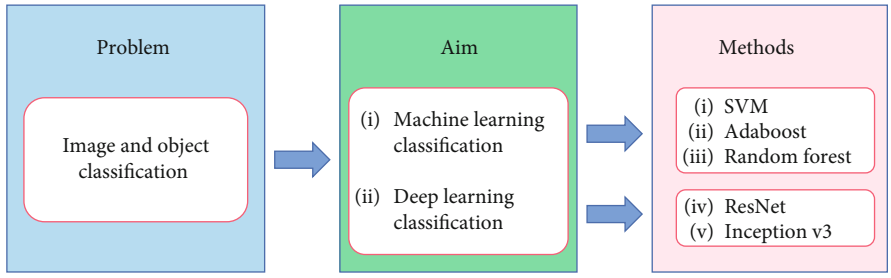


Figure 21: Classifier design methods.

6.1.4. Classifier Design Methods. In the gastric histopathology images, there are two kinds of classification designs, machine learning and deep learning. Machine learning classifier mainly uses SVM, Adaboost, and random forest. Deep learning uses convolutional neural networks for classification, such as ResNet and Inception-V3. Figure 21 shows the classifier design methods of the gastric cancer histopathological images.

(1) Machine Learning Classifier Design Methods. SVM [64] is a dichotomous model. By mapping data to a high-dimensional space and finding a hyperplane in the high-dimensional space, the distance between separated data classes is maximized, so as to achieve the purpose of optimal classification. This work [48] uses the SVM classifier and optimizes kernel function in the SVM classifier to improve classification accuracy.

The Adaboost [65] algorithm is a lifting method, which combines several weak classifiers into a strong classifier. Adaboost trains one weak classifier at a time and then iterates to train the next weak classifier after training. The work in [27] uses the Adaboost classifier to classify three feature groups of color, texture, and shape and finally obtains the optimal classification result.

Random forest [66] is an ensemble learning method in machine learning. This method integrates multiple decision trees through ensemble learning. Limitations can be avoided by integrating multiple models. This work [32] uses a random forest classifier to classify the shape features of the reticular map made up of nuclei.

(2) Deep Learning Classifier Design Methods. ResNet [67] applies the idea of residuals to the convolutional neural network. Different from the traditional convolutional neural network which directly represents the mapping relationship between input and output, it represents the residual between input and output through multiple parameterization layers. The authors in [48, 55] use an encoder based on ResNet to process the gastric histopathological image. Compared with other neural network structures, ResNet can provide end-to-end services for users and retain more detailed information on gastric cancer histopathological images.

Inception-V3 [68] uses convolutional kernels of different sizes in the convolutional layer to improve the perception of the network and proposes batch standardization to alleviate the problem of gradient disappearance. It uses better convolution kernel decomposition methods on the basis of Inception-V2 to make computational complexity more efficient, reduce representational bottlenecks of features, and avoid information loss. The work [54] has optimized the Inception-V3 network architecture to reduce network parameters, avoid overfitting, and improve network effectiveness.

6.2. The Potential Methods in Gastric Histopathology Image Analysis

6.2.1. Image Preprocessing Methods. In image preprocessing [69], methods in other fields can also be applied in the field of gastric cancer histopathology.

In [70], the authors develop an image denoising network based on CNN and optimize the denoising network by utilizing the advantages of complex numerical operations to increase the tightness of convolution.

In [71], the authors propose a new image denoising method using a new loss function, which pays attention to the perceived visual quality and applies the loss function to

the jump connection network to obtain images with high precision details. In the histopathological images of gastric cancer, the operation of sectioning and staining will affect the image quality and generate noise. The image denoising method in [70, 71] can be applied to the histopathology of gastric cancer to improve the image quality.

In [72], the authors propose a data expansion method. In the histopathology of gastric cancer, due to the problem of data collection, the amount of data is relatively small, so this data expansion method can be applied to the histopathology of gastric cancer to increase the dataset and improve the experimental quality.

In [73], the authors propose a method for contrast enhancement of retinal images, which uses a shearlet transform and adaptive gamma correction-based singular value equalization hybrid technique combined with a contrast-limited adaptive histogram equalization technique. In the histopathological images of gastric cancer, HE staining is usually used in the image, which has distinct color features. The image contrast enhancement technology in [73] can make the color features more prominent and improve the image quality.

6.2.2. Feature Extraction Methods. The work in [74] summarizes the feature extraction methods. Feature extraction is also an important step in gastric cancer histopathological images, and many feature extraction techniques in [74] can be applied to gastric cancer histopathological images.

In [75], the authors extract image feature vectors based on computed radiography and obtain 279 image descriptors. Through the Hellwig method, they find the feature combination set of the overall index with the maximum information capacity. Then, by testing the classification results of 11 classifiers, they select the two feature sets with the highest accuracy. This method can also be applied to the feature selection of gastric cancer histopathological images to select the most effective feature.

In [76], the authors propose a new unsupervised feature selection method, which calculates the dependencies between features and avoids selecting redundant features. The experimental operation efficiency is improved. In the histopathological images of gastric cancer, the experimental data are large, the system execution time is long, and the memory demand is large during the experiment. The feature selection method mentioned in [76] can also be applied to the histopathology of gastric cancer to improve the operating environment of the system and the operation efficiency of the experiment.

6.2.3. Segmentation Methods. The work in [77] summarizes the segmentation techniques applied in various fields. In the histopathology of gastric cancer, image segmentation is often used, and the current mainstream segmentation techniques can also be applied in the histopathology of gastric cancer.

In medical images, the authors in [78] propose a new polyp segmentation method based on multidepth codec network combination. The network can extract the features of different effective receptive fields and multisize images to represent the multilevel information, and it can also extract the effective information features from the missing pixels in the training stage. The method of [78] can be applied to gastric cancer histopathological images,

In [79], the authors propose an image segmentation method based on a scalable multichannel weighted region model. In order to improve the performance of image Mosaic, a new edge detection function is proposed. The model can also be used to segment images in gastric cancer histopathology.

In [80], the work proposes a radiology-based deep supervised U-net for the segmentation of prostate and prostatic lesions. The U-net network can also be used to segment the normal and cancerous areas in the histopathological images of gastric cancer.

In [81], the work proposes an image segmentation method combining low-level operation, affine probabilistic graph, and multigraph. This method is used for segmentation of computed tomography (CT) images. Experimental results show that compared with atlas selection and nonrigid registration, this method has better performance in the whole region of interest. This method can also be applied to histopathological image segmentation of gastric cancer and obtain good results.

6.2.4. Classifier Design Methods. In [82], the authors summarize various classifier techniques, many of which can be applied to gastric cancer histopathological images.

In [83], the authors used a combination of convolutional neural network and machine learning classifier to classify traffic density and compared CNN, CNN-SVM, CNN-RF, and CNN-Xgboot to select an appropriate model. This method can also be applied to gastric cancer histopathological images. Features are extracted through deep learning network and then classified by machine learning method.

In [84], the authors propose a new method for classification of breast masses based on the deep learning model. It can automatically process small two-dimensional radiofrequency signals and their amplitude samples. A deep learning classification model of breast masses was constructed using radiofrequency data. This classification model can also be used for the classification of gastric cancer pathology.

In [85], the authors propose a Deep CNN based on the Dolphin Echolocation-based Sine Cosine Algorithm. The algorithm uses Dolphin-SCA-based fuzzy fusion model for segmentation, and then, the statistical attributes of weight, mean value, variance, and skewness are used for feature calculation. Finally, the convolutional neural network is used for classification. This method can also be applied to the histopathology of gastric cancer.

In [86], the authors propose a dictionary learning method based on multiclass loss feedback discrimination for support vector machines. The framework learns the discriminant dictionary while training the SVM, which makes the features extracted from the learner dictionary better matched with the SVM. This classification model can also be used for the classification of gastric cancer pathology.

6.3. The Gastric Histopathology Methods for Other Potential Fields

6.3.1. Image Preprocessing Methods. In the histopathology of gastric cancer, the image preprocessing method is mainly

data enhancement. The data enhancement techniques can apply to other areas, such as database enhancements in the field of microbiology. In [87], the microbial data are expanded by combining geometric transformation and GAN network. In the pathological images of breast cancer, the number of datasets is also insufficient. The data enhancement technique is also applicable to increase the dataset to make the experiment more accurate. In [88], the authors use the geometric transformation method to increase the number of datasets, reduce the overfitting situation, and make the experiment more accurate.

6.3.2. Feature Extraction Methods. In the histopathology of gastric cancer, CRF and GLCM methods are used for feature extraction. These methods are also widely used in the field of microbiology because the cell morphology and size of gastric cancer histopathology are similar to that of microorganisms. In [89], a CNN-CRF network framework is adopted to extract features, which CRF is used for feature postprocessing. Texture feature is also a common feature extraction method in breast cancer pathological images. In [90], the GLCM method is used for feature extraction.

6.3.3. Segmentation Methods. In gastric cancer histopathology, image segmentation includes threshold segmentation based on color and edge detection based on texture. Compared with the gastric cancer histopathological image, the color contrast between the microorganisms and the background in the microbial image is also relatively bright, and threshold segmentation can be adopted. The microbe's texture and background are also quite different, so edge detection can be used. In [91], the authors use an image segmentation method related to color and texture. In the pathology of cervical cancer [92], because cervical cancer cells are similar to gastric cancer cells, the segmentation method of gastric cancer can also be used in the pathological images of cervical cancer, and the minimum-model method can also be used in cervical cancer.

6.3.4. Classifier Design Methods. In the histopathology of gastric cancer, SVM and RF methods were used for image classification. These methods can also be applied to microbial images and pathological images of cervical cancer. In [93], the methods of microbial classification are summarized and it is found that the methods applied to gastric cancer histopathological images can be applied to microbial images. In [94], SVM classifier is used to classify cervical cancer.

7. Conclusion and Future Work

This paper reviews the methods of histopathological image analysis of gastric cancer, including preprocessing, feature extraction, segmentation, and classification. In preprocessing, in order to solve the problem of lack of data, the method of data enhancement is adopted. The data enhancement method is mainly image rotation and geometric transformation. In terms of feature extraction, it is summarized from two aspects: the machine learning method and the deep learning method. The machine learning method includes color feature, texture feature, and morphological feature. In terms of image segmentation, there are many papers adopting the machine learning method, mainly adopting edge detection, the segmentation method, and the U-net convolutional neural network for deep learning. In terms of image classification, machine learning classifiers such as SVM and RM are mainly applied. Deep learning networks employ some classical convolutional neural network structures, such as ResNet and Inception-V3. Others are frameworks combining machine learning and deep learning, extracting features by deep learning network, and then classified by machine learning classifier.

In the future, there is still space for improvement in the histopathological image analysis of gastric cancer. First of all, there are few papers related to image analysis methods in the field of gastric cancer. Researchers can develop a new network model combining with gastric cancer histopathological images to analyze gastric cancer histopathological images. The following techniques, for example, "DoDNet: Learning to Segment Multiorgan and Tumors from Multiple Partially Labeled Datasets" [95], "3D Cascaded Convolutional Networks for Multivertebrae Segmentation" [96], and "PGL: Prior-Guided Local Self-supervised Learning for 3D Medical Image Segmentation" [97], can be applied to the pathological image processing of gastric cancer. Secondly, in the field of gastric cancer histopathology, there is a lack of complete, clear, and accurately labeled pathological images, so the establishment of more perfect data can provide great help for the experiment. Finally, in terms of feature extraction and classifier design, there are many novel techniques that can be applied to gastric cancer histopathological image analysis, which is a promising and valuable research direction.

Disclosure

Chen Li works as co-first author and corresponding author in this paper. Shiliang Ai and Chen Li are co-first authors.

Conflicts of Interest

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

Acknowledgments

The authors thank Miss Zixian Li and Mr. Guoxian Li for their important discussion. We acknowledge the financial support from the National Natural Science Foundation of China (No. 61806047), the Fundamental Research Funds for the Central Universities (No. N2019003), and the China Scholarship Council (2018GBJ001757).

References

[1] O. M. Yis, G. Bugdayci, M. B. Pehlivan, and M. Basol, "Roles of the systemic inflammatory response biomarkers in the diagnosis of cancer patients with solid tumors," *Experimental Biomedical Research*, vol. 2, no. 1, pp. 37–43, 2019.

[2] A. O.-O. Chan, B. C.-Y. Wong, and S.-K. Lam, "Gastric cancer: past, present and future," *Canadian Journal of Gastroenterology*, vol. 15, no. 7, pp. 469–474, 2001.

[3] T. M. Elsheikh, R. M. Austin, D. F. Chhieng et al., "American Society of Cytopathology workload recommendations for automated Pap test screening: developed by the productivity and quality assurance in the era of automated screening task force," *Diagnostic Cytopathology*, vol. 41, no. 2, pp. 174–178, 2013.

[4] R. Lozano, "Comparison of computer-assisted and manual screening of cervical cytology," *Gynecologic Oncology*, vol. 104, no. 1, pp. 134–138, 2007.

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

[6] H. Sharma, "Appearance-based necrosis detection using textural features and SVM with discriminative thresholding in histopathological whole slide images," in *IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)*, IEEE, 2015.

[7] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," *The British Journal of Radiology*, vol. 78, Supplement 1, pp. s3–s19, 2005.

[8] E. Bengtsson and P. Malm, "Screening for cervical cancer using automated analysis of PAP-smears," *Computational and Mathematical Methods in Medicine*, vol. 2014, Article ID 842037, 12 pages, 2014.

[9] L. Zhang, H. Kong, C. Ting Chin et al., "Automation-assisted cervical cancer screening in manual liquid-based cytology with hematoxylin and eosin staining," *Cytometry Part A*, vol. 85, no. 3, pp. 214–230, 2014.

[10] K. Y. Win, S. Choomchuay, K. Hamamoto, M. Raveesunthornkiat, L. Rangsirattanakul, and S. Pongsawat, "Computer aided diagnosis system for detection of cancer cells on cytological pleural effusion images," *BioMed Research International*, vol. 2018, Article ID 6456724, 21 pages, 2018.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[12] W. Rogers, B. Ryack, and G. Moeller, "Computer-aided medical diagnosis: literature review," *International Journal of Bio-Medical Computing*, vol. 10, no. 4, pp. 267–289, 1979.

[13] A. N. Ramesh, C. Kambhampati, J. R. T. Monson, and P. J. Drew, "Artificial intelligence in medicine," *Annals of the Royal College of Surgeons of England*, vol. 86, no. 5, pp. 334–338, 2004.

[14] C. Demir and B. Yener, *Automated Cancer Diagnosis Based on Histopathological Images: A Systematic Survey*, Rensselaer Polytechnic Institute, Tech. Rep, 2005.

[15] A. Vellido and P. J. G. Lisboa, "Neural networks and other machine learning methods in cancer research," in *International Work-Conference on Artificial Neural Networks*, Springer, Berlin, Heidelberg, 2007.

[16] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: a review of literature," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 47, 2018.

[17] T. J. Fuchs and J. M. Buhmann, "Computational pathology: challenges and promises for tissue analysis," *Computerized Medical Imaging and Graphics*, vol. 35, no. 7-8, pp. 515–530, 2011.

[18] J. Gilbertson and Y. Yagi, "Histology, imaging and new diagnostic work-flows in pathology," *Diagnostic Pathology*, vol. 3, Supplement 1, p. S14, 2008.

[19] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: a review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[20] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, IEEE, 2016.

[21] M. Sonka, V. Hlavac, and R. Boyle, *Image pre-processing*, Image Processing, Analysis and Machine Vision, Springer, Boston, MA, 1993.

[22] E. Garcia, "Automatic lymphocyte detection on gastric cancer IHC images using deep learning," in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2017.

[23] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, and P. Hufnagl, "Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology," *Computerized Medical Imaging and Graphics*, vol. 61, pp. 2–13, 2017.

[24] Y. Li, "Deep learning based gastric cancer identification," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, 2018.

[25] J. Qu, N. Hiruta, K. Terai, H. Nosato, M. Murakawa, and H. Sakanashi, "Gastric pathology image classification using stepwise fine-tuning for deep neural networks," *Journal of healthcare engineering*, vol. 2018, Article ID 8961781, 13 pages, 2018.

[26] I. Guyon, *Feature Extraction: Foundations and Applications*, vol. 207, Springer, 2008.

[27] H. Sharma, N. Zerbe, D. Heim et al., "A multi-resolution approach for combining visual information using nuclei segmentation and classification in histopathological images," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*, Berlin, Germany, 2015.

[28] S. A. Korkmaz, A. Akcicek, H. Binol, and M. F. Korkmaz, "Recognition of the stomach cancer images with probabilistic HOG feature vector histograms by using HOG features," in *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*, IEEE, 2017.

[29] S. A. Korkmaz, H. Binol, A. Akcicek, and M. F. Korkmaz, "A expert system for stomach cancer images with artificial neural network by using HOG features and linear discriminant analysis: HOG_LDA_ANN," in *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*, IEEE, 2017.

[30] S. A. Korkmaz and H. Binol, "Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection," *Journal of Molecular Structure*, vol. 1156, pp. 255–263, 2018.

[31] E. Cosatto, P.-F. Laquerre, C. Malon et al., "Automated gastric cancer diagnosis on H&E-stained sections; ltraining a classifier on a large scale with multiple instance machine learning," in *Medical Imaging 2013: Digital Pathology*, vol. 8676, SPIE, 2013.

[32] H. Sharma, N. Zerbe, C. Boger, S. Wienert, O. Hellwich, and P. Hufnagl, "A comparative study of cell nuclei attributed relational graphs for knowledge description and categorization in histopathological gastric cancer whole slide images," in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2017.

[33] L. Gan, W. Lv, X. Zhang, and X. Meng, "Improved PCA + LDA applies to gastric cancer image classification process," *Physics Procedia*, vol. 24, pp. 1689–1695, 2012.

[34] R. Geetha, S. Sivasubramanian, M. Kaliappan, S. Vimal, and S. Annamalai, "Cervical cancer identification with synthetic minority oversampling technique and PCA analysis using random forest classifier," *Journal of Medical Systems*, vol. 43, no. 9, 2019.

[35] D. Kashyap, A. Somani, J. Shekhar et al., "Cervical cancer detection and classification using independent level sets and multi SVMs," in *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2016.

[36] 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.

[37] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, 2008.

[38] W. Liu, Z. Sun, J. Chen, and C. Jing, "Raman spectroscopy in colorectal cancer diagnostics: comparison of PCA-LDA and PLS-DA models," *Journal of Spectroscopy*, vol. 2016, Article ID 1603609, 6 pages, 2016.

[39] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 1, pp. 100–132, 1985.

[40] Y. Li, X. Xie, S. Liu, X. Li, and L. Shen, "GT-Net: a deep learning network for gastric tumor diagnosis," in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 2018.

[41] M. Sun, G. Zhang, H. Dang, X. Qi, X. Zhou, and Q. Chang, "Accurate gastric cancer segmentation in digital pathology images using deformable convolution and multi-scale embedding networks," *IEEE Access*, vol. 7, pp. 75530–75541, 2019.

[42] Y. Nan, "Partial labeled gastric tumor segmentation via patch-based reiterative learning," 2017, http://arxiv.org/abs/1712.7488.

[43] C. Sun, C. Li, J. Zhang, F. Kulwa, and X. Li, "Hierarchical conditional random field model for multi-object segmentation in gastric histopathology images," *Electronics Letters*, vol. 56, no. 15, pp. 750–753, 2020.

[44] anonymous, "Gastric segmenting," *Electronics Letters*, vol. 56, pp. 747–747, 2020.

[45] R. M. Cormack, "A review of classification," *Journal of the Royal Statistical Society: Series A (General)*, vol. 134, no. 3, pp. 321–353, 1971.

[46] S. Wienert, D. Heim, K. Saeger et al., "Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach," *Scientific Reports*, vol. 2, no. 1, 2012.

[47] B. Liu, M. Zhang, T. Guo, and Y. Cheng, "Classification of gastric slices based on deep learning and sparse representation," in *2018 Chinese Control and Decision Conference (CCDC)*, IEEE, 2018.

[48] B. Liu, K. Yao, M. Huang, J. Zhang, Y. Li, and R. Li, "Gastric pathology image recognition based on deep residual networks," in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, IEEE, 2018.

[49] Z. Zhu, X. Ding, D. Zhang, and L. Wang, "Weakly-supervised balanced attention network for gastric pathology image localization and classification," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2020.

[50] J. Kloeckner, T. K. Sansonowicz, Á. L. Rodrigues, and T. W. N. Nunes, "Multi-categorical classification using deep learning applied to the diagnosis of gastric cancer," *Jornal Brasileiro de Patologia e Medicina Laboratorial*, vol. 56, 2020.

[51] S. Wang, Y. Zhu, L. Yu et al., "RMDL: recalibrated multi-instance deep learning for whole slide gastric image classification," *Medical Image Analysis*, vol. 58, article 101549, 2019.

[52] Z. Song, S. Zou, W. Zhou et al., "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning," *Nature Communications*, vol. 11, no. 1, article 4294, 2020.

[53] S. C. Kosaraju, J. Hao, H. M. Koh, and M. Kang, "Deep-Hipo: multi-scale receptive field deep learning for histopathological image analysis," *Methods*, vol. 179, pp. 3–13, 2020.

[54] O. Iizuka, F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, and M. Tsuneki, "Deep learning models for histopathological classification of gastric and colonic epithelial tumours," *Scientific Reports*, vol. 10, no. 1, article 1504, 2020.

[55] K.-O. Cho, S. H. Lee, and H.-J. Jang, "Feasibility of fully automated classification of whole slide images based on deep learning," *The Korean Journal of Physiology & Pharmacology*, vol. 24, no. 1, pp. 89–99, 2020.

[56] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[57] T. Carron and P. Lambert, "Color edge detector using jointly hue, saturation and intensity," in *Proceedings of 1st International Conference on Image Processing*, vol. 3, IEEE, 1994.

[58] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval," in *Proceedings. International Conference on Image Processing*, vol. 2, IEEE, 2002.

[59] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, IEEE, 2005.

[60] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, pp. 610–621, 1973.

[61] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis—a brief tutorial," *Institute for Signal and information Processing*, vol. 18, 1998.

[62] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[63] 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*, Springer, Cham, 2015.

[64] W. S. Noble, "What is a support vector machine?," *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[65] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[66] A. Liaw and M. Wiener, "Classification and regression by RandomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[67] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2016.

[68] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision,"

in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2016.

[69] A. Rehman and T. Saba, "Retracted article: Analysis of advanced image processing to clinical and preclinical decision making with prospectus of quantitative imaging biomarkers," *The Artificial Intelligence Review*, vol. 42, no. 4, article 1045, 2014.

[70] Y. Quan, Y. Chen, Y. Shao, H. Teng, Y. Xu, and H. Ji, "Image denoising using complex-valued deep CNN," *Pattern Recognition*, vol. 111, article 107639, 2021.

[71] A. F. M. S. Uddin, T. Chung, and S.-H. Bae, "A perceptually inspired new blind image denoising method using L1 and perceptual loss," *IEEE Access*, vol. 7, pp. 90538–90549, 2019.

[72] I.-J. Yu, W. Ahn, S. H. Nam, and H. K. Lee, "BitMix: data augmentation for image steganalysis," *Electronics Letters*, vol. 56, no. 24, pp. 1311–1314, 2020.

[73] G. Palanisamy, N. B. Shankar, P. Ponnusamy, and V. P. Gopi, "A hybrid feature preservation technique based on luminosity and edge based contrast enhancement in color fundus images," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 2, pp. 752–763, 2020.

[74] S. Ding, H. Zhu, W. Jia, and C. Su, "A survey on feature extraction for pattern recognition," *Artificial Intelligence Review*, vol. 37, no. 3, pp. 169–180, 2012.

[75] Z. Omiotek, O. Stepanchenko, W. Wójcik, W. Legieć, and M. Szatkowska, "The use of the Hellwig's method for feature selection in the detection of myeloma bone destruction based on radiographic images," *Biocybernetics and biomedical engineering*, vol. 39, no. 2, pp. 328–338, 2019.

[76] H. Lim and D.-W. Kim, "Pairwise dependence-based unsupervised feature selection," *Pattern Recognition*, vol. 111, article 107663, 2021.

[77] Z. Wang, E. Wang, and Y. Zhu, "Image segmentation evaluation: a survey of methods," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5637–5674, 2020.

[78] N.-Q. Nguyen and S.-W. Lee, "Robust boundary segmentation in medical images using a consecutive deep encoder-decoder network," *IEEE Access*, vol. 7, pp. 33795–33808, 2019.

[79] H. Jiang, S. Piao, M. Z. Qadir, and Q. Guo, "M-WRSF model for medical image segmentation," *Electronics Letters*, vol. 55, no. 7, pp. 386–389, 2019.

[80] P. Hambarde, S. Talbar, A. Mahajan, S. Chavan, M. Thakur, and N. Sable, "Prostate lesion segmentation in MR images using radiomics based deeply supervised U-Net," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1421–1435, 2020.

[81] C. Platero and M. C. Tobar, "A multiatlas segmentation using graph cuts with applications to liver segmentation in CT scans," *Computational and Mathematical Methods in Medicine*, vol. 2014, Article ID 182909, 16 pages, 2014.

[82] A. Carreño, I. Inza, and J. A. Lozano, "Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework," *Artificial Intelligence Review*, vol. 53, no. 5, pp. 3575–3594, 2020.

[83] K.-H. N. Bui, O. Hyeonjeong, and H. Yi, "Traffic density classification using sound datasets: an empirical study on traffic flow at asymmetric roads," *IEEE Access*, vol. 8, pp. 125671–125679, 2020.

[84] P. Jarosik, Z. Klimonda, M. Lewandowski, and M. Byra, "Breast lesion classification based on ultrasonic radio-frequency signals using convolutional neural networks," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 977–986, 2020.

[85] S. Kumar and D. P. Mankame, "Optimization driven deep convolution neural network for brain tumor classification," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1190–1204, 2020.

[86] B.-Q. Yang, X.-P. Guan, J.-W. Zhu, C.-C. Gu, K.-J. Wu, and J.-J. Xu, "SVMs multi-class loss feedback based discriminative dictionary learning for image classification," *Pattern Recognition*, vol. 112, article 107690, 2021.

[87] H. Xu, C. Li, M. M. Rahaman et al., "An enhanced framework of generative adversarial networks (EF-GANs) for environmental microorganism image augmentation with limited rotation-invariant training data," *IEEE Access*, vol. 8, pp. 187455–187469, 2020.

[88] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, "Automatic nuclei segmentation in H&E stained breast cancer histopathology images," *PLoS One*, vol. 8, no. 7, article e70221, 2013.

[89] J. Zhang, C. Li, F. Kulwa et al., "A multiscale CNN-CRF framework for environmental microorganism image segmentation," *BioMed Research International*, vol. 2020, Article ID 4621403, 27 pages, 2020.

[90] A. D. Belsare, M. M. Mushrif, M. A. Pangarkar, and N. Meshram, "Classification of breast cancer histopathology images using texture feature analysis," in *TENCON 2015-2015 IEEE Region 10 Conference*, IEEE, 2015.

[91] F. Kulwa, C. Li, X. Zhao et al., "A state-of-the-art survey for microorganism image segmentation methods and future potential," *IEEE Access*, vol. 7, pp. 100243–100269, 2019.

[92] B. C. Patel and G. R. Sinha, "An adaptive K-means clustering algorithm for breast image segmentation," *International Journal of Computer Applications*, vol. 10, no. 4, pp. 35–38, 2010.

[93] C. Li, K. Wang, and X. Ning, "A survey for the applications of content-based microscopic image analysis in microorganism classification domains," *Artificial Intelligence Review*, vol. 51, no. 4, pp. 577–646, 2019.

[94] Y. Rejani and S. T. Selvi, "Early detection of breast cancer using SVM classifier technique," 2009, http://arxiv.org/abs/912.2314.

[95] J. Zhang, "DoDNet: learning to segment multi-organ and tumors from multiple partially labeled datasets," 2020, http://arxiv.org/abs/10217.

[96] L. Xia, L. Xiao, G. Quan, and W. Bo, "3D cascaded convolutional networks for multi-vertebrae segmentation," *Current Medical Imaging*, vol. 16, no. 3, pp. 231–240, 2020.

[97] Y. Xie, "PGL: prior-guided local self-supervised learning for 3D medical image segmentation," 2020, http://arxiv.org/abs/2011.12640.

Hindawi
BioMed Research International
Volume 2021, Article ID 5561125, 18 pages
https://doi.org/10.1155/2021/5561125

Research Article

Multichannel Retinal Blood Vessel Segmentation Based on the Combination of Matched Filter and U-Net Network

Yuliang Ma,[1] **Zhenbin Zhu**,[1] **Zhekang Dong**,[2] **Tao Shen**,[3] **Mingxu Sun**,[3] **and Wanzeng Kong**[4]

[1]*Institute of Intelligent Control and Robotics, Hangzhou Dianzi University, Hangzhou, 310018 Zhejiang, China*
[2]*School of Electronics and Information, Hangzhou Dianzi University, Hangzhou, 310018 Zhejiang, China*
[3]*School of Electrical Engineering, University of Jinan, Jinan, 250022 Shandong, China*
[4]*Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province, Hangzhou, 310018 Zhejiang, China*

Correspondence should be addressed to Yuliang Ma; mayuliang@hdu.edu.cn

Received 14 January 2021; Revised 7 April 2021; Accepted 5 May 2021; Published 26 May 2021

Academic Editor: Changming Sun

Aiming at the current problem of insufficient extraction of small retinal blood vessels, we propose a retinal blood vessel segmentation algorithm that combines supervised learning and unsupervised learning algorithms. In this study, we use a multiscale matched filter with vessel enhancement capability and a U-Net model with a coding and decoding network structure. Three channels are used to extract vessel features separately, and finally, the segmentation results of the three channels are merged. The algorithm proposed in this paper has been verified and evaluated on the DRIVE, STARE, and CHASE_DB1 datasets. The experimental results show that the proposed algorithm can segment small blood vessels better than most other methods. We conclude that our algorithm has reached 0.8745, 0.8903, and 0.8916 on the three datasets in the sensitivity metric, respectively, which is nearly 0.1 higher than other existing methods.

1. Introduction

The human eyes consist of the following parts: cornea, pupil, iris, vitreous, and retina. Abnormalities in any of these tissue structures may cause vision defects or even blindness. Among them, the study of retinal structure and its blood vessels is significant [1]. The extraction of retinal blood vessels and the characterization of morphological properties, such as diameter, shape, distortion, and bifurcation, can be used to screen, evaluate, and treat different ocular abnormalities [2]. Evaluation of retinal vascular properties, such as changes in width, is used to analyze hypertension, while bifurcation points and tortuosity can help identify cardiovascular disease and diabetic retinopathy [3].

The retinal vessel extraction methods, including pattern recognition, are classified into five core classes [4]. The pattern recognition techniques are generally divided into two categories: supervised learning and unsupervised learning. The supervised learning method needs to use manual segmentation images of ophthalmologists for training. This method requires many training images, and the training time is longer than that of other methods, but this method has an excellent generalized effect and can be applied to other images of the same type. Compared with supervised learning, nonsupervised learning methods, such as matched filtering, mathematical morphology operations, blood vessel tracking, and clustering, do not require corresponding image labels but analyze and process based on the existing data. These two types of methods have been applied and innovated by many researchers in recent years.

1.1. Unsupervised Learning Methods. Literature [5] proposed a new kernel-based technique, viz, Fréchet PDF-based matched filter. The new method performs a better matching between the vessel profile and Fréchet template. Literature [6] improved the extraction method of blood vessels, using a series of morphological operations to extract small blood vessels, and finally fused with the segmented image to supple-

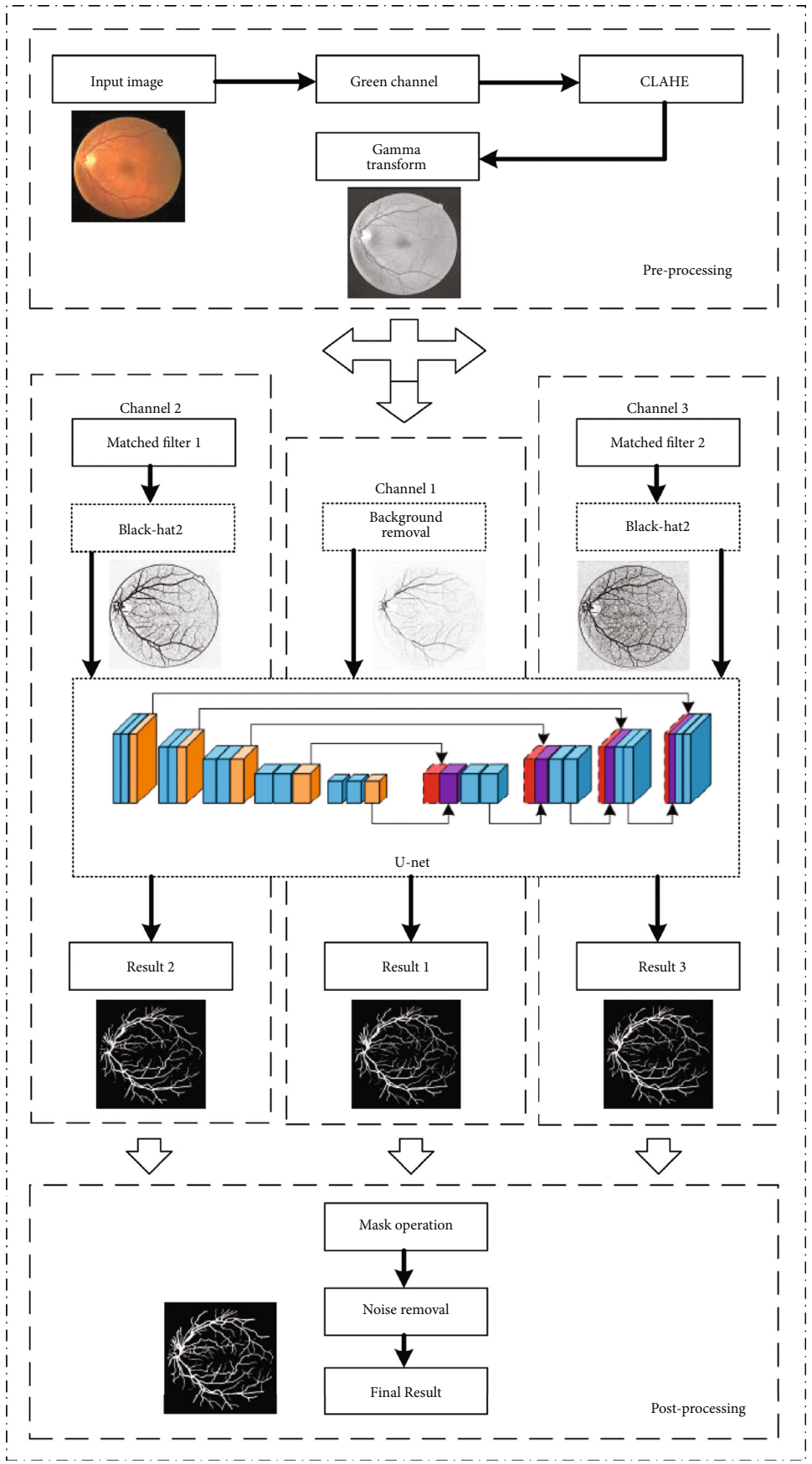


Figure 1: Proposed model.

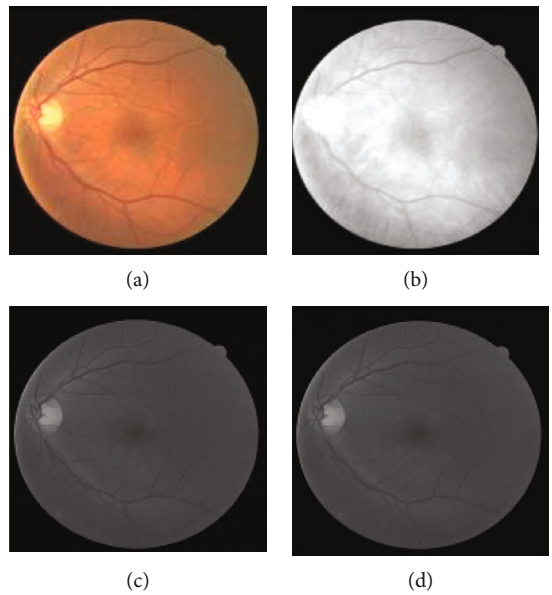


Figure 2: Color fundus image and its different RGB channels: (a) original RGB image; (b) red channel; (c) green channel; (d) blue channel.

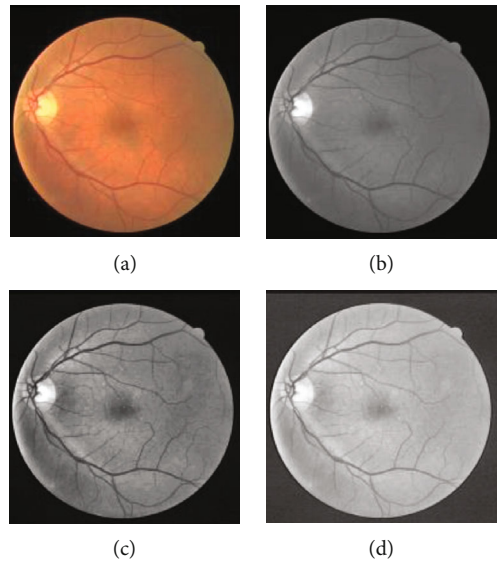


Figure 3: Typical images after each preprocessing step: (a) original RGB image; (b) red channel; (c) image after CLAHE operation; (d) image after gamma correction.

ment the small blood vessels. Compared with other algorithms, it can segment as many tiny blood vessels as possible. However, the steps of the algorithm are too complicated, and although the final segmentation effect obtains the smallest blood vessels, the small blood vessels are in an intermittent state as a whole, and they are not well connected with thicker blood vessels. Literature [7] proposed a new matched filtering method, which applies contrast-limited adaptive histogram equalization and Gaussian second-derivative-based matched filter in preprocessing and uses an entropy-based optimal threshold method performing binarization. This algorithm effectively improves the sensitivity metric of segmentation, but like literature [6], it does not perform well with accuracy. Literature [8] proposed an automatic segmentation method of retinal blood vessels using a matched filter and fuzzy C-means clustering. The algorithm uses contrast-limited adaptive histogram equalization to enhance the contrast of the image. After using Gabor and Frangi filters for noise removal and background removal, the fuzzy C-means are used to extract the initial vascular network, and the integrated level set method is used to refine segmentation further. The algorithm has good sensitivity and specificity. The problem is that the ability to segment small blood vessels is limited, and many segmentation details are missed. Literature [9] proposed a novel method to extract the retinal blood vessel using local contrast normalization and a second-order detector.

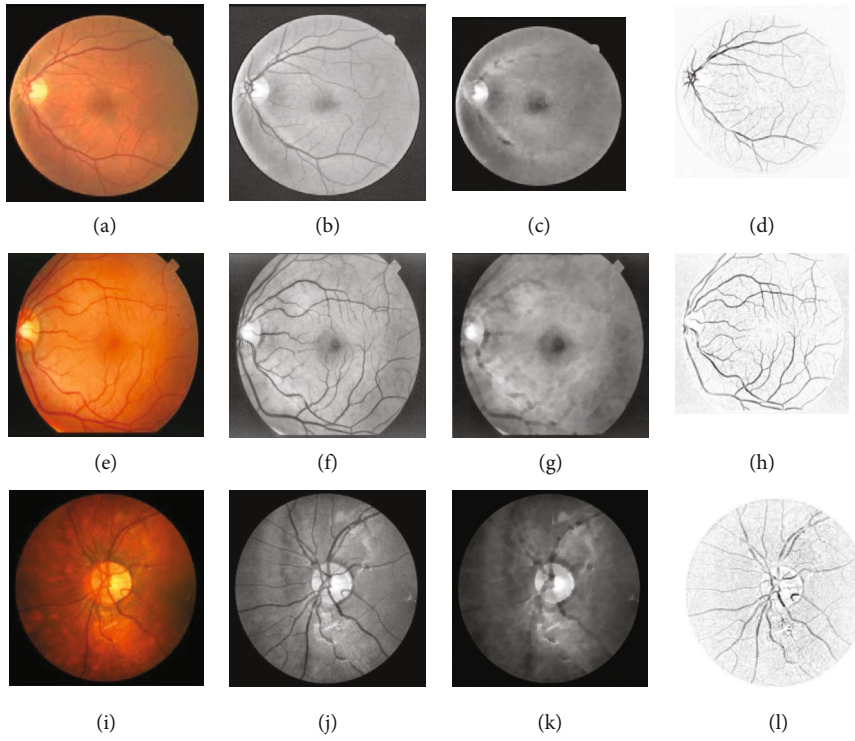


Figure 4: (a, e, i) The fundus images, (b, f, j) green channel images applying CLAHE and gamma transformation, (c, g, k) background extracted by close operation, and (d, h, l) the final results. The different samples of (a–d) DRIVE, (e–h) STARE, and (i–l) CHASE_DB1.

The proposed methodology achieves higher accuracy in vessel segmentation than existing techniques. Literature [10] proposed a novel matched filter approach with the Gumbel probability distribution function as its kernel. The reason to achieve the higher accuracy is due to a better matching filter with the Gumbel PDF-based kernel.

1.2. Supervised Learning Methods. Literature [11] proposed a method using deep conventional neural networks and a hysteresis threshold method to detect the vessels accurately. The proposed method gives good performance in which more tiny vessels are detected. Literature [12] proposed a multilevel CNN model applied for automatic blood vessel segmentation in retinal fundus images. A novel max-resizing technique is proposed to improve the generalization of the training procedure for predicting blood vessels from retinal fundus images. Literature [13] proposed a new segment-level loss used with the pixel-wise loss to balance the importance between thick vessels and thin vessels in the training process. Literature [14] proposed a cross-connected convolutional neural network (CcNet) to automatically segment retinal vessel trees. The cross connections between a primary path and a secondary path fuse the multilevel features. This method has relatively advanced performances, including competitive strong robustness and segmentation speed. Literature [15] proposed a method for retinal vessel segmentation using patch-based fully convolutional networks. Literature [16] applied dilated convolutions in a deep neural network to improve the segmentation of retinal blood vessels from fundus images. Literature [17] proposed a new improved algorithm based on the U-Net network model. The algorithm integrates the Inception-Res structure module and the Dense-Inception structure module into the U-Net structure. The algorithm dramatically deepens the depth of the network but does not add additional training parameters. It has good segmentation performance in the image segmentation of retinal blood vessels and has strong generalization ability. Literature [18] proposed a new hybrid algorithm for retinal vessel segmentation on fundus images. The proposed algorithm applies a new directionally sensitive blood vessel enhancement before sending fundus images to U-Net. Literature [19] proposed a supervised method based on a pretrained fully convolutional network through transfer learning. This method simplifies the typical retinal vessel segmentation problem into regional semantic vessel element segmentation tasks. Generally, unsupervised methods are less complex and suffer from relatively lower accuracy than supervised methods [13].

To solve the problem of insufficient segmentation of small blood vessels in most papers, we have devised a new automatic segmentation framework for retinal vessels based on improving U-Net and a multiscale matched filter. The creative points of this paper are summarized as follows:

(1) We proposed an improved black hat algorithm to enhance the characteristics of blood vessels and reduce the interference of other tissues

(2) An algorithm combining a multiscale matched filter and U-Net neural network is proposed. This paper mainly uses the improved U-Net convolutional

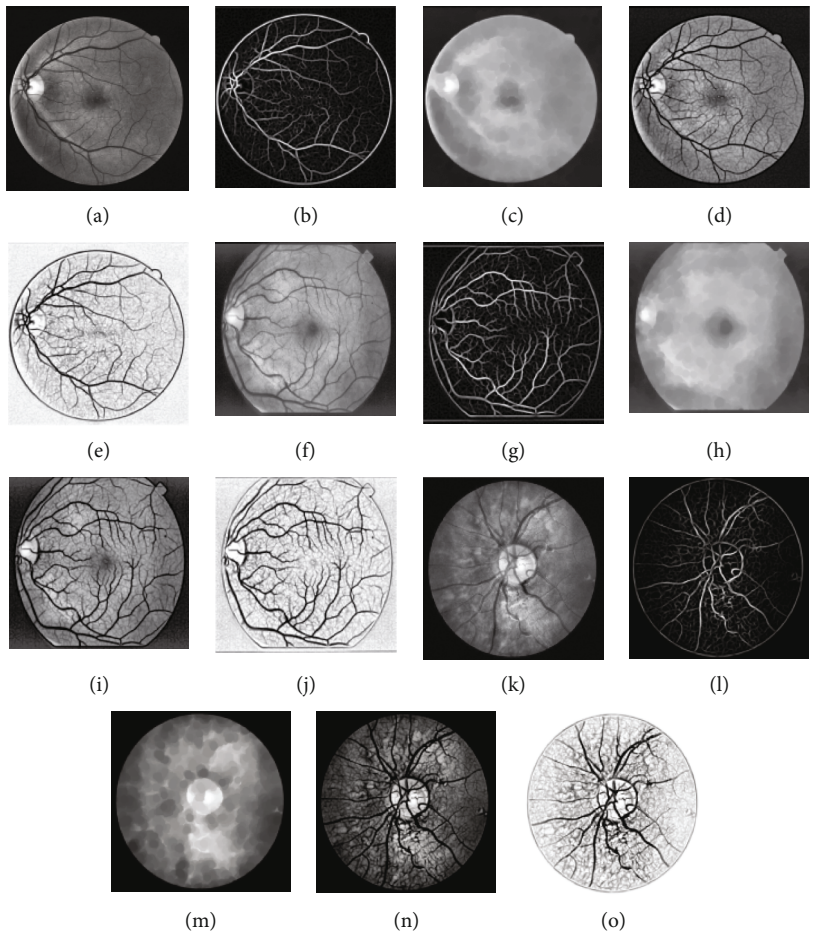


Figure 5: (a, f, k) The grayscale images after preprocessing operation, (b, g, l) large-scale matched filtered images, background extracted by (c, h, m) close operation and (d, i, n) subtraction operation, and (e, j, o) final results. The different samples of (a–e) DRIVE, (f–j) STARE, and (k–o) CHASE_DB1.

neural network combined with a multiscale matched filter to perform multichannel blood vessel segmentation processing on the retinal fundus image

(3) We have devised a new loss function to train the improved U-Net neural network to solve pixel imbalance in the image better

The rest of this paper is organized as follows. Section 2 outlines the proposed method and datasets. The performance of the proposed method and the discussion are described in detail in Section 3. A conclusion is drawn in Section 4.

2. Materials and Methods

2.1. System Overview. The proposed algorithm consists of three steps: preprocessing datasets, training U-Net in 3 channels, and postprocessing. This algorithm's main feature extraction framework is based on the improved U-Net model, using three feature extraction channels. It is mainly to perform a whole feature extraction of the image in channel 1 so that some morphological operations are performed in the preprocessing part to reduce image artifacts and noise. On the remaining two channels, matched filters are used to extract retinal vessels of different scales, and then, the improved U-Net model is used to extract features, and the OR-type operator is used to fuse the final output image. Experimental results verify that the image processed by multichannel matched filtering is better than the unprocessed image. The overall flowchart is shown in Figure 1.

2.2. Datasets. To verify the effectiveness of the algorithm in this paper, this paper chooses three commonly used public datasets for training and testing: DRIVE, STARE, and CHASE_DB1 datasets. These datasets include a wide range of challenging images. The DRIVE contains 40 color retinal fundus images divided into a training set and a testing set. The plane resolution of DRIVE is 565×584. The STARE contains 20 color retinal fundus images with a resolution of 605×700 pixels. Unlike the DRIVE, this dataset does not have a training set and a testing set. The CHASE_DB1 contains 28 color retinal fundus images with a resolution of 960×999 pixels, and the training set and testing set are also not divided. Each image in these three datasets has a label of retinal blood vessel image segmented manually by two professional physicians. We randomly selected 5 images in the STARE dataset as test images (im0002, im0077, im0163,

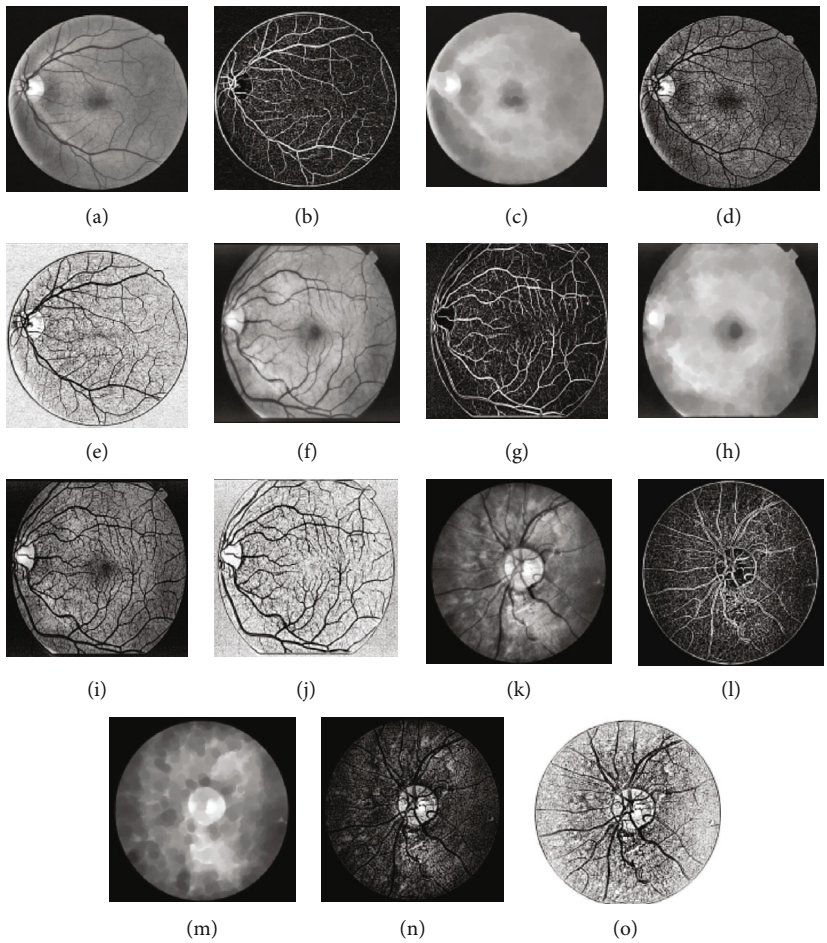


Figure 6: (a, f, k) The grayscale images after preprocessing operation, (b, g, l) small-scale matched filtered images, background extracted by (c, h, m) close operation and (d, i, n) subtraction operation, and (e, j, o) final results. The different samples of (a–e) DRIVE, (f–j) STARE, and (k–o) CHASE_DB1.

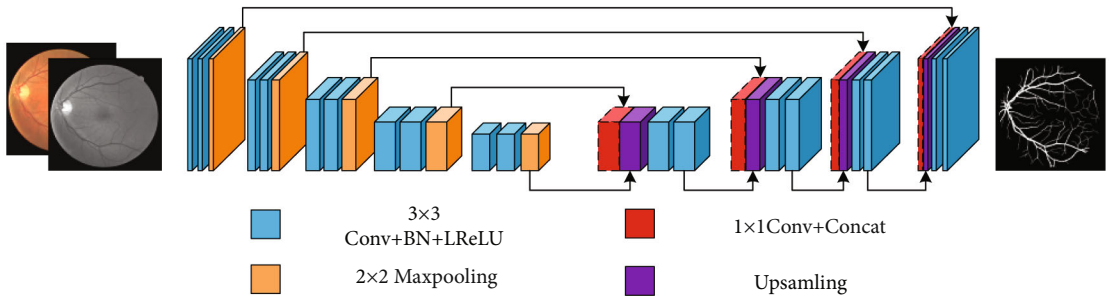


Figure 7: U-Net model architecture.

im0255, and im0291), and the remaining 15 images were set as the training set. In CHASE_DB1, we select the last 8 images as the test set and the remaining 20 images as the training set. Note that mask images of STARE and CHASE_DB1 are not available, so we extracted the green channel of the images and then used some morphological algorithms and threshold algorithm to obtain the mask images.

2.3. Preprocessing. In this paper, the green channel is selected as the input image of the preprocessing part. This is because the retinal blood vessels presented by the green channel have better contrast with the background compared with the red channel and the blue channel [20, 21], as shown in Figure 2.

It can be seen from Figure 2 that the appearance of blood vessels on the green channel of the color image consists of

Input: Train images X, ground truth G
Input: Initial epochs $E \longleftarrow$ 30, batch size $\longleftarrow$ 1, learning rate lr$\longleftarrow$ 0.01
Input: Initialize best loss bl$\longleftarrow$ Inf
Output: Predicted images P, U-Net parameter
1. $X_{\text{pre}} \longleftarrow$ preprocessing (X)
2. $X_{\text{enh}} \longleftarrow$ enhancement (X_{pre})
3. **for** $e \longleftarrow$ 0 to E **do**
4. **if** $e = 1/3 * E$ **then**
5. lr$\longleftarrow 0.1 *$ lr
6. **else if** e $= 2/3 *$ E **then**
7. lr$\longleftarrow 0.1 *$ lr
8. **end if**
9. $N \longleftarrow$ compute the number of train images X
10. par$\longleftarrow$ initial parameter of U-Net
11. **while** $n \longleftarrow 1 < N + 1$ **do**
12. $Y^n \longleftarrow$ Unet $(X_{\text{enh}}{}^n$, par)
13. loss$^n \longleftarrow$ Dice$(Y^n, \text{G}^n) + \lambda$Cross_entropy$(Y^n, \text{G}^n)$
14. **if** loss$^n <$ bl **then**
15. bl $\longleftarrow$ loss^n
16. par $\longleftarrow$ new par
17. **end if**
18. par $\longleftarrow$ SGD(par, lr)
19. **end while**
20. **end for**
21. $P \longleftarrow$ Unet$(X_{\text{enh}}$, par)
22. **return** P, par

Algorithm 1: Training of U-Net with dynamic learning rate.

more information compared to that on the red and blue channel images, but the overall image is still dark, and the contrast is not obvious. In order to improve this situation, adaptive histogram threshold processing (CLAHE) [22] and gamma transformation are performed on the extracted green channel grayscale image, as shown in Figure 3. In this part of the process, CLAHE is used to enhance the contrast between the nonvessels and blood vessels, and gamma transformation is used to adjust and reduce the background noise in the image. We can see Tables 1–3 in Supplementary Materials for a comprehensive comparison of blood vessel enhanced algorithms, and these data can prove that the CLAHE method improves the general performance of the proposed method.

2.4. Multichannel Feature Extraction

2.4.1. Channel 1. In order to retain all the blood vessel feature information of the image as much as possible, some morphological operations are used in channel 1 to remove background noise, and then, the U-Net network is used for feature extraction. For the artifacts caused by uneven illumination in the image and nonvascular structures, we use the morphological closing operation algorithm to estimate the background and then perform the result using the mathematical operation shown in equation (1).

It can be seen intuitively from Figure 4 that the brighter video disc structure in the original image is removed, and most of the artifacts are also processed.

$$\begin{cases} g(x,y) = 255 - \left(I_{\text{close}}(x,y) - I(x,y) + \dfrac{1}{m*n} \displaystyle\sum_{x=1}^{m} \sum_{y=1}^{n} I_{\text{close}}(x,y) \right), \\ f(x,y) = \dfrac{255}{\max\left(g(x,y)\right) - \min\left(g(x,y)\right)} * \left| g(x,y) - \min\left(g(x,y)\right) \right|, \end{cases} \tag{1}$$

where $f(x,y)$ is the processed image and $I_{\text{close}}(x,y)$ is the image after a morphological closing operation. We select disk type structuring elements for the closing operator having a radius of eleven pixels. $I(x,y)$ is the original image; m and n are the image pixel size.

2.4.2. Channel 2. By analyzing the gray image of retinal blood vessels, it can be found that the cross-sectional gray intensity of blood vessels is distributed in an inverted Gaussian curve, the gray value of the center line of the blood vessel is low, and the gray value at the edge of the blood vessel is high [5]. Aiming at this remarkable feature of retinal blood vessel images, Chaudhuri et al. [23] designed a Gaussian matched filter and used its distribution to simulate the grayscale intensity distribution of blood vessel cross sections and filter the blood vessels in sections. In this paper, the matched filters are used in channel 2 and channel 3 to separately enhance and extract the large and small blood vessels to realize the comprehensive segmentation of retinal blood vessels.

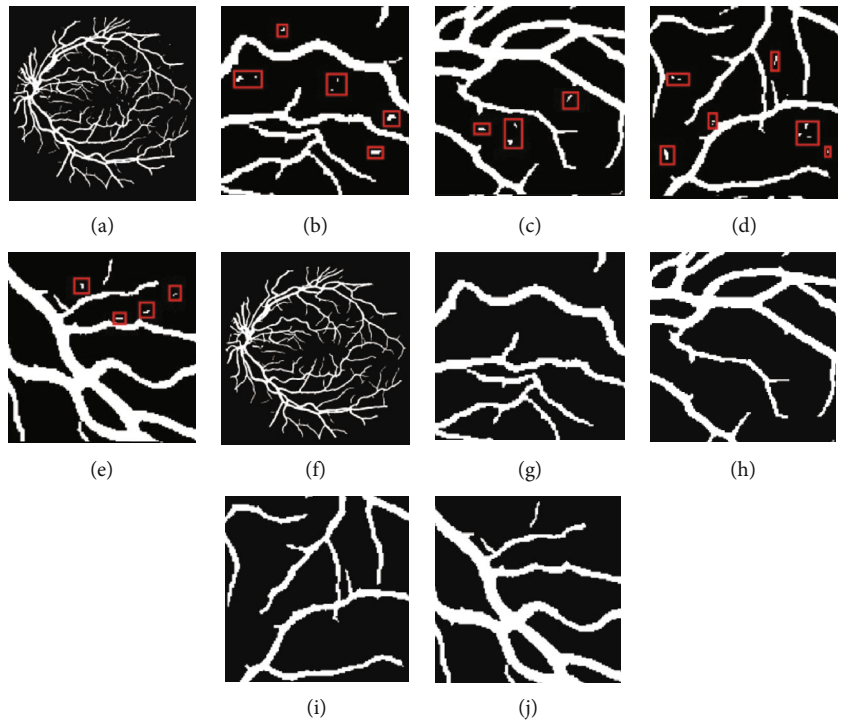


Figure 8: Comparison of postprocessing: (a–e) segmentation image without postprocessing; (f–j) segmentation image applying postprocessing.

Define the two-dimensional Gaussian kernel function as

$$K(x,y) = -e^{-\left(x^2/2s^2\right)}, \quad |y| \le \frac{l}{2}, \tag{2}$$

where s is the width of the Gaussian kernel and l is the length of the Gaussian kernel. The blood vessel starts from the center of the optic disc and extends in multiple directions. Rotating the Gaussian kernel is used to filter the multidirectional blood vessels.

Assuming that $p(x,y)$ is a discrete point in the kernel function, the rotation matrix is

$$g_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix}. \tag{3}$$

$\theta_i (0 \le \theta_i \le p)$ is the angle of the i-th kernel function, and the coordinate value of $p(x,y)$ after rotation is $\bar{p}_i = (u,v)$; then, the i-th template kernel function is

$$K_i(x,y) = -e^{-\left(u^2/2s^2\right)}, \quad \forall \bar{p}_i \in N, \tag{4}$$

where N is the template field, and the value range is

$$N = \left\{ (u,v), |u| \le 3s, |v| \le \frac{l}{2} \right\}. \tag{5}$$

In actual algorithm applications, it is often necessary to consider the mean value of the correlation coefficient of the template filter, as shown in

$$m_i = \sum_{\bar{p}_i \in N} \frac{K_i(x,y)}{A}. \tag{6}$$

Among them, A represents the number of points in the template area. So, the final template kernel function is

$$K_i'(x,y) = K_i(x,y) - m_i, \quad \forall \bar{p}_i \in N. \tag{7}$$

This paper improves and optimizes the dependence of Gaussian matched filter response on a vessel diameter. The image enhancement result using large-scale Gaussian matched filtering in channel 2 is shown in Figure 5, where the parameters are set to $l = 10.8$, $s = 1.9$, and 8 directions which means $i = [1, 2, \cdots, 8]$ in equation (3). It can be seen from the image that the algorithm has a better segmentation effect for thicker blood vessels and strong antinoise, but it has a poor segmentation effect on small blood vessels, and there is a problem that the smaller blood vessels cannot be distinguished from the background, and the blood vessels are easily broken. In order to solve this problem, this paper proposes an improved method based on the black hat algorithm, which can effectively reduce the influence of background noise by subtracting the original image before matching filter processing and the obtained image after processing to enhance the characteristics of blood vessels. We performed a series of processing transformations as shown in equations (8) and (9) on the images processed by large-scale matched filtering. We call

Table 1: The parameters of the U-Net architecture.

Block name	Layer name	Image size	Parameters
DoubleConv	Conv (ksize = 3, pad = 1)		$(3 * 3 * C_1 + 1) * C_2$
	BN + LReLU		$2 * C_2$
	Conv (ksize = 3, pad = 1)		$(3 * 3 * C_2 + 1) * C_2$
	BN + LReLU		$2 * C_2$
			$9 * (C_1 + C_2) * C_2 + 6 * C_2$
Input		$1 \times 576 \times 576$	0
Encoder block_1	DoubleConv_1	$64 \times 576 \times 576$	$9 * (1 + 64) * 64 + 6 * 64 = 37824$
	Maxpooling (ksize = 2)	$64 \times 576 \times 576$	0
Encoder block_2	DoubleConv_2	$128 \times 288 \times 288$	$9 * (64 + 128) * 128 + 6 * 128 = 221952$
	Maxpooling (ksize = 2)	$128 \times 288 \times 288$	0
Encoder block_3	DoubleConv_3	$256 \times 144 \times 144$	$9 * (128 + 256) * 256 + 6 * 256 = 886272$
	Maxpooling (ksize = 2)	$256 \times 144 \times 144$	0
Encoder block_4	DoubleConv_4	$512 \times 72 \times 72$	$9 * (256 + 512) * 512 + 6 * 512 = 3542016$
	Maxpooling (ksize = 2)	$512 \times 72 \times 72$	0
Encoder block_5	DoubleConv_5	$512 \times 36 \times 36$	$9 * (512 + 512) * 512 + 6 * 512 = 4721664$
	Maxpooling (ksize = 2)	$512 \times 36 \times 36$	0
Decoder block_1	Upsampling (bilinear)	$512 \times 72 \times 72$	0
	Concat	$1024 \times 72 \times 72$	0
	DoubleConv_6	$256 \times 72 \times 72$	$9 * (1024 + 256) * 256 + 6 * 256 = 2950656$
Decoder block_2	Upsampling (bilinear)	$256 \times 144 \times 144$	0
	Concat	$512 \times 144 \times 144$	0
	DoubleConv_7	$128 \times 144 \times 144$	$9 * (512 + 128) * 128 + 6 * 128 = 738048$
Decoder block_3	Upsampling (bilinear)	$128 \times 288 \times 288$	0
	Concat	$256 \times 288 \times 288$	0
	DoubleConv_8	$64 \times 288 \times 288$	$9 * (256 + 64) * 64 + 6 * 64 = 184704$
Decoder block_4	Upsampling (bilinear)	$64 \times 576 \times 576$	0
	Concat	$128 \times 576 \times 576$	0
	DoubleConv_9	$64 \times 576 \times 576$	$9 * (128 + 64) * 64 + 6 * 64 = 110976$
Output	Conv (ksize = 1)	$1 \times 576 \times 576$	$1 * 1 * 64 + 1 = 65$

this algorithm black hat2.

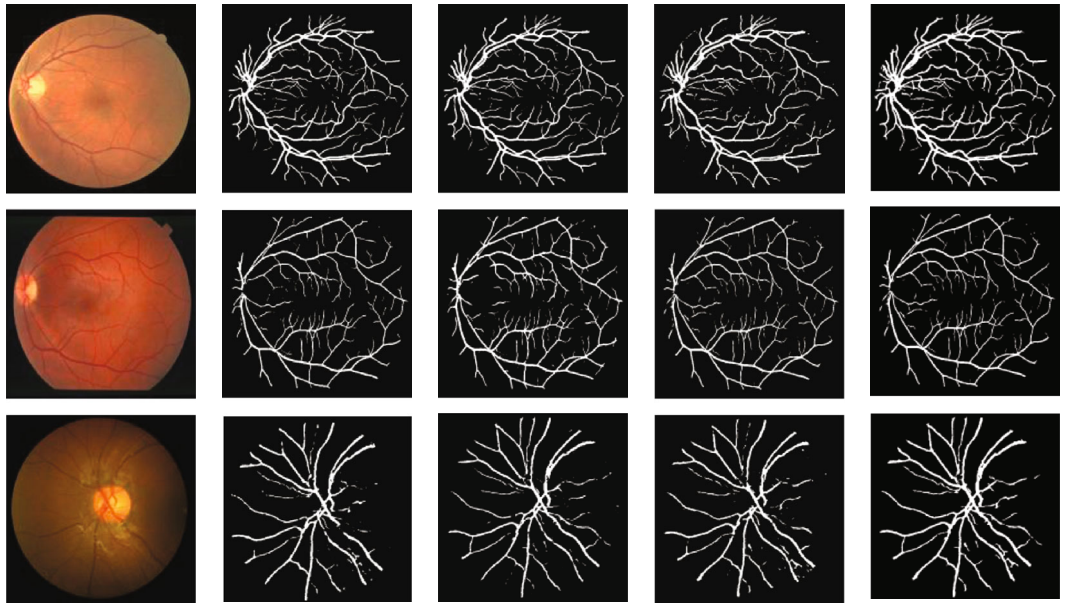
$$B_{\text{hat}}(f) = (f(x, y) \bullet b(u, v)) - f(x, y), \tag{8}$$

$$g(x, y) = 255 - f(x, y) - 2 * B_{\text{hat}}(f), \tag{9}$$

where $\bullet$ is the morphological closing operation and $b(u, v)$ is disk type structuring element, $B_{\text{hat}}(f)$ is the black hat transformation, $f(x, y)$ is the original image, and $g(x, y)$ is the final processed image.

2.4.3. Channel 3. This paper uses a small-scale Gaussian matched filter to enhance the image of small blood vessels, as shown in Figure 6. After many experiments, the parameters of the matched filter are set as $l = 5$, $s = 0.1$, and 18 directions which means $i = [1, 2, \cdots, 18]$ in equation (3). Using small-scale filters can effectively enhance the small blood vessels in the image, but at the same time, it also enhances much striped noise in the image, and the enhancing effect on the thick blood vessels with central reflection is poor. To reduce the background noise, the black hat2 algorithm used in channel 2 is also used to remove the background in channel 3.

2.5. U-Net Model. In image semantic segmentation using deep learning, the U-Net network model is the most widely used, which is improved based on the classic full convolutional network (FCN) [24]. U-Net is an image-to-image pixel-level classification network, and its network structure is apparent, as shown in Figure 7. U-Net is different from other standard segmentation networks: U-Net uses an entirely different feature fusion method—splicing. U-Net stitches the features together in the channel dimension. This



(a) Retinal image (b) Channel 1 (c) Channel 2 (d) Channel 3 (e) Fusion result

Figure 9: Performance of each channel's segmentation result.

method fuses the in-depth features extracted from the image with the shallow features to form thicker features, while the fusion operation of FCN only uses corresponding point addition and does not obtain thicker features.

Unlike the structure in the original literature [24], this paper sets the padding value of 1 in each layer's convolution operation, and the convolution kernel size is $3 * 3$. The purpose is to ensure that the output and input image sizes are consistent and avoid the size increasing operation in the output layer. It is essentially a binary classification operation in the output layer of U-Net. We use an adaptive threshold segmentation algorithm for processing in this paper. The idea of this algorithm is not to calculate the global image threshold but to calculate the local threshold according to different areas of the image, so for different areas of the image, the algorithm can adaptively calculate different thresholds and perform binary segmentation. The specific calculation process is shown in

$$T = -b + \frac{1}{(2m+1)\times(2n+1)}\sum_{i=0}^{n}\sum_{j=0}^{m} g(x \pm i, y \pm j), \quad (10)$$

where b is the fixed parameter, $(2\,m+1)\times(2n+1)$ is the area, and T is the area's threshold.

This paper proposes a new loss function that combines the Dice coefficient with the two-class cross-entropy loss function. The Dice coefficient is widely used in the evaluation of image segmentation. In order to facilitate the formation of the minimized loss function, as shown in

$$L_{\text{dice}} = 1 - \frac{2|X \cap Y|}{|X| + |Y|}, \quad (11)$$

where $X \cap Y$ represents the common elements of the prediction graph and the label graph, X and Y represent the number of elements of the prediction graph and the label. In order to facilitate the calculation, approximate $|X \cap Y|$ as the dot product between the predicted probability map and the label, and add the elements in the result. $|X|$ and $|Y|$ are quantified by summing the squares of each element. As shown in

$$L_{\text{dice}} = 1 - \frac{2\sum_i^N p(k,i)q(k,i)}{\sum_i^N p^2(k,i) + \sum_i^N q^2(k,i)}, \quad (12)$$

where N is the number of pixels, $p(k,i) \in [0,1]$ and $q(k,i) \in [0,1]$ are the predicted probabilities and true labels of the pixel belonging to category k.

The cross-entropy loss function used to optimize the network is shown as

$$L_r = -\sum_i^N \left[\left(1 - \frac{\text{TP}}{N_{\text{p}}}\right) y \log(p) + \left(1 - \frac{\text{TN}}{N_{\text{n}}}\right)(1-y)\log(1-p)\right], \quad (13)$$

where TP and TN are the numbers of true positive and true negative pixels, respectively; N_{p} and N_{n} are the numbers of segmented pixels and nonsegmented pixels, respectively; y is the label value ($y = 1$, segmentation target; $y = 0$, background); and p is the predicted probability value of the pixel.

Table 2: Segmentation results of improvements on DRIVE.

Channel	Se	Sp	ACC	AUC
Channel 1	0.8174	0.9768	0.9626	0.8971
Channel 2	0.8008	0.9741	0.9587	0.8875
Channel 3	0.8113	0.9748	0.9633	0.8931
3 channels of fusion	0.8745	0.9624	0.9546	0.9185

Table 3: Comparison of the proposed method with other methods on the DRIVE dataset.

Method	Se	Sp	ACC	AUC
Khan et al. (2016) [9]	0.7373	0.9670	0.9501	0.8522
Khan et al. (2016) [29]	0.780	0.972	0.952	0.876
Soomro et al. (2017) [11]	0.746	0.917	0.946	0.8315
Ngo and Han (2017) [12]	0.7464	0.9836	0.9533	0.8650
Biswal et al. (2017) [30]	0.71	0.97	0.95	0.84
Yan et al. (2018) [13]	0.7653	0.9818	0.9542	0.8736
Oliveira et al. (2018) [15]	0.8039	0.9804	0.9576	0.8922
Wang et al. (2019) [25]	0.7648	0.9817	0.9541	0.8733
Guo et al. (2019) [31]	0.7800	0.9806	0.9551	0.8803
Feng et al. (2019) [14]	0.7625	0.9809	0.9528	0.8717
Ribeiro et al. (2019) [32]	0.7880	0.9819	0.9569	0.8850
Dharmawan et al. (2019) [18]	0.8314	0.9726	—	0.902
Saroj et al. (2020) [5]	0.7307	0.9761	0.9544	0.8534
Dash and Senapati (2020) [33]	0.7403	0.9905	0.9661	0.8654
Biswas et al. (2020) [16]	0.7823	0.9814	0.9561	0.8819
Budak et al. (2020) [34]	0.7439	0.9900	0.9685	0.8670
2nd human observer	0.7760	0.9724	0.9472	0.8742
Proposed method	0.8745	0.9624	0.9546	0.9185

Table 4: Comparison of the proposed method with other methods on the STARE dataset.

Method	Se	Sp	ACC	AUC
Khan et al. (2016) [9]	0.7359	0.9708	0.9502	0.8534
Khan et al. (2016) [35]	0.7728	0.9649	0.9518	0.8689
Khan et al. (2017) [36]	0.778	0.966	0.951	0.872
Soomro et al. (2017) [11]	0.748	0.922	0.948	0.835
Biswal et al. (2017) [30]	0.70	0.97	0.95	0.835
BahadarKhan et al. (2017) [37]	0.758	0.963	0.946	0.861
Yan et al. (2018) [13]	0.7581	0.9846	0.9612	0.8714
Oliveira et al. (2018) [15]	0.8315	0.9858	0.9694	0.9087
Wang et al. (2019) [25]	0.7523	0.9885	0.9640	0.8704
Guo et al. (2019) [31]	0.8201	0.9828	0.9660	0.9015
Feng et al. (2019) [14]	0.7709	0.9848	0.9633	0.8779
Dharmawan et al. (2019) [18]	0.7924	0.9827	—	0.8876
Saroj et al. (2020) [5]	0.7278	0.9724	0.9509	0.8501
Tamim et al. (2020) [38]	0.7806	0.9825	0.9632	0.8816
2nd human observer	0.8952	0.9384	0.9349	0.9168
Proposed method	0.8903	0.9744	0.9699	0.9323

A coefficient λ is introduced to define the new loss function Loss, as shown in

$$\text{Loss} = L_{\text{dice}} + \lambda L_r. \tag{14}$$

Notably, the coefficient λ is set to 0.5 in this work, and the flowchart of U-Net is summarized in Algorithm 1.

Table 5: Comparison of the proposed method with other methods on the CHASE_DB1 dataset.

Method	Se	Sp	ACC	AUC
Biswal et al. (2017) [30]	0.76	0.97	—	0.865
Yan et al. (2018) [13]	0.7633	0.9809	0.9610	0.8721
Oliveira et al. (2018) [15]	0.7779	0.9864	0.9653	0.8822
Wang et al. (2019) [25]	0.7730	0.9792	0.9603	0.8761
Guo et al. (2019) [31]	0.7888	0.9801	0.9627	0.8845
Soomro et al. (2019) [39]	0.8020	0.968	0.891	0.885
Tamim et al. (2020) [38]	0.7585	0.9846	0.9577	0.8716
Joshua et al. (2020) [40]	0.7796	0.9864	0.9722	0.8830
2nd human observer	0.7686	0.9779	0.9560	0.8733
Proposed method	0.8916	0.9596	0.9561	0.9256

2.6. Postprocessing. In the postprocessing, since the final segmentation image merges the three segmentation images, the noise in the resulting image is also superimposed on all the noises of the three images. Noises will undoubtedly have a significant impact on the actual effect of the segmented image, so this paper addresses this issue in the final postprocessing step. In this paper, a morphological algorithm is used to calculate the size of the connected area of the image. The 8-adjacent connection method is adopted to eliminate the area with the connected area less than 25 pixels, which is to reclassify the area pixels as background. This paper selects a test image in the DRIVE dataset for experimental comparison, and the comparison images are shown in Figure 8.

2.7. Experimental Design

2.7.1. U-Net Implementation Details. The U-Net model used in this paper is slightly different from the structure in literature [24]. In order to keep the input and output image sizes of the model consistent, the convolution structure is adjusted accordingly. The specific model structure parameters are shown in Table 1.

In training, we set the epoch to 30 and the initial learning rate lr to 0.01, and then, the learning rate is set to update in a three-stage formula, as shown in

$$\text{lr} = \begin{cases} 0.01, & \text{epoch} > 10, \\ 0.001, & 10 < \text{epoch} \leq 20, \\ 0.0001, & 20 < \text{epoch} \leq 30. \end{cases} \tag{15}$$

Setting a larger learning rate at the beginning is to make the model obtain the vicinity of the optimal global parameters faster, and this operation can reduce the training time of the model. After training for a particular epoch, the learning rate needs to be reduced accordingly in order to make the parameters closer to the optimal value in subsequent updates. The stochastic gradient descent (SGD) algorithm is used in the optimization of the loss function.

2.7.2. Training Image Preparation. We randomly select 15 images from STARE and the first 20 images from CHASE_

Table 6: Segmentation results of all test images of the three datasets.

Image	ACC	Se	Sp	AUC
DRIVE				
01_test	0.946	0.928	0.947	0.938
02_test	0.952	0.914	0.956	0.935
03_test	0.955	0.817	0.970	0.894
04_test	0.959	0.868	0.968	0.918
05_test	0.958	0.838	0.971	0.904
06_test	0.958	0.811	0.973	0.892
07_test	0.954	0.851	0.964	0.907
08_test	0.958	0.820	0.971	0.896
09_test	0.959	0.849	0.969	0.909
10_test	0.957	0.863	0.965	0.914
11_test	0.945	0.870	0.952	0.911
12_test	0.958	0.875	0.966	0.920
13_test	0.953	0.859	0.963	0.911
14_test	0.954	0.901	0.959	0.930
15_test	0.951	0.917	0.954	0.935
16_test	0.954	0.889	0.961	0.925
17_test	0.958	0.845	0.968	0.907
18_test	0.954	0.913	0.958	0.935
19_test	0.954	0.937	0.956	0.946
20_test	0.955	0.925	0.957	0.941
Avg.	0.955	0.875	0.962	0.918
STARE				
im0002	0.972	0.839	0.981	0.910
im0077	0.967	0.966	0.961	0.964
im0163	0.961	0.976	0.960	0.968
im0255	0.970	0.872	0.979	0.926
im0291	0.980	0.798	0.990	0.894
Avg.	0.970	0.890	0.974	0.932
CHASE_DB1				
11L	0.946	0.937	0.947	0.942
11R	0.942	0.950	0.942	0.946
12L	0.953	0.878	0.959	0.919
12R	0.958	0.872	0.965	0.918
13L	0.958	0.884	0.963	0.923
13R	0.956	0.850	0.963	0.907
14L	0.970	0.895	0.968	0.931
14R	0.966	0.867	0.971	0.919
Avg.	0.956	0.892	0.960	0.926

DB1 as their respective training set. Due to the limited number of images in the existing dataset, to avoid the overfitting phenomenon in the model training, we perform data expansion processing on the training set of each dataset. Thanks to the translation invariance of the convolutional structure, the images in the training set in this paper were flipped horizontally and vertically and rotated 180 degrees to increase the amount of data 4 times.

2.7.3. Measuring Metrics. In order to evaluate the segmentation performance of this algorithm, we use the following metrics to perform a comprehensive evaluation of the segmentation result. These metrics are accuracy (ACC), sensitivity (Se), specificity (Sp), and AUC and calculated as follows:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP}}, \tag{16}$$

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{17}$$

$$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{18}$$

$$\mathrm{AUC} = \frac{1}{2}\left(\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} + \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}\right), \tag{19}$$

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative. Se is the sensitivity, which indicates the degree of classification of blood vessels and nonvascular pixels. In this paper, higher sensitivity indicates that more tiny blood vessels can be detected. Sp is specificity, which is used to express the ability of the algorithm to recognize nonvascular pixels. ACC is the accuracy of algorithm segmentation, reflecting the gap between the algorithm segmentation result and the natural result. AUC is the area under the ROC curve, and we adopt another calculation method to get the AUC, as shown in equation (19) [11].

Besides, we also use two other evaluation metrics to measure the effect of segmentation: MCC and CAL.

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{TP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP}) \times (\mathrm{TP} + \mathrm{FN}) \times (\mathrm{TN} + \mathrm{FP}) \times (\mathrm{TN} + \mathrm{FN})}}. \tag{20}$$

MCC is a correlation coefficient between the segmentation output of the algorithm and ground truth. It comprehensively considers TP, TN, FP, and FN, which is a relatively balanced metric. Finally, it is more suitable for an imbalanced class ratio.

CAL can be expressed as the product of C, A, and L as follows:

$$f(C, A, L) = C \times A \times L. \tag{21}$$

Suppose S and S_{G} are the segmentation result and the corresponding ground truth, respectively. These functions are defined as follows:

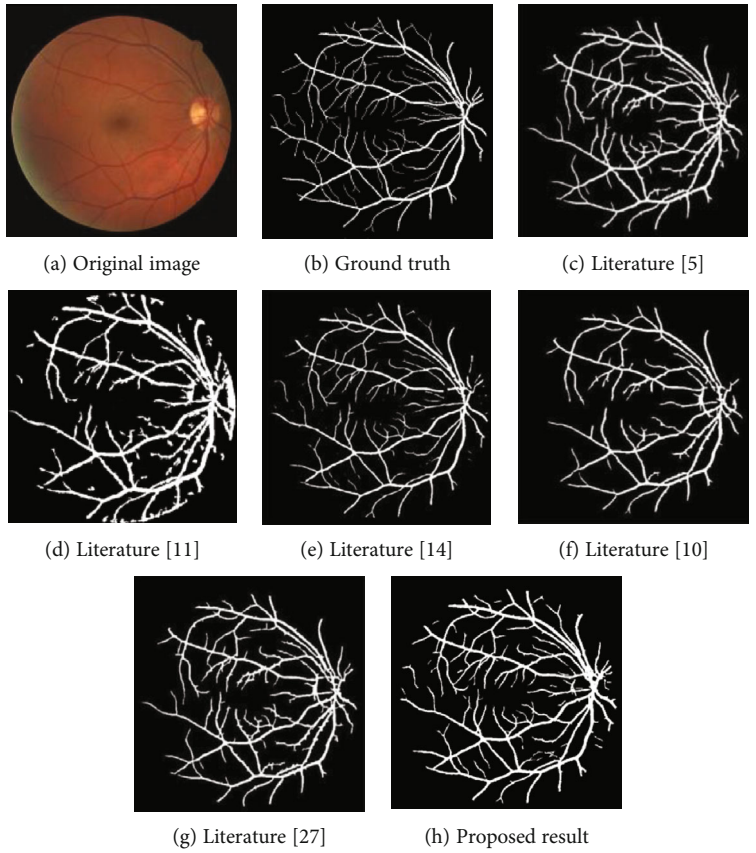
(1) Connectivity (C): it evaluates the fragmentation degree between S and S_G by comparing the number of connected components:

$$C = 1 - \min\left(1, \frac{|\#_C(S_{\mathrm{G}}) - \#_C(S)|}{\#(S_{\mathrm{G}})}\right), \tag{22}$$

where $\#_C(\bullet)$ means the number of connected components,

Table 7: MCC and CAL metrics of existing techniques on the three datasets.

Method	DRIVE		STARE		CHASE_DB1	
	MCC	CAL	MCC	CAL	MCC	CAL
Azzopardi et al. (2015) [41]	0.719	0.721	0.698	0.709	0.656	0.608
Orlando et al. (2016) [42]	0.740	0.675	0.726	0.665	0.689	0.571
Dharmawan et al. (2017) [18]	07991	0.8834	0.7959	0.8181	—	—
Yang et al. (2018) [43]	0.725	—	0.662	—	—	—
Strisciuglio et al. (2019) [44]	0.729	0.728	0.698	0.709	0.663	0.620
Khan et al. (2020) [45]	0.739	0.696	0.707	0.566	0.629	0.547
2nd human observer	0.770	0.771	0.741	0.622	0.626	0.722
Proposed method	0.756	0.796	0.796	0.837	0.566	0.733



(a) Original image (b) Ground truth (c) Literature [5]

(d) Literature [11] (e) Literature [14] (f) Literature [10]

(g) Literature [27] (h) Proposed result

Figure 10: Comparison of different methods on the DRIVE dataset.

while $\#(\bullet)$ means the number of vessel pixels in the considered binary image.

(2) Area (A): it evaluates the degree of intersecting area between S and S_G and is defined as

$$A = \frac{\#((\delta_\alpha(S) \cap S_G) \cup (\delta_\alpha(S_G) \cap S))}{\#(S \cup S_G)}, \tag{23}$$

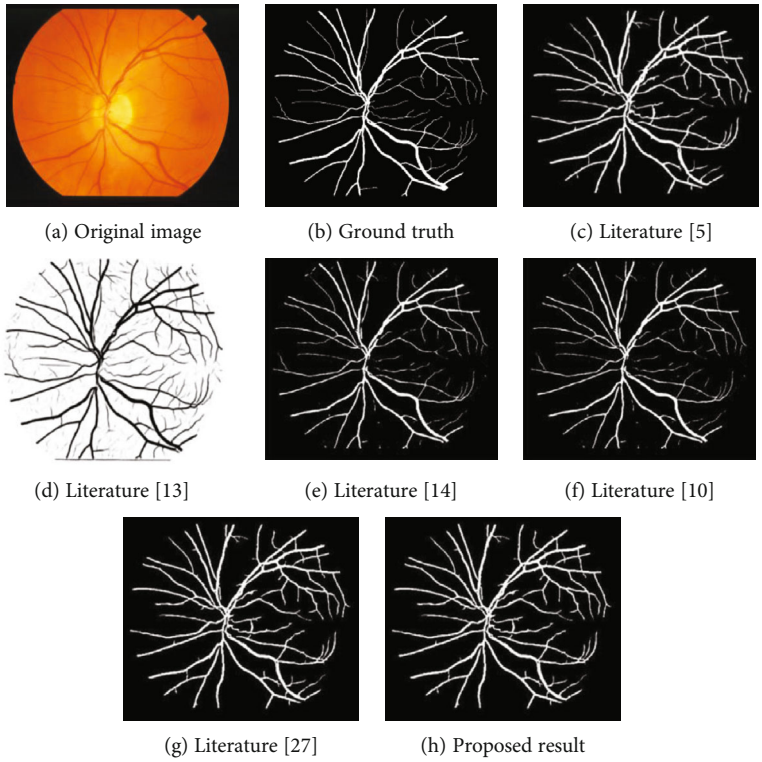
where $\delta_\alpha(\cdot)$ is a morphological dilation using a disc of α pixels in radius. We set $\alpha = 2$.

(3) Length (L): it evaluates the equivalent degree between S and S_G by computing the total length:

$$L = \frac{\#\left(\left(\varphi(S) \cap \delta_\beta(S_G)\right) \cup \left(\left(\delta_\beta(S) \cap \varphi(S_G)\right)\right)\right)}{\#(\varphi(S) \cup \varphi(S_G))}, \tag{24}$$

where $\varphi(\cdot)$ is the homotopic skeletonization and $\delta_\beta(\bullet)$ is a morphological dilation with a disc of β pixel in radius. We set $\beta = 2$.

According to [26], the CAL metric is essential to quantify thick and thin vessels more equally.



(a) Original image (b) Ground truth (c) Literature [5]

(d) Literature [13] (e) Literature [14] (f) Literature [10]

(g) Literature [27] (h) Proposed result

Figure 11: Comparison of different methods on the STARE dataset.

3. Results and Discussion

As shown in Figure 9, one test image is selected from each of the three datasets to display the segmentation results of each channel and the fusion results. It can be seen that some of the intermittent blood vessels of each channel are reconnected after fusion, and the number of small blood vessels in the fusion map is significantly higher than that of each channel segmentation map.

The DRIVE dataset is selected as the experimental object and compares the three channels' metric data in this paper. The results show that the overall fusion effect of the three channels is better than the segmentation results of every single channel; in particular, the sensitivity has been dramatically improved, as shown in Table 2.

To illustrate this paper's segmentation effect, we list various metrics on the DRIVE, STARE, and CHASE_DB1 datasets of different papers in recent years in Tables 3–5. It can be seen that the algorithm in this paper is superior to most similar papers in sensitivity and AUC metrics. To have a more comprehensive understanding of the overall segmentation effect of the test set, we show the relevant indicators of the prediction results of all test set images in Table 6. The other essential metrics are MCC and CAL, and they achieved by the proposed method has been contrasted with existing segmentation techniques on the DRIVE, STARE, and CHASE_DB1 datasets shown in Table 7.

We selected image 19_test from the test set of the DRIVE dataset to display the segmentation results, as shown in Figure 10. Literature [5, 27] segmented some small blood vessels, but it is still slightly insufficient compared to this paper's segmentation diagram. Literature [10] lacks many details, and the small blood vessels are not segmented. The segmentation result of literature [11] contains a lot of edge noise, and there are many intermittent blood vessels. Compared with the existing segmentation methods, the segmentation results in this paper have a good performance in terms of the integrity of the whole blood vessels and the segmentation of small blood vessels.

As shown in Figure 11, we select the test results of the image im0163 in the STARE dataset for comparison. It can be shown that the segmentation results of this paper are similar to those of literature [13, 14], but the background noise in literature [13] is not eliminated. Compared with literature [5, 10, 27], the algorithm in this paper illuminates the optic disc structure in the original image as much as possible in the preprocessing part, so the problem that is incorrectly dividing part of the optic disc structure into blood vessels like these papers did not appear in the final segmentation result.

The CHASE_DB1 dataset is not used in most of the papers about retinal blood vessel segmentation. One of the reasons is that the dataset contains half of the abnormal images, which may cause some interference to the trained segmentation model. Meanwhile, this dataset is also a new and challenging dataset compared to the classic DRIVE and STARE datasets. We selected four images image_12R, image_13L, image_13R, and image_14L from the test set of the CHASE_DB1 dataset to compare the segmentation results in order to verify the generalizability of the proposed algorithm, as shown in Figure 12. The segmentation result of the algorithm in literature [19] has much noise, and some blood vessels are not effectively separated. Literature [28]

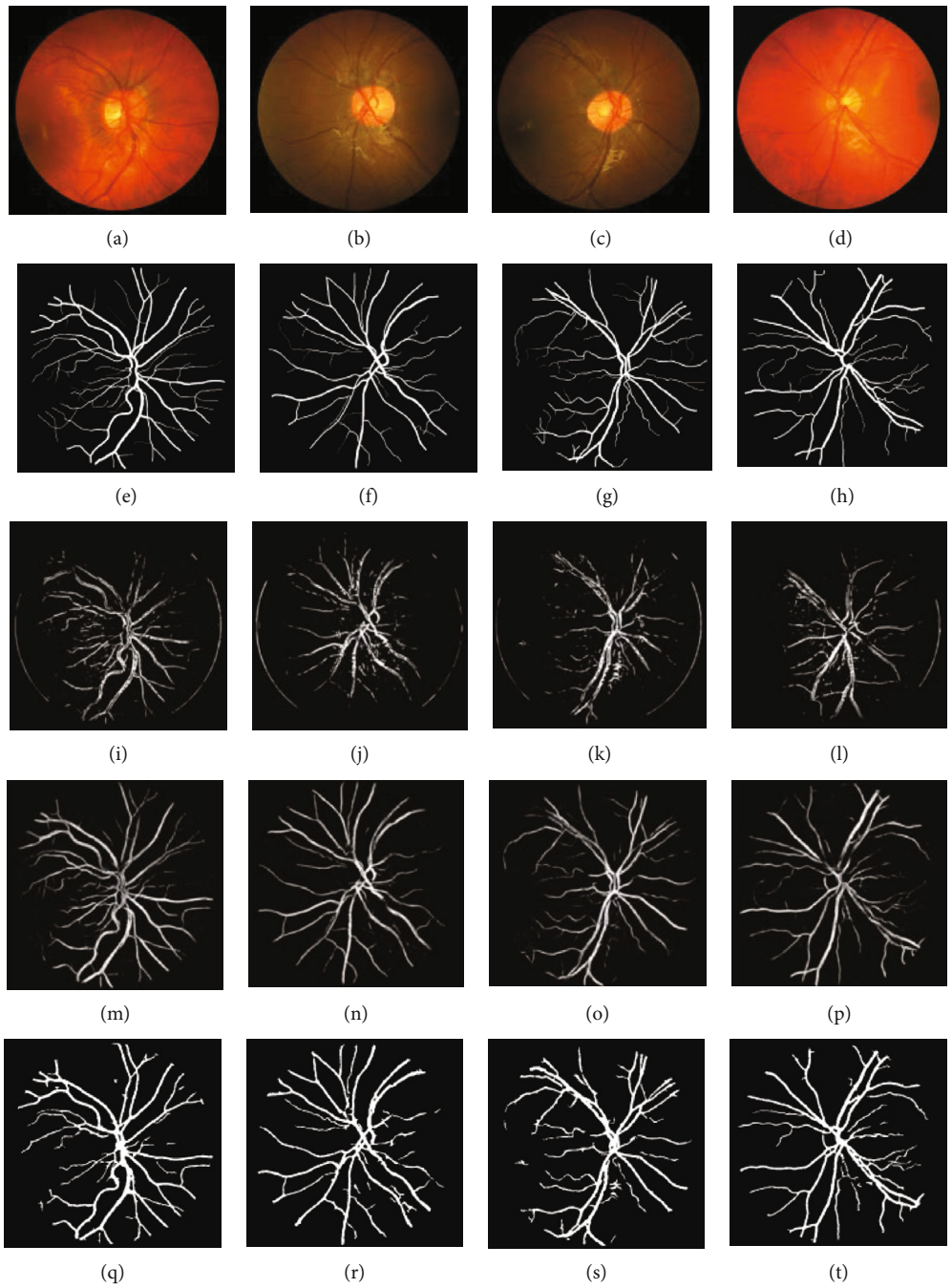


Figure 12: (a, e, i, m, q) Image_12R, (b, f, j, n, r) image_13L, (c, g, k, o, s) image_13R, and (d, h, l, p, t) image_14L from the CHASE_DB1 dataset. (a–d) Original images, (e–h) ground truth, (i–l) literature [19], (m–p) literature [28], and (q–t) proposed segmentation images.

does an excellent job in the segmentation of small blood vessels, but there is a problem that some blood vessels are not connected. Due to the postprocessing in this paper, the segmentation result on this dataset contains less noise and guarantees the continuity of most blood vessels. However, compared with the manual label, some tiny blood vessels cannot be completely segmented from the image background.

The source codes of the proposed framework have been running on the PC (Intel Core i5-6300HQ CPU, 2.30 GHz, 12.0 GB RAM, NVIDIA GTX 950M GPU). DRIVE, STARE, and CHASE_DB1 have spent 11.3 h, 7.1 h, and 16.4 h on training separately in each channel. The average testing time of test images was 1.34 s. Table 8 shows the parameter comparison of the proposed method with other methods based on U-Net, which can help us compare the framework complexity of different methods. Note that the parameters are not equal to the training time because some methods use slices of a train image as input of the network. For example, literature [19] has 42421 slices as the training set, which means it needs more time to train the network.

Table 8: Network comparison of the proposed method with other methods based on U-Net.

Method	Input size	Epoch	Training images	Parameters
Ronneberger et al. (2015) [24]	572 ∗ 572 ∗ 1	N/A	N/A	28.94 M
Yan et al. (2018) [13]	128 ∗ 128 ∗ 1	Over 30	26052 (4)	30.96 M
Jiang et al. (2018) [19]	500 ∗ 500 ∗ 3	30	42421 (4)	58.31 M
Soomro et al. (2019) [39]	Original image	N/A	3959 (4)	4.71 M
Joshua et al. (2020) [40]	512 ∗ 512 ∗ 3	50	113 (4)	0.64 M
Zhang et al. (2020) [17]	256 ∗ 256 ∗ 1	120	116 (3)	3.86 M
Budak et al. (2020) [34]	48 ∗ 48 ∗ 3	Over 30	Over 45840 (2)	0.97 M
Proposed method	576 ∗ 576 ∗ 1	30	220 (3)	13.39 M

$x\,(y)$ means x training images of y datasets.

4. Conclusion

This paper proposes a new retinal blood vessel segmentation method, which combines a multiscale matched filter with a U-Net neural network model of deep learning. First of all, we use an improved morphological image algorithm to effectively reduce the impact of image background in feature extraction. Additionally, in order to avoid ignoring the characteristics of small blood vessels, this paper performs multichannel feature extraction and segmentation on retinal blood vessel images. Finally, the segmented images of the three channels are merged, and various characteristics of retinal blood vessels are obtained as much as possible. In the training of the U-Net model, we used the loss function weighted by the Dice coefficient and the binary cross-entropy to solve the image pixel imbalance problem. The algorithm of this paper is tested on the existing public datasets DRIVE, START, and CHASE_DB1. The experimental results show that there is better performance in four metrics compared with similar papers. The average sensitivity of the algorithm in this paper reached 0.8745, 0.8903, and 0.8916 on the DRIVE, STARE, and CHASE_DB1 datasets, respectively. This result is nearly 0.1 higher than the average sensitivity of other papers. The improvement of the sensitivity metric also reflects that the algorithm in this paper has a good performance in extracting small blood vessels. The focus of this paper is to combine the advantages of unsupervised algorithms and supervised algorithms. We did not make too many improvements to the U-Net network. Therefore, how to prune the deep learning network model structure will be an interesting research direction in the future.

Data Availability

The three public open-source datasets used to support this study are available at http://www.isi.uu.nl/Research/Databases/DRIVE/, http://cecas.clemson.edu/~ahoover/stare/, and https://blogs.kingston.ac.uk/retinal/chasedb1/.

Conflicts of Interest

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

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant nos. 62071161 and 62001149, Key R&D Projects of Shandong under Grant no. 2019JZZY021005, Natural Science Foundation of Shandong Province under Grant no. ZR2020MF067, and Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province.

Supplementary Materials

Table 1: channel 1 results of DRIVE test images. Table 2: channel 1 results of STARE test images. Table 3: channel 1 results of CHASE_DB1 test images. *(Supplementary Materials)*

References

[1] X. You, Q. Peng, Y. Yuan, Y. M. Cheung, and J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi- supervised approach," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2314–2324, 2011.

[2] F. Jusoh, H. Haron, R. Ibrahim, and M. Azemin, *An Overview of Retinal Blood Vessels Segmentation//Advanced Computer and Communication Engineering Technology*, Springer, Cham, 2016.

[3] R. GeethaRamani and L. Balasubramanian, "Retinal blood vessel segmentation employing image processing and data mining techniques for computerized retinal image analysis," *Biocybernetics and Biomedical Engineering*, vol. 36, no. 1, pp. 102–118, 2016.

[4] K. B. Khan, A. A. Khaliq, A. Jalil et al., "A review of retinal blood vessels extraction techniques: challenges, taxonomy, and future trends," *Pattern Analysis and Applications*, vol. 22, no. 3, pp. 767–802, 2019.

[5] S. K. Saroj, R. Kumar, and N. P. Singh, "Frechet PDF based matched filter approach for retinal blood vessels segmentation," *Computer Methods and Programs in Biomedicine*, vol. 194, article 105490, 2020.

[6] R. Ghoshal, A. Saha, and S. Das, "An improved vessel extraction scheme from retinal fundus images," *Multimedia Tools and Applications*, vol. 78, no. 18, pp. 25221–25239, 2019.

[7] N. P. Singh and R. Srivastava, "Extraction of retinal blood vessels by using an extended matched filter based on second derivative of Gaussian," *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, vol. 89, no. 2, pp. 269–277, 2019.

[8] N. Memari, A. R. Ramli, M. I. B. Saripan, S. Mashohor, and M. Moghbel, "Retinal blood vessel segmentation by using matched filtering and fuzzy c-means clustering with integrated level set method for diabetic retinopathy assessment," *Journal of Medical and Biological Engineering*, vol. 39, no. 5, pp. 713–731, 2019.

[9] M. A. U. Khan, T. A. Soomro, T. M. Khan, D. G. Bailey, J. Gao, and N. Mir, "Automatic retinal vessel extraction algorithm based on contrast-sensitive schemes," in *2016 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pp. 1–5, New Zealand, 2016.

[10] N. P. Singh and R. Srivastava, "Retinal blood vessels segmentation by using Gumbel probability distribution function based matched filter," *Computer Methods and Programs in Biomedicine*, vol. 129, pp. 40–50, 2016.

[11] T. A. Soomro, A. J. Afifi, J. Gao et al., "Boosting sensitivity of a retinal vessel segmentation algorithm with convolutional neural network," in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–8, Sydney, New South Wales, Australia, 2017.

[12] L. Ngo and J. H. Han, "Multi-level deep neural network for efficient segmentation of blood vessels in fundus images," *Electronics Letters*, vol. 53, no. 16, pp. 1096–1098, 2017.

[13] Z. Yan, X. Yang, and K. T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1912–1923, 2018.

[14] S. Feng, Z. Zhuo, D. Pan, and Q. Tian, "CcNet: a cross-connected convolutional network for segmenting retinal vessels using multi-scale features," *Neurocomputing*, vol. 392, pp. 268–276, 2020.

[15] A. Oliveira, S. Pereira, and C. A. Silva, "Retinal vessel segmentation based on fully convolutional neural networks," *Expert Systems with Applications*, vol. 112, pp. 229–242, 2018.

[16] R. Biswas, A. Vasan, and S. S. Roy, "Dilated deep neural network for segmentation of retinal blood vessels in fundus images," *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, vol. 44, no. 1, pp. 505–518, 2020.

[17] Z. Zhang, C. Wu, S. Coleman, and D. Kerr, "DENSE-INception U-net for medical image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 192, article 105395, 2020.

[18] D. A. Dharmawan, D. Li, B. P. Ng, and S. Rahardja, "A new hybrid algorithm for retinal vessels segmentation on fundus images," *IEEE Access*, vol. 7, pp. 41885–41896, 2019.

[19] Z. Jiang, H. Zhang, Y. Wang, and S. B. Ko, "Retinal blood vessel segmentation using fully convolutional network with transfer learning," *Computerized Medical Imaging and Graphics*, vol. 68, pp. 1–15, 2018.

[20] D. Marín, A. Aquino, M. E. Gegundez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 146–158, 2011.

[21] M. A. Latha, N. C. Evangeline, and S. Sankaranarayanan, "Colour image segmentation of fundus blood vessels for the detection of hypertensive retinopathy," in *2018 Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII)*, pp. 206–212, Chennai, India, 2018.

[22] F. Zana and J. C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1010–1019, 2001.

[23] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on Medical Imaging*, vol. 8, no. 3, pp. 263–269, 1989.

[24] 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, Munich, Germany, 2015.

[25] X. Wang, X. Jiang, and J. Ren, "Blood vessel segmentation from fundus image by a cascade classification framework," *Pattern Recognition*, vol. 88, pp. 331–341, 2019.

[26] Z. Yan, X. Yang, and K. T. T. Cheng, "A skeletal similarity metric for quality evaluation of retinal vessel segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 4, pp. 1045–1057, 2018.

[27] N. P. Singh and R. Srivastava, "Weibull probability distribution function-based matched filter approach for retinal blood vessels segmentation," in *Advances in Computational Intelligence*, pp. 427–437, Springer, Singapore, 2017.

[28] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: a deformable network for retinal vessel segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149–162, 2019.

[29] K. B. Khan, A. A. Khaliq, M. Shahid, and S. Khan, "An efficient technique for retinal vessel segmentation and denoising using modified ISODATA and CLAHE," *International Islamic University Malaysia Engineering Journal*, vol. 17, no. 2, pp. 31–46, 2016.

[30] B. Biswal, T. Pooja, and N. Bala Subrahmanyam, "Robust retinal blood vessel segmentation using line detectors with multiple masks," *IET Image Processing*, vol. 12, no. 3, pp. 389–399, 2017.

[31] S. Guo, K. Wang, H. Kang, Y. Zhang, Y. Gao, and T. Li, "BTS-DSN: deeply supervised neural network with short connections for retinal vessel segmentation," *International Journal of Medical Informatics*, vol. 126, pp. 105–113, 2019.

[32] A. Ribeiro, A. P. Lopes, and C. A. Silva, "Ensemble learning approaches for retinal vessel segmentation," in *2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG)*, pp. 1–4, Lisbon, Portugal, 2019.

[33] S. Dash and M. R. Senapati, "Enhancing detection of retinal blood vessels by combined approach of DWT, Tyler Coye and Gamma correction," *Biomedical Signal Processing and Control*, vol. 57, article 101740, 2020.

[34] Ü. Budak, Z. Cömert, M. Çıbuk, and A. Şengür, "DCCMED-Net: densely connected and concatenated multi encoder-decoder CNNs for retinal vessel extraction from fundus images," *Medical Hypotheses*, vol. 134, article 109426, 2020.

[35] K. B. Khan, A. A. Khaliq, and M. Shahid, "B-COSFIRE filter and VLM based retinal blood vessels segmentation and denoising," in *2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, pp. 132–137, Quetta, Pakistan, 2016.

[36] K. B. Khan, A. A. Khaliq, and M. Shahid, "A novel fast GLM approach for retinal vascular segmentation and denoising," *Journal of Information Science and Engineering*, vol. 33, no. 6, pp. 1611–1627, 2017.

[37] K. BahadarKhan, A. A Khaliq, and M. Shahid, "A morphological Hessian based approach for retinal blood vessels segmentation and denoising using region based Otsu thresholding," *PloS one*, vol. 11, no. 7, article e0158996, 2016.

[38] N. Tamim, M. Elshrkawey, G. Abdel Azim, and H. Nassar, "Retinal blood vessel segmentation using hybrid features and multi-layer perceptron neural networks," *Symmetry*, vol. 12, no. 6, p. 894, 2020.

[39] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, L. Zheng, and M. Paul, "Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation," *Expert Systems with Applications*, vol. 134, pp. 36–52, 2019.

[40] A. O. Joshua, F. V. Nelwamondo, and G. Mabuza-Hocquet, "Blood vessel segmentation from fundus images using modified U-Net convolutional neural network," *Journal of Image and Graphics*, vol. 8, no. 1, pp. 21–25, 2020.

[41] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable COSFIRE filters for vessel delineation with application to retinal images," *Medical Image Analysis*, vol. 19, no. 1, pp. 46–57, 2015.

[42] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 16–27, 2017.

[43] Y. Yang, F. Shao, Z. Fu, and R. Fu, "Blood vessel segmentation of fundus images via cross-modality dictionary learning," *Applied Optics*, vol. 57, no. 25, pp. 7287–7295, 2018.

[44] N. Strisciuglio, G. Azzopardi, and N. Petkov, "Robust inhibition-augmented operator for delineation of curvilinear structures," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5852–5866, 2019.

[45] K. B. Khan, M. S. Siddique, M. Ahmad, and M. Mazzara, "A hybrid unsupervised approach for retinal vessel segmentation," *BioMed Research International*, vol. 2020, 20 pages, 2020.

Hindawi
BioMed Research International
Volume 2021, Article ID 5584004, 15 pages
https://doi.org/10.1155/2021/5584004

Research Article

Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images

Venkatesan Chandran,[1] **M. G. Sumithra**,[1] **Alagar Karthick**,[2] **Tony George**,[3] **M. Deivakani**,[4] **Balan Elakkiya**,[5] **Umashankar Subramaniam**,[6] **and S. Manoharan**[7]

[1]*Department of Electronics and Communication Engineering, KPR Institute of Engineering and Technology, Avinashi road, Coimbatore, 641407 Tamilnadu, India*
[2]*Renewable Energy Lab, Department of Electrical and Electronics Engineering, KPR Institute of Engineering and Technology, Avinashi road, Coimbatore, 641407 Tamilnadu, India*
[3]*Department of Electrical and Electronics Engineering, Adi Shankara Institute of Engineering and Technology Mattoor, Kalady, Kerala 683574, India*
[4]*Department of Electronics and Communication Engineering, PSNA College of Engineering and Technology, Dindigul, 624622 Tamilnadu, India*
[5]*Department of Electronics and Communication Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Tamilnadu 600062, India*
[6]*Department of Communications and Networks, Renewable Energy Lab, College of Engineering, Prince, Sultan University, Riyadh 12435, Saudi Arabia*
[7]*Department 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 S. Manoharan; manoharan.subramanian@ambou.edu.et

Received 5 February 2021; Revised 31 March 2021; Accepted 20 April 2021; Published 4 May 2021

Academic Editor: Changming Sun

Traditional screening of cervical cancer type classification majorly depends on the pathologist's experience, which also has less accuracy. Colposcopy is a critical component of cervical cancer prevention. In conjunction with precancer screening and treatment, colposcopy has played an essential role in lowering the incidence and mortality from cervical cancer over the last 50 years. However, due to the increase in workload, vision screening causes misdiagnosis and low diagnostic efficiency. Medical image processing using the convolutional neural network (CNN) model shows its superiority for the classification of cervical cancer type in the field of deep learning. This paper proposes two deep learning CNN architectures to detect cervical cancer using the colposcopy images; one is the VGG19 (TL) model, and the other is CYENET. In the CNN architecture, VGG19 is adopted as a transfer learning for the studies. A new model is developed and termed as the Colposcopy Ensemble Network (CYENET) to classify cervical cancers from colposcopy images automatically. The accuracy, specificity, and sensitivity are estimated for the developed model. The classification accuracy for VGG19 was 73.3%. Relatively satisfied results are obtained for VGG19 (TL). From the kappa score of the VGG19 model, we can interpret that it comes under the category of moderate classification. The experimental results show that the proposed CYENET exhibited high sensitivity, specificity, and kappa scores of 92.4%, 96.2%, and 88%, respectively. The classification accuracy of the CYENET model is improved as 92.3%, which is 19% higher than the VGG19 (TL) model.

1. Introduction

Cervical cancer is the second most deadly condition for women in the medical world following breast cancer and later believed that cervical cancer remains incurable in the later stages. Much recent progress has been made to improve the disease detection rate by using an image. Statistics by the World Health Organization (WHO) revealed that cervical cancer is the fourth most prevalent cancer globally, with a reporting rate of 5,70,000 new cases in 2018, accounting for 7.5% of all women cancer deaths [1]. Over 3,11,000 cervical cancer deaths per year were reported at around 85% in low- and intermediate-income countries, and the early diagnosis of cervical cancer offers a way of saving a life. Women with HIV are sixfold more likely to develop cervical cancer than women without HIV, and it is estimated that 5% of all cervical cancer cases are related to HIV. A variety of considerations have redefined screening effectiveness, which includes the access to equipment, consistency of screening tests, adequate supervision, and detection and treatment of lesions detected [2]. Despite severe medical and science advancements, this disease is not completely curable, mainly if diagnosed in a developing state. Prevention and screening services, therefore, play a crucial role in the fight against cervical cancer. The screening of cervical cancer follows a typical workflow: HPV testing, cytology or PAP smear testing, colposcopy, and biopsy. Several tools supported the workflow which have been created to make it more effective, practical, and inexpensive. The PAP smear image screening is mostly employed for the treatment of cervical cancer, but it requires a greater number of microscopic examinations to diagnosis of cancer and noncancer patients, and also it is time consuming and requires trained professionals, but there is a chance of missing the positive cases by using the conventional screening method. The PAP smear and HPV testing are very costly treatment, and it also provides lower sensitivity. On the other side, the colposcopy treatment is widely used in the developing countries. To overcome the shortcomings in PAP smear images and HPV testing, the colposcopy screening is used. Both cervical and other cancers are more likely to be treated in the early stage, but the lack of signs and symptoms at this stage hinders the early diagnosis. Cervical cancer deaths can be avoided by successful screening schemes and can lead to lowered sickness and impermanence [3]. In low- and middle-income nations, cervical cancer screening facilities are very sparse because of a shortage of qualified and educated health care staff and insufficient healthcare funding to fund screening systems [4].

Colposcopy is a popular surgical procedure to prevent cervical cancer. Timely identification and classification of this type of cancer may significantly improve the patient's eventual clinical care. Several works have been taken various approaches for collecting details from images in digital colposcopy. These studies' key aim is to provide health practitioners with tools during colposcopy exams irrespective of their level of competence. Previous studies have been developed in diagnosis using computer-aided systems for a range of tasks, including improvement and evaluation of image quality, regional segmentation, picture identification, identification of unstable regions and patterns, transition zone type classification (TZ) type, and cancer risk classification [5]. CAD instruments help improve the picture of cervical colposcopy and areas of concern segments and identify certain anomalies. These methods help clinicians to make diagnostic choices, but they should have adequate experience and expertise to make an appropriate diagnosis. The appearance of pathological regions may indicate such neoplasms; so in a colposcopy analysis, the detection of these lesions may be very critical. These abnormal areas include acetowhite, abnormal vascularization, mosaic areas, and punctures [6, 7]. Most literature surveys recommended a mechanism to spot irregular areas in conventional colposcopy images. Most works include inconsistent zone segmentation, including exclusion from specular reflection, segmentation of the cervix, acetowhite field segmentation [8], mosaic regions recognition, vasculature and puncture, and classification [9].

Deep learning has made significant advances in different applications such as computer vision, natural language processing, forecasting, and battery health monitoring [10]. Medical image processing, including classification, identification, segmentation, and registration, plays an essential role in disease diagnosis. Medical images such as MRI, CT, and ultrasound images and blood smear images [11], make up the vast majority of the image data processed. Deep learning's multilayer neural network perception mechanism can learn more abstract features in images and is expected to address the issues that plague conventional medical CAD systems. However, the deep learning techniques should be supported with an extensive database, especially for positive cases. To overcome this issue, many transfer learning and ensemble learning approaches are discussed in the previous work. The convolution neural network (CNN) is used to identify MI signals in an efficient computer-aided diagnosis (CAD) framework for urban healthcare in smart cities [12]. The novel feature extraction protocol followed by the genetic algorithm is proposed to detect arrhythmia to improve the performance using several tiers [13]. The structure is as follows: Section 2 discusses the related work connected with cervical screening, Section 3 elaborates the proposed architecture of CYENET to cervical screening, Section 4 interprets the results obtained out of the implementation, and Section 5 drawn the conclusion and future scope of this work.

2. Related Work

Several algorithms were utilized for machine learning, and their segmentation refining was matched to a cervical cancer classifier in which random forests showed the best output [14]. Also, robust refinement methods have been used to manage, and unattended learning approaches to the different image or superpixel patches from extracted objects methods include Adaboost detectors [15], SVM supports [16], or Gaussian mixture models [17]. A novel Markov random field segmentation based on superpixels was proposed and implemented for nonoverlapping cells [18]. The multifilter SVM is executed, and the parameters were set for the identification of cervical cells [19]. It was suggested that cervical cell classification using artificial neural networks (ANN) was built and tested with a precision of 78% [20]. Unbalanced medical evidence for the variety of cervical cancer without any parameter change was addressed using an unsupervised approach

[21]. The particle swarm optimization (PSO) with KNN membership values outperformed all other fundamental classification models [22]. The cervical cancer cell is classified using shape and texture characteristics of the segmentation and classification method and Gabor characteristics. It was found that a greater accuracy of 89% was obtained for both normal and cancer cell classification [23]. The extracted features from CNN were classified using the least square support vector machine (LSSVM) and produced more remarkable results, one of the suggested model's reference components [24]. Radial basis function- (RBF-) SVM also obtained a strong outcome and outperformed logistic regression and random forest methods [25]. Based on the features, it was found that the accuracies were ranged from 90 to 95%.

New deep architectures such as ResNet, Inception, and tree models [26] have recently shown promising results in many applications and detect cancer cells. As one of the deep learning methods, the convolutional neural networks is the commonly used technique to identify and recognize cervical cancer [27]. Early cervical cancer cell identification and classification method based on CNN's was developed to extract deep learned features from the cervical images [28]. The extreme learning machine (ELM) was used to categorize the input images. The CNN paradigm was used for fine-tuning and transfer learning. Alternatives to classifiers based on the ELM, the multilayered perceptron (MLP), and the automotive encoder (AE) were also studied. It was reported that the stacked soft-max autoencoder reported a 97.25% precision on the cervical cancer dataset [29]. It was concluded that a tentative effort was made to tackle the issue of patient risk prediction using the applications for machine learning to grow cervical cancer. The machine learning software with cervical screening was used to tackle the problem of predicting the patient's risk [30]. They concentrated on the transition of information between linear classifiers to related activities to predict the patient's risk. Since the related risk factors in the population are highly sparsely influenced, the techniques for reducing dimensionality can boost the power of predictive machine learning models [31]. However, several projects benefit from reducing dimensionality and classification by using suboptimal methods in which each part is learned separately [32]. For the efficient collection and classification of cell properties in cervical smeared images [33], a quantum hybrid- (QH-) innovative approach was combined with adaptive search capability of the quantum-behaved particle swarm optimization (QPSO) method with the intuitionist reasonableness of the standard fuzzy k-nearest neighboring (fuzzy k-NN) algorithm (known simply as Q-fuzzy approach).

A model was suggested for the cervical cancer prediction model (CCPM) that produces an early prediction of cervical cancer with input risk factors [34]. CCPM eliminates outliers first by employing outlier identification methods such as Density-Based Spatial Noise Cluster (DBSCAN) and isolation Forest (iForest) by balancing the number of cases in the dataset. This approach has shown greater accuracy in cervical cancer forecasting. To design an integrated cervical cell diagnostic and screening device, the authors have developed a new Regionally Growing Extraction Function (RGBFE) to extract diagnostic features from the images [35]. Data from the cervical cell images with extracted features were supplied into the intelligent diagnostic component. Precancerous phases were forecasted using a new architecture called the Hybrid Multilayered Perceptron (H2MLP) network using an artificial neural network is created. The cells are classified into normal, low-quality intraepithelesis (LSIL), and high-quality intraepithelesis (HSIL). Improved screening systems are also inaccessible in developing countries, owing to the difficulty and time-consuming nature of manually screening irregular cells from a cervical cytology specimen. This system focused on transfer learning, and pretrained and densely connected convolutional networks are used to suggest a computer-aided diagnostic (CAD) method for automated cervical image classification to assess CIN2 or higher level lesions in the cervical imaging (ImageNet and Kaggle). The effect of various training strategies on model results, including scratch random initialization (RI), pretrained model (FT) tuning, different size of training data, and K-fold cross validation, was evaluated. Experimental findings demonstrated accuracy of 73.08% for 600 test images [36]. The summary of the literature related to the screening of cervical cancer is provided in Table 1. Owing to the millions of cells that a pathologist must examine, Pap smear screening takes longer days for analyses. Deep learning models were used to identify all cells and other materials present in the Pap smear image screening. The system is often difficult to classify since two cells overlap. To address the need for this problem, meticulously annotated data is required; developing this form of the medical field dataset is very difficult. Considering the challenges mentioned above, a novel deep learning model for cervical cancer screening via colposcopy is proposed. The significant aspects of using colposcopy images for cervical cancer screening are that it provides more focus to the patients because it is a simple and noninvasive procedure (no need to introduce instruments into the body). When compared to the other tests, the colposcopy dataset array is sparse. The automated classification of cervical cancer from colposcopy images helps mass screening for medical professionals to quickly determine whether further diagnostic checks are necessary. This paper presents the computerized system for cervical cancer prediction using colposcopy images. The critical contribution of the article is as follows:

(i) This research is aimed at developing automatic cervical cancer detection from colposcopy images using the proposed deep convolutional neural network named CYENET. Unlike previous work reported in the literature, this proposed method does not require segmentation and feature engineering stages; it can also extract the discriminative features using ensemble approaches

(ii) The transfer learning approach is used by fine-tuning the VGG19 model, which is widely used for medical image processing to predict accuracy. Besides the extensive experiment on the cancerous and noncancerous colposcopy images to effectively demonstrate the proposed CYENET (colposcopy ensemble network) and pretrained VGG model with recently proposed methods, and our proposed method achieves

Table 1: Summary of the related works for screening cervical cancer.

S.no	Methods	Dataset	Advantages	Disadvantages
1	Inception V3 model [1]	Herlev dataset	(i) High accuracy (ii) Good universality Low complexity	(i) The deep network needs further study to investigate cervical cells.
2	Transfer learning, pretrained DenseNet [2]	Fujian Maternal and child health hospital Kaggle	(i) More feasibility and effective	(i) Limited data
3	CNN-extreme learning machine- (ELM-) based system [6]	Herlev dataset	(i) Fast learning (ii) Easy convergence (iii) Less randomized	(i) More complexity (ii) Need more investigation
4	Gene-assistance module, voting strategy [7]	Chinese hospital and Universitario De Caracas, Venezuela	(i) More scalable and practical	(i) Limited datasets
5	Random forest and Adaboost [14]	Radiotherapy dataset	(i) Better treatment planning	(i) Need to extract features (ii) Painful treatment
6	ColpoNet [16]	Colposcopy images	(i) Better accuracy (ii) Efficient classification	(i) Need to improve accuracy by extracting relevant information
7	CNN Model [17]	Papanicolaou-stained cervical smear dataset	(i) Better sensitivity and specificity	(i) Reported 1.8% false-negative images
8	Fourier transform and machine learning methods. [18]	Microscopic images	(i) Fully automatic system (ii) Saving precious time for the microscopist	(i) The level of complexity is more
9	CNN-SVM model [21]	Herlev and one private dataset	(i) Good robustness (ii) Highest accuracy	(i) Need improvement to adjust parameter (ii) Need of hand-crafted features
10	Stacked Autoencoder [27]	UCI database	(i) High accuracy (ii) Reduced data dimension	(i) Training time is very high due to reducing the dimension
11	PSO with KNN algorithm [33]	Cervical smear images	(i) Better accuracy (ii) Good feature selection	(i) Time-consuming due to two-phase feature selection
12	Ensemble model [34]	PAP smear image	(i) For 2 class problem achieves the accuracy of 96% (ii) For 7 class problem achieves an accuracy of 78%	(i) Overall of cells are difficult to identify
13	Multimodal deep network [37]	National Cancer Institute	(i) Good correlation (ii) High accuracy (iii) Learn better complementary features	(i) More complexity in image fusion

better accuracy as compared with the existing method in terms of classifying cervical cancer from colposcopy images

(iii) The convolutional neural network from scratch is designed to automate screening the cervical images by using an optimized architecture with an ensemble approach named CYENET (colposcopy ensemble network) deep learning architecture with a significant increase in diagnostic accuracy

(iv) Intel ODT dataset is used for experimentation. The data augmentation technique is performed on the colposcopy images to prevent the trained model's overfitting problem. This technique is an efficient strategy to learn the particular features to achieve superior accuracy

(a) Another significant contribution of this paper is the use of occlusion sensitivity maps to visualize the picture characteristics of cervical cancers for classification purposes

3. Materials and Methods

A colposcopy image is an essential aid in early cancer diagnosis. The assessment and identification of people with irregular cytology who need further care or follow-up depend on the transition zone colposcopic examination (TZ). The title

of the TZ is also an essential aspect of this study. Intra- and interobserver heterogeneity in the colposcopy perception of distinctive properties is considered to be relatively strong, but the observer heterogeneity of the TZ form and squamous column junction (SCJ) visibility evaluation and the quantitative calculation of the intra- and interobserver similarities of TZ contour tracing [37] are hardly studied. A TZ has been graded as type 1 because it is fully ectocervical (without any endocervical portion). Type 2 and Type 3 transition areas still have an endocervical component. When the latest SCJ was fully visible in TZ, it was considered a type 2. If even using external instruments, the new SCJ was not fully visible, and it was listed as type 3 [38]. It is used to assess a patient with pathological cytology, although it is not a final diagnostic examination. Variations may be made by the same colposcopies or by various colposcopies. The biggest downside of using colposcopy as a diagnostic instrument is the clinician's expertise and experience. Different experiments demonstrated good sensitivity and low accuracy in colposcopy diagnosed invasive and preinvasive cervix lesions [39].

The ensemble learning approach is employed using seven machine learning algorithms that are stacked together for automated detection for hepatocellular carcinoma [40] and using collaborative representation classification with boosting technique for classifying the hyperspectral image [41]. The flow map of the proposed automated method for detecting early cervical cancer is shown in Figure 1. The region between the original and the new SCJ is described colposcopically as the TZ [42]. The recognition of the TZ is essential information that all colposcopies require. Next, to identify the TZ as type 1, 2, or 3, you must find the new boundary between squamous and columnar epithel. Figure 2 displays example pictures from the dataset with type 1, 2, and 3 classifications. The center image in the green is the image taken by passing the green light to improve the cervical part's visibility.

3.1. Deep Convolutional Neural Network Model. The CNN models have been popular in many image processing applications, including medical image analysis. Detecting cervical cancer in the colposcopy images is an obvious computer vision problem. When comparing deep learning with conventional features, the neural network, especially convolutional, is used to distinguish cases type 1, type 2, and type 3. The test is to diagnose cervical lesions using deep convolutional neural networks (moderate). The proposed VGG 19 (TL) model is fine-tuned to classify three cervical cancer classes by freezing the top layers and tested with the cervical image dataset. We proposed a CYENET architecture by incorporating the essential advantages of depth and parallel convolutional filter, to enhance the extraction of specific cervical cancer features from colposcopy images. The proposed model consists of two types of convolution layers, i.e., traditional convolution layers at the beginning of a network preceded by one single convolution filter and multiple convolution layers to extract various features from the same data. Multiple convolutional filters are used to remove the biased parts to reduce the overfitting effect. This proposed model involves three phases: (1) data preprocessing, (2) CNN model training, and (3) classification results. The CYENET model consists of 15 convolutional layers, 12 activation layers, five max pooling layers, and four cross channel normalization layers. The test data are entered into the trained model, and the output parameters are measured. The initial strata are inspired by Google net architecture, several layers for manipulating functionality, and two fully connected layers with Softmax classification layers seen in Figure 3. The network description of the CYENET model is provided in Table 2 that refers to the convolution layer and max-pooling layer, varying the filter size in the parallel convolutional block.

3.2. Dataset and Preprocessing. The dataset consists of 5679 colposcopy photographs obtained from the cervical screening data collection by Intel and Smartphone ODT. The data is classified by considering the transition zone visible in the diagnostic study's specific picture [36]. The dataset is preprocessed to delete all the cases' ethical details. Firstly, the data are divided into three categories through diagnostic records: type 1, type 2, and type 3. A referenced pretrained dataset identifies the area of interest (ROI) of the cervical images due to minimal professionalism with MATLAB image labeler applications' assistance. The central region in which the lesion occurs is the ROI area called the clinic's transition zone (TZ). The original picture is obtained first with annotations, marks, and ROI.

The total images are 691 cases of type 1, 3126 cases of type 2, and 1862 cases of type 3. By observing the entire dataset, it is found that the dataset is imbalanced due to its unequal distribution of images. Due to dataset imbalancing, the model maybe leads to an overfitting issue. To overcome this issue, the oversampling technique is adopted. The oversampling method is known to repeat the type 1 and type 3 images arbitrarily and equal to the number of images in type 2. The cumulative images in the data collection after the oversampling technique are 9378 images. Secondly, data enhancement methods are used to optimize the volume of training data. Using the data augmentation method, the model robustness is increased, and the overfitting problem is reduced. The input image is augmented by rotating, adjusting the brightness, cropping, and randomly increasing the dataset. After the image augmentation process, the total image size is increased to 11266. All transformed image data is eventually dimensioned to 227×227 for CNN to fit the model. The dataset is divided into the training data with 7498 images, validation data with 1,884, and testing data with 1,884 photos. Figures 4(a) and 4(b) display the data augmentation technique to increase the data before the input of data to train neural networks.

3.3. Model Parameters. In this work, the two-deep learning model is used to diagnose cervical lesions through the colposcopic images. The transfer learning VGG_19 is fine-tuned for the proposed method, and CYENET architecture developed from scratch. The standard neural network framework uses a single type of CNN filter with an input data size varying from 1×1 to 5×5. The filter convolved with the input data to produces the same input data with a discriminatory feature map. The multilayer convolutional filter design's motivation is fundamental that incorporating several convolutional filters to extracts the discrimination-based multilayer

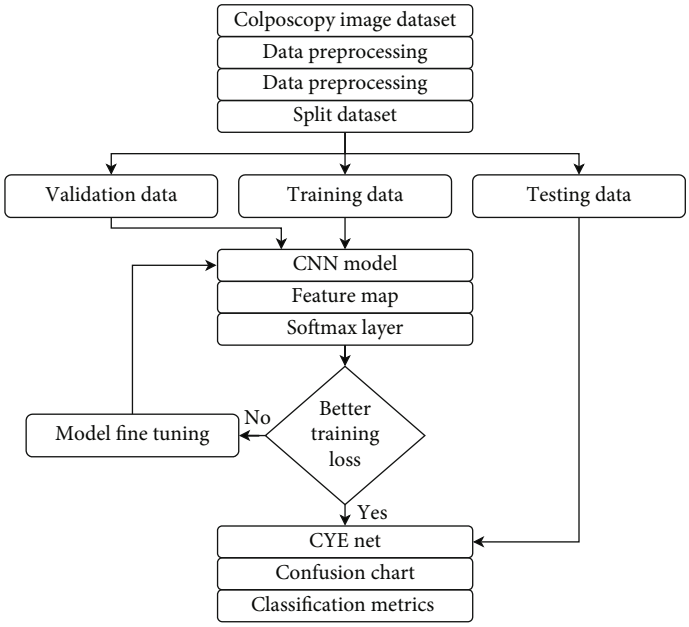


Figure 1: Flow chart of the proposed CYENET model for diagnosis of cervical cancer.

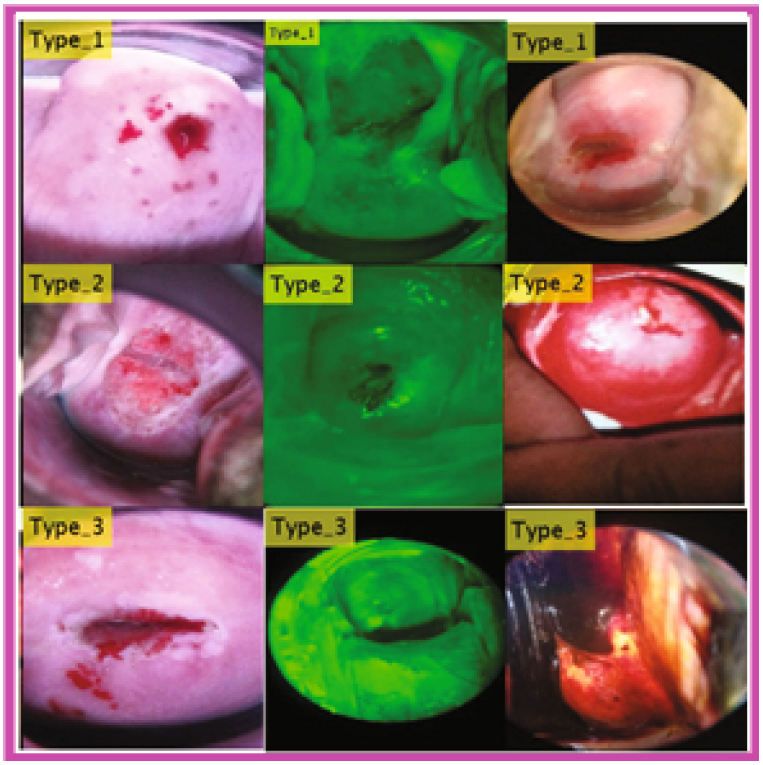


Figure 2: Dataset samples of type 1, type 2, and type 3 classes.

features. It extends further clusters from the same data. The three different kernel sizes are included in the training timing with 1×1, 3×3, and 5×5 to extract specific features. The proposed CYENET architecture and model parameters are fixed as an epoch of 50, batch size of 64, Adam optimization algorithm with a learning rate of 0.0001, and a decaying learning rate of 0.01 using piecewise technique every ten epochs. Before training, the data are shuffled at each point to

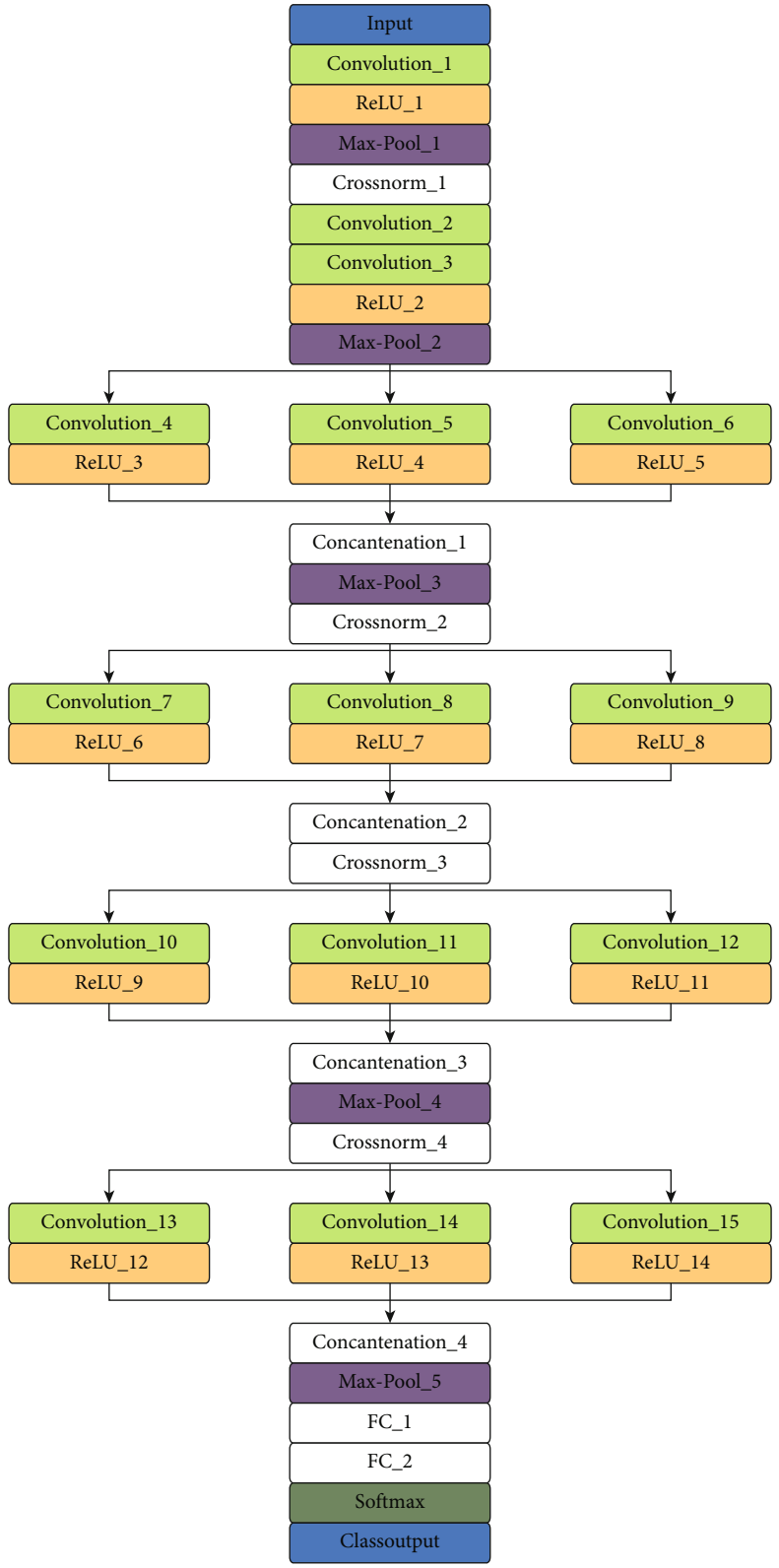


Figure 3: Strucutre of the proposed CYENET model.

Table 2: Description of network architecture of the CYENET model.

Layer No.	Layer type	Filter size	Stride	No. of filters	FC units	Input	Output
1	Convolution 1	5×5	2×2	64	—	$3 \times 227 \times 227$	$64 \times 112 \times 112$
2	Max-pool_1	3×3	2×2	—	—	$64 \times 112 \times 112$	$64 \times 56 \times 56$
3	Convolution 2	1×1	1×1	64	—	$64 \times 56 \times 56$	$64 \times 56 \times 56$
4	Convolution 3	3×3	1×1	128	—	$64 \times 56 \times 56$	$128 \times 56 \times 56$
5	Max-pool_2	3×3	2×2	—	—	$128 \times 56 \times 56$	$128 \times 28 \times 28$
6	Parallel convolution 1	1×1, 3×3, 5×5	1×1	$32 \oplus 64 \oplus 128$	—	$128 \times 28 \times 28$	$224 \times 28 \times 28$
7	Max-pool_3	3×3	2×2	—	—	$224 \times 28 \times 28$	$224 \times 14 \times 14$
8	Parallel convolution 2	1×1, 3×3, 5×5	1×1	$32 \oplus 64 \oplus 128$	—	$224 \times 14 \times 14$	$224 \times 14 \times 14$
9	Parallel convolution 3	1×1, 3×3, 5×5	1×1	$32 \oplus 64 \oplus 128$	—	$224 \times 14 \times 14$	$224 \times 14 \times 14$
10	Max-pool_4	3×3	2×2	—	—	$224 \times 14 \times 14$	$224 \times 7 \times 7$
11	Parallel convolution 4	1×1, 3×3, 5×5	1×1	$32 \oplus 64 \oplus 128$	—	$224 \times 7 \times 7$	$224 \times 7 \times 7$
12	Max-pool_5	5×5	1×1	—	—	$224 \times 7 \times 7$	$224 \times 2 \times 2$
13	Fully connected 1	—	—	—	512		
14	Fully connected 2	—	—	—	3		

bring about a normalizing effect during training. Additional discriminative features are extracted by each convolutional layer which is adding an advantage in prediction. Figure 5(a) shows the activation map for type 1 cases extracted from the single filter from the convolutional layer 1and Figure 5(b). The activation map for type1 instances extracted from the 64 filters from the convolutional layer 1. It has been done to understand what features our CNN model is extracting for the detection of particular classes.

Mathematical equations that decide the performance of a neural network are activation functions. The functionality is attached to each neuron in the network to determine whether or not it should be triggered ("fired"), depending on input relevance with the model prediction. ReLU is a piecewise linear function that, if the input is positive, outputs directly; otherwise, it outputs zero. The ReLU activation is used due to its faster converges and avoids easy saturation. It overcomes the problem faced by logistic regression and tan hyperbolic function of an inability to output the values greater than 1. The ReLU activation function is used in all the hidden layers. It is defined as

$$f(x) = \max(0, x), \tag{1}$$

where x is the input of the neuron. The ReLU activation function is programmed to exit the limitless activation function. The concatenation layer is used to concatenate the different features provided by the other kernel. After each concatenation layer, the local response normalization is employed to carry out the channel wise normalization of the activation function to reduce the model's overfitting problem. The local response normalization can be done in two ways: (i) within the channel and (ii) across the channel. In this proposed method, the local response normalization is carried out as cross channel normalization for pixel-wise normalization in the particular layer. It is given by equation (2).

$$x_i = \frac{x_i}{\left(k + \left(\alpha \sum_j x_j^2\right)\right)^{\beta}}. \tag{2}$$

In equation two, the terms k, α, and $\beta \in R$ are hyperparameters, and x_i is the input pixel value. The A_x pooling layer is used to minimize dimensionality after performing normalization. The max-pooling layer is used to reduce the dimension of features extracted from the convolutional layer and reduce the model's computation complexity by only keeping the channel's maximum pixel values with the specified kernel size 2×2. After the max-pooling layer 5, fully connected layer 1 with 128 output nodes with a drop out ratio of 0.5 is connected that follows the FC1 layer. The fully connected layer 2 with three output nodes is associated with the dropout ratio of 0.3% to reduce the overfitting problem. The softmax layer outputs each class's probabilities concerning the ground truth marks of the training and validation performance. The colposcopic images' three-class output are type 1, type 2, and type 3 to reduce the model's computation complexity instead of having 100 to 1000 nodes. The softmax activation function is indicated as

$$f_i(z) = \frac{e^{z_i}}{\sum_g e^{z_g}}. \tag{3}$$

In equation (3), f_i is the i^{th} part of the class scores f and z vector, abd a vector of arbitrary real-valued scores is squashed 0 to 1 with the probability ofthe prediction rate. The categorical crossentropy function is used as the cost function to determine the error between the predicted and observed classes. The

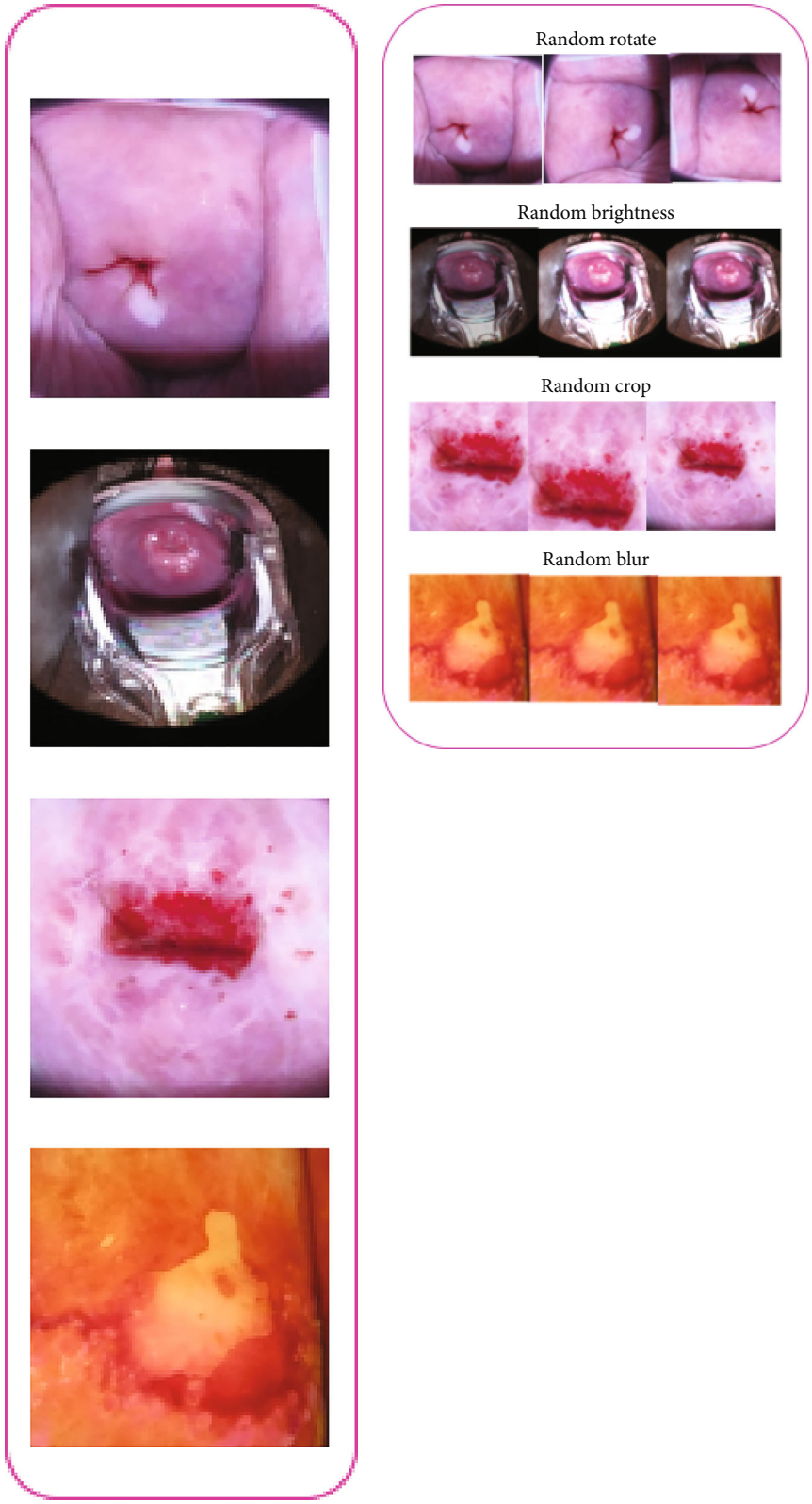
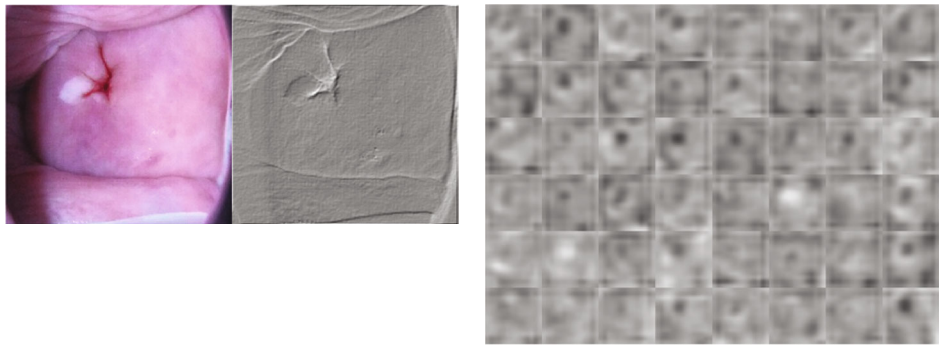


Figure 4: (a) Sample input images and (b) augmented images using different techniques.



(a) (b)

Figure 5: (a) Feature map of the convolutional layer with (a) 1 filter and (b) 64 filter.

categorical crossentropy function is given in equation (4).

$$H_p(q) = -\sum_{i=1}^{N} y_i.\log\left((\widehat{y}_i)\right). \tag{4}$$

In equation (4), $(\widehat{y}_i)$, the i^{th} scalar value is in the model output, y_i is the corresponding target value, and N is the number class label (0 for type 1, 1 for type 2, and 2 for type 3). We investigated the two models CYENET and VGG 19, in the proposed process. The CYENET is developed from scratch, and the model VGG 19 is explored through the adaptation of the transfer learning process. Both the model is trained to classify the type of cervical cancer from the colposcopic images.

4. Results and Discussion

The experiment is implemented in MATLAB 2020b, performed on a 24 GB Quadro NVIDIA RTX 6000 workstation computer with an Intel i9 processor. Experimental data is derived from the Kaggle dataset [36]. The colposcopy cervical cancer dataset is split into 80% training, 10% validation, and 10% testing. Approximately 7498 training images and 1884 validation images are used for the training and validation process. The depth of the layer, initial learning rate, optimizer, momentum value, and L2 regularization value are calculated from the Bayesian optimization. The number of epochs is fixed as 50 for training the model. The model is trained with a multi-GPU environment, batch size of 64, and initial learning rate of 0.0001. As discussed in Section 3, CYENET and VGG 19 with fine-tuning are trained with the same image dataset with fixed parameters. The precision, sensitivity, specificity, and Cohen's kappa score are evaluated to analyze the deep learning model. The confusion matrix is also used to test the models since it deals with a multiclass classification problem. The confusion matrix is used to analyze the classification model's performance in Figure 6, the training accuracy of the proposed method VGG_19, and the CYENET model trained against the training dataset with epoch 50.

The training accuracy gradually increased concerning the epoch's number and reached the training accuracy of 97.1% for the CYENET model and 87% for the VGG_19 (TL) model. The validation plot for the proposed CYENET and the precisely tuned VGG 19 against the epoch is shown in Figure 7. The accuracy of the model is undoubtedly growing regarding the number of epochs the model is trained. After 23 epochs, the proposed CYENET model achieves the validation accuracy value of 91.3%. Simultaneously, the sophisticated VGG-19 model achieved some early oscillation in the accuracy due to the chosen learning rate of 0.0001. The VGG 19 model obtained a validation accuracy of 68.8%. The results indicate that cervical screening from colposcopic images of the CYENET model performs better than the VGG19 model due to its more robust and more straightforward architecture.

The training and validation loss curve of the proposed model CYENET and VGG 19 are shown in Figure 8. The model convergence of the proposed network is determined by the shift in the validation loss curve. Compared to the VGG 19, CYENET converges very quickly with a loss value of 0.2982, and the model VGG 19 converges to 0.9885 loss values. In comparison, the validation model of the VGG 19 is unstable, CYENET is stable, and the loss curve is smoother.

Figure 9 displays the confusion matrix for the proposed CYENET model with test data, which shows the cumulative number of images projected with accurate label correspondence to the predicted label data from the confusion matrix. The CYENET confusion matrix includes true positive (T_{rue} positive), false positive (F_{alse} positive), true negative (T_{rue} negative), and false negative (F_{alse} negative). Table 2 reports accuracy, sensitivity, specificity, positive predicted value (PPV), and negative predicted value (NPV) as our evaluation metrics. Sensitivity and specificity are the most accurate assessment metrics for classifier completeness computed from the confusion matrix in medical images.

$$\text{Accuracy} = \frac{T_{\text{rue}}\,\text{Positive} + F_{\text{alse}}\,\text{Negative}}{T_{\text{rue}}\,\text{Positive} + T_{\text{rue}}\,\text{Negative} + F_{\text{alse}}\,\text{Positive} + F_{\text{alse}}\,\text{Negative}}, \tag{5}$$

$$\text{Sensitivity} = \frac{T_{\text{rue}}\,\text{Positive}}{T_{\text{rue}}\,\text{Positive} + F_{\text{alse}}\,\text{Negative}}, \tag{6}$$

$$\text{Specificity} = \frac{T_{\text{rue}}\,\text{Negative}}{T_{\text{rue}}\,\text{Negative} + F_{\text{alse}}\,\text{Positive}}, \tag{7}$$

$$\text{PPV} = \frac{T_{\text{rue}}\,\text{Positive}}{T_{\text{rue}}\,\text{Positive} + F_{\text{alse}}\,\text{Positive}}, \tag{8}$$

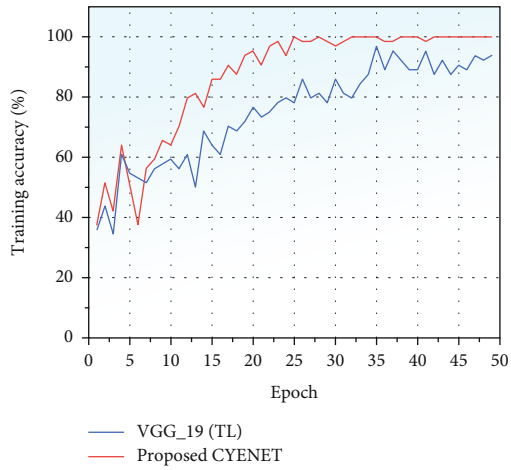


Figure 6: Training accuracy plot for the CYENET and VGG 19 (TL) model.

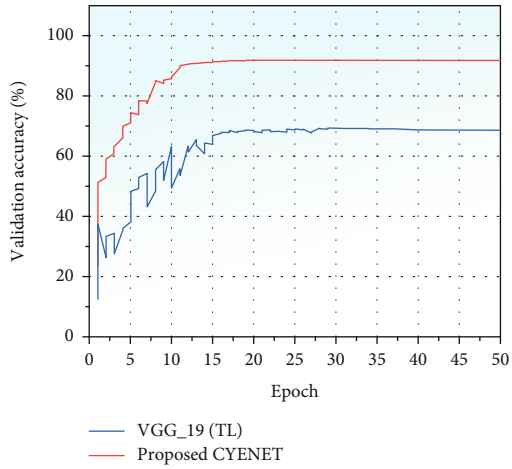


Figure 7: Validation accuracy plot for the CYENET and VGG 19 (TL) model.

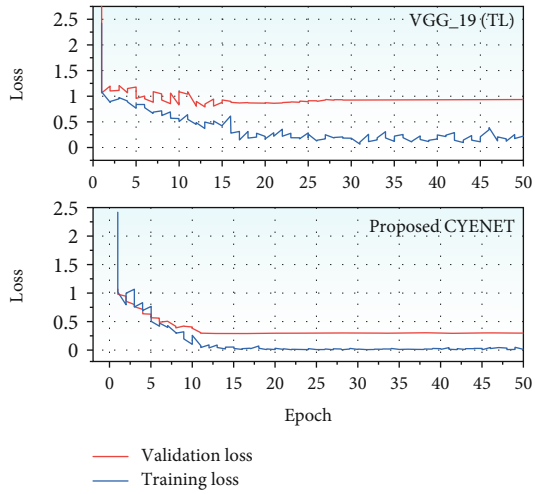


Figure 8: Training and validation loss curve for the CYENET and VGG_19 (TL) model.

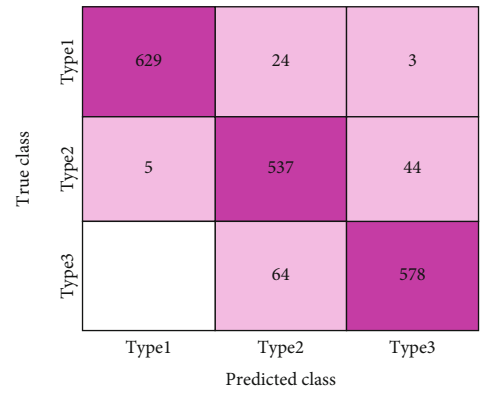


Figure 9: Confusion chart of the proposed CYENET.

$$\text{NPV} = \frac{T_{\text{rue}}\ \text{Negative}}{T_{\text{rue}}\ \text{Negative} + F_{\text{alse}}\ \text{Negative}}. \tag{9}$$

Sensitivity is the percentage of people who test positive out of all those who have the disease. The proportion of people who test negative among all those who do not have the disease is the specificity of a test. The PPV is the possibility that a person will have the disease after receiving a positive test result. The NPV is the possibility that a person will not have the disease after receiving a negative test result. Table 3 shows the test results of the proposed model tested with 1884 test images. The above experimental result CYENET model outperformed all the other models in the table trained on the colposcopic images. The DenseNet-121 and DenseNet-169 achieved lower accuracy with 72.42% and 69.79%, respectively. The model performance is influenced by the size of the dataset and also the depth of the layer. The deep architecture may decline its overall model classification performance due to the problem of interclass similarity. The Inception-Resnet-v2 model provides a lower specificity of 70.6% due to the dataset imbalance. The model is prone to image characteristics such as contrast, brightness, tone, and quality of the image capturing devices. The SVM method discussed in the performance table achieves an accuracy of 63.27% and the lowest sensitivity value of 38.46%. The model is trained on both hand-crafted features and features extracted from the CNN model. It is a time-consuming difficult task to perform in real-time even though the cost is nominal. The colponet model based on the CNN architecture provides an accuracy of 81.0% for classifying cervical cancer from the colposcopy images. The difference between the training accuracy and validation accuracy of the component model is very high. The model's training time is very high where the model is trained for 3000 epochs and provides the convergence loss of 1.12, which is very for the application of medical image processing. The proposed CYENET model is designed and trained to achieve an overall testing accuracy of 92.30% by considering all these disadvantages. The proposed model uses a different filter size to extract distinct features and works well for unexpected data. It offers a 92.40% sensitivity and a 96.20% specificity, which improves sensitivity and specificity by approximately 25% compared with Inception-Resnet-v2 in [45]. The proposed method has trouble distinguishing the

Table 3: Comparative experiment results of proposed architecture with different models.

Model name	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Ref
DenseNet-121	72.42	59.86	76.83	48.39	84.52	[43]
DenseNet-169	69.79	65.00	71.48	44.84	85.31	[43]
Colponet	81.0	—	—	—	—	[16]
SVM	63.27	38.46	71.85	32.43	76.87	[44]
Inception-Resnet-v2	69.3	66.70	70.6	47.20	84.00	[45]
CYENET	92.30	92.40	96.20	92.00	95.00	Present study
VGG19 (TL)	73.30	33.00	79.00	70.00	88.00	Present study

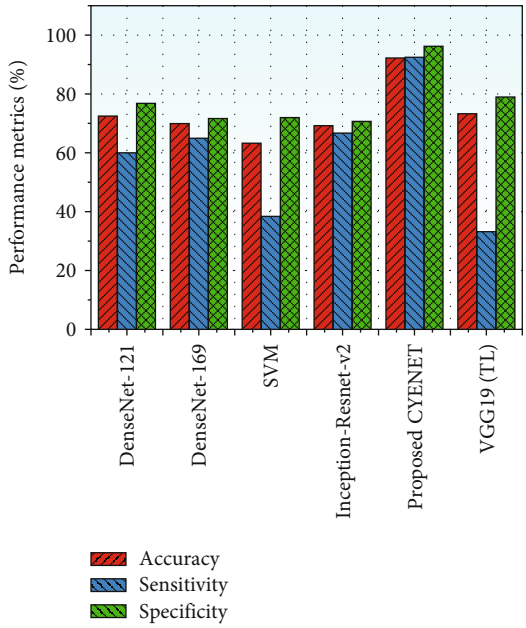


Figure 10: Performance metric comparison of systems.

false positive samples from those with fewer false positives. The CYENET model evaluates cervical epithelial features rather than morphological ones, and its false-negative epidermal features are close to true negatives. The improved sensitivity and precision indicate that positive and negative samples are predicted wisely. Since the model proposed is compared to traditional metrics such as precision, sensitivity, and specificity, sometimes, the above metrics for multiclass problems does not make sufficient to prove the model's general ability.

By taking this into consideration, the F1 score of the model is calculated by the harmonic mean of the accuracy and reminder. Still class imbalances in the dataset influence the f1 score, but Cohen's Kappa metrics are seen to have the right measure to tackle multiple class issues as well as class imbalances which the statistical standards to find the agreement between two parties. The suggested models CYENET and VGG 19 (TL), both calculated with the colposcopy images of the cervical cancer diagnosis, are measured using F1 measurements and Cohen's kappa. The F1 score of the proposed CYENET and VGG 19 (TL) is 92.0% and 44.80%, respectively, and Cohen's Kappa score of 88% and 53.5%, respectively. The proposed model CYENET is superior to the literature models and even to the proposed model VGG 19 (TL). Figure 10 shows the graphical representation of the model discussed in Table 2.Due to the existence of several distractors such as pubic hair, intrauterine instruments, the speculum, and even human parts, the proposed method for cervical cancer screening using colposcopy can suffer. Another issue with the proposed approach if the captured images are out of focus and prediction accuracy will be reduced.

Figure 11 indicates the positive and negative expected values (PPV) of CYENET and VGG 19 models. By fixing the probability (prevalence) of infection to 0.05, the positive predicted value and negative predicted value of the CYENET model are calculated with sensitivity and specificity of 92.40% and 96.20%, respectively, for varying probabilities of infection shown in Figure 11(b), and the VGG 19 (TL) model achieves the sensitivity and specivity of 33.0% and 79.0%, respectively, for varying probability demonstrated in Figure 11(a). The incidence graph helps the medical practitioners to classify groups with a previous risk of diagnosis with cervical cancer.

The overall run time of the proposed model CYENET is 3 minutes 32 seconds, and for VGG19 5 minutes 24 seconds, the batch size of 64 is provided in Table 4. The total number of parameters for the CYENET is 8465376, and the total number of parameters for the VGG19 is 123642856. Still, the top layers are frozen to reduce the number of trainable parameters. Among the compared models, the densenet architecture proves to be having a significant training time due to its dense nature.

4.1. Occlusion Sensitivity Map Visualization. We used occlusion sensitivity maps [42] to determine the colposcopy images' aspects that are most appropriate for the CYENET classification decision in this experiment. Occlusion sensitivity is a simple technique for deciding which deep neural network uses image features to make a classification decision. Precisely, occlusion sensitivity measures the variation in likelihood score for a given class as a function of mask location by systematically occluding various portions of the input picture with an occluding mask (usually a grey square). Figure 12 depicts several cervical cancer input colposcopy images with occlusion sensitivity maps superimposed on them. The occlusion sensitivity maps indicate that the colposcopy images' parts contribute more to the score for cervical cancer classes and which factors contribute less or none at all. It can be seen from the occlusion maps that CYENET was able to distinguish regions with speculum and other opacities. Compared to the Grad-CAM process, the visualization results support our argument that occlusion sensitivity maps are intuitive and interpretable.

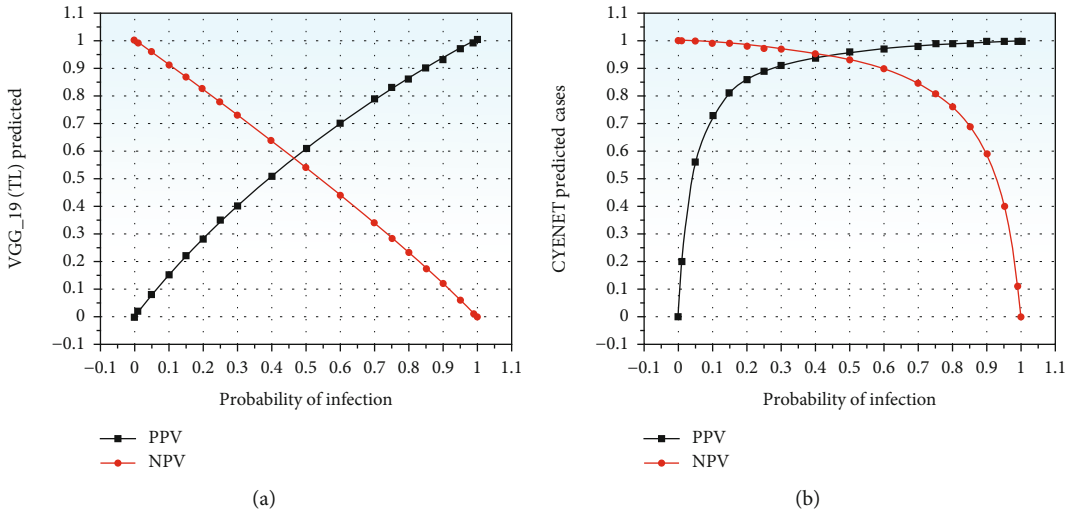


Figure 11: PPV and NPV curve of VGG_19 (TL) (a) and CYENET (b).

Table 4: Comparative results of proposed architecture with several parameters and run time.

Model name	Number of parameters	Run time (per epoch)
DenseNet-121 [43]	7978856	21 min 10 s
DenseNet-169 [43]	28681000	24 min 59 s
Colponet [16]	6977000	16 min 27 s
Inception-Resnet-v2 [45]	55843161	15 min 36 s
CYENET	8465376	3 min 32 s
VGG19 (TL)	123642856	5 min 24 s

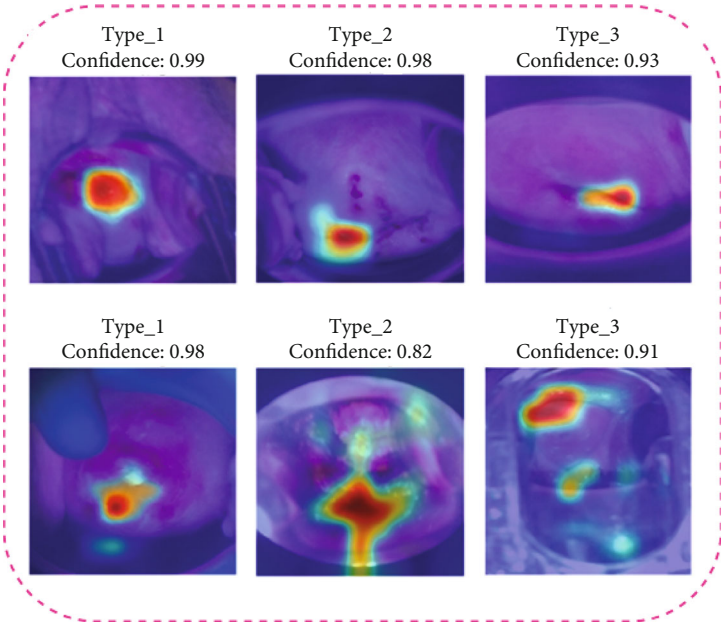


Figure 12: Occlusion sensitivity map for test data.

5. Conclusion

A new deep learning architecture name CYENET is proposed for classifying the cervical cancer type from colposcopic images. The image dataset is balanced using the oversampling technique for improving the classification results. Two models are presented in this paper. One is using a transfer learning approach with VGG19 architecture. The other is a dedicated new model called CYENET for cervical cancer type classification using the ODT colposcopy image dataset. Both the models are evaluated using classification accuracy, sensitivity, specificity, Cohen's Kappa score, and F1-measure. The VGG19 (TL) model's sensitivity and specificity are 33% and 79%, respectively, with Cohen's Kappa score of 53.5%. The classification accuracy for VGG19 was 73.3%. Relatively satisfied results are obtained for VGG (TL). From the kappa score of the VGG19 model, we can interpret that it comes under the category of moderate classification.

Similarly, the proposed CYENET exhibited high sensitivity, specificity, and kappa scores of 92.4%, 96.2%, and 88%, respectively. The classification accuracy of the CYENET model is improved as 92.3%, which is 19% higher than the VGG19 (TL) model. Comparing the results of CYENET with previously reported results of the work, CYENET is an effective and promising prospect as a diagnosis assist tool for clinicians. The proposed method of cervical cancer classification can benefit a target population that does not need invasive intervention. The proposed CYENET has better classification efficiency and can assist medical professionals and skilled healthcare practitioners in increasing the diagnostic sensitivity and accuracy of cervical cancer detection through colposcopy screening as a result. In the future, the theoretical deep learning model will be checked for different datasets. The approach can also be enhanced by combining some advanced image processing techniques and CNN algorithms to create a diagnostic system for cervical precancerous new data.

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.

References

[1] N. Dong, L. Zhao, C. H. Wu, and J. F. Chang, "Inception v3 based cervical cell classification combined with artificially extracted features," *Applied Soft Computing*, vol. 93, p. 106311, 2020.

[2] T. Zhang, Y. M. Luo, P. Li et al., "Cervical precancerous lesions classification using pre-trained densely connected convolutional networks with colposcopy images," *Biomedical Signal Processing and Control*, vol. 55, p. 101566, 2020.

[3] W. Hua, T. Xiao, X. Jiang et al., "Lymph-vascular space invasion prediction in cervical cancer: exploring radiomics and deep learning multilevel features of tumor and peritumor tissue on multiparametric MRI," *Biomedical Signal Processing and Control*, vol. 58, p. 101869, 2020.

[4] T. I. Yusufaly, K. Kallis, A. Simon et al., "A knowledge-based organ dose prediction tool for brachytherapy treatment planning of patients with cervical cancer," *Brachytherapy*, vol. 19, no. 5, pp. 624–634, 2020.

[5] J. Shao, Z. Zhang, H. Liu et al., "DCE-MRI pharmacokinetic parameter maps for cervical carcinoma prediction," *Computers in Biology and Medicine*, vol. 118, article 103634, 2020.

[6] A. Ghoneim, G. Muhammad, and M. S. Hossain, "Cervical cancer classification using convolutional neural networks and extreme learning machines," *Future Generation Computer Systems*, vol. 102, pp. 643–649, 2020.

[7] J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: an ensemble approach," *Future Generation Computer Systems*, vol. 106, pp. 199–205, 2020.

[8] S. I. Kim, S. Lee, C. H. Choi, M. Lee, J. W. Kim, and Y. B. Kim, "Prediction of disease recurrence according to surgical approach of primary radical hysterectomy in patients with early-stage cervical cancer using machine learning methods," *Gynecologic Oncology*, vol. 159, no. 2020, pp. 185-186, 2020.

[9] M. Nayak, S. Das, U. Bhanja, and M. R. Senapati, "Elephant herding optimization technique based neural network for cancer prediction," *Informatics in Medicine Unlocked*, vol. 21, article 100445, 2020.

[10] V. Chandran, C. K. Patil, A. Karthick, D. Ganeshaperumal, R. Rahim, and A. Ghosh, "State of charge estimation of lithium - ion battery for electric vehicles using machine learning algorithms," *World Electric Vehicle Journal*, vol. 12, no. 1, p. 38, 2021.

[11] M. Suriya, V. Chandran, and M. G. Sumithra, "Enhanced deep convolutional neural network for malarial parasite classification," *International Journal of Computers and Applications*, pp. 1–10, 2019.

[12] A. Alghamdi, M. Hammad, H. Ugail et al., "Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities," *Multimedia Tools and Applications*, 2020.

[13] M. Hammad, A. M. Iliyasu, A. Subasi, E. S. L. Ho, and A. A. A. El-Latif, "A multitier deep learning model for arrhythmia detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[14] Q. Meng, "Machine learning to predict local recurrence and distant metastasis of cervical cancer after definitive radiotherapy," *International Journal of Radiation Oncology • Biology • Physics*, vol. 108, no. 3, article e767, 2020.

[15] J. Shan, R. Jiang, X. Chen et al., "Machine learning predicts lymph node metastasis in early-stage oral tongue squamous cell carcinoma," *Journal of Oral and Maxillofacial Surgery*, vol. 78, no. 12, pp. 2208–2218, 2020.

[16] S. K. Saini, V. Bansal, R. Kaur, and M. Juneja, "ColpoNet for automated cervical cancer screening using colposcopy images," *Machine Vision and Applications*, vol. 31, no. 3, pp. 1–15, 2020.

[17] P. Sanyal, P. Ganguli, and S. Barui, "Performance characteristics of an artificial intelligence based on convolutional neural network for screening conventional Papanicolaou-stained cervical smears," *Medical Journal, Armed Forces India*, vol. 76, no. 4, pp. 418–424, 2020.

[18] B. R. Jany, A. Janas, and F. Krok, "Automatic microscopic image analysis by moving window local Fourier transform and machine learning," *Micron*, vol. 130, article 102800, 2020.

[19] V. Karunakaran, V. N. Saritha, M. M. Joseph et al., "Diagnostic spectro-cytology revealing differential recognition of cervical cancer lesions by label-free surface enhanced Raman fingerprints and chemometrics," *Biologie et Médecine*, vol. 29, p. 102276, 2020.

[20] F. A. D. Jia, S. B. Zhengyi, and T. C. C. Zhang, "CNN-SVM network abstract," *Neurocomputing*, 2020.

[21] A. Dongyao Jia, B. Zhengyi Li, and C. Chuanwang Zhang, "Detection of cervical cancer cells based on strong feature CNN-SVM network," *Neurocomputing*, vol. 411, pp. 112–127, 2020.

[22] J. Ren, A. Zhang, and X. Wang, "Jo ur na l P re," *Pharmacological Research*, no. article 104743, 2020.

[23] M. F. Ijaz, M. Attique, and Y. Son, "Data-driven cervical cancer prediction model with outlier detection and over-sampling methods," *Sensors*, vol. 20, no. 10, pp. 2809–2822, 2020.

[24] E. M. L. Ruiz, T. Niu, M. Zerfaoui et al., "A novel gene panel for prediction of lymph-node metastasis and recurrence in patients with thyroid cancer," *Surgery*, vol. 167, no. 1, pp. 73–79, 2020.

[25] D. Stelzle, L. F. Tanaka, K. K. Lee et al., "Estimates of the global burden of cervical cancer associated with HIV," *The Lancet Global Health*, vol. 9, no. 2, pp. e161–e169, 2021.

[26] D. Goksuluk, G. Zararsiz, S. Korkmaz et al., "MLSeq: machine learning interface for RNA-sequencing data," *Computer Methods and Programs in Biomedicine*, vol. 175, pp. 223–231, 2019.

[27] K. Adem, S. Kiliçarslan, and O. Cömert, "Classification and diagnosis of cervical cancer with stacked autoencoder and softmax classification," *Expert Systems with Applications*, vol. 115, pp. 557–564, 2019.

[28] A. Goncalves, P. Ray, B. Soper et al., "Bayesian multitask learning regression for heterogeneous patient cohorts," *Journal of Biomedical Informatics: X*, vol. 4, article 100059, 2019.

[29] B. Deepa and M. G. Sumithra, "An intensity factorized thresholding based segmentation technique with gradient discrete wavelet fusion for diagnosing stroke and tumor in brain MRI," *Multidimensional Systems and Signal Processing*, vol. 30, no. 4, pp. 2081–2112, 2019.

[30] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, "A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images," *Computer Methods and Programs in Biomedicine*, vol. 164, pp. 15–22, 2018.

[31] C. Zhang, W. Leng, C. Sun et al., "Urine proteome profiling predicts lung cancer from control cases and other tumors," *eBioMedicine*, vol. 30, pp. 120–128, 2018.

[32] K. Matsuo, S. Purushotham, A. Moeini et al., "A pilot study in using deep learning to predict limited life expectancy in women with recurrent cervical cancer," *American Journal of Obstetrics and Gynecology*, vol. 217, no. 6, pp. 703–705, 2017.

[33] A. M. Iliyasu and C. Fatichah, "A quantum hybrid PSO combined with fuzzy k-NN approach to feature selection and cell classification in cervical cancer detection," *Sensors*, vol. 17, no. 12, p. 2935, 2017.

[34] A. Sarwar, V. Sharma, and R. Gupta, "Hybrid ensemble learning technique for screening of cervical cancer using Papanicolaou smear image analysis," *Personalized Medicine Universe*, vol. 4, pp. 54–62, 2015.

[35] N. Karjane and D. Chelmow, "New cervical cancer screening guidelines, again," *Obstetrics and Gynecology Clinics of North America*, vol. 40, no. 2, pp. 211–223, 2013.

[36] C. Schwaiger, M. Aruda, S. Lacoursiere, and R. Rubin, "Current guidelines for cervical cancer screening," *Journal of the American Academy of Nurse Practitioners*, vol. 24, no. 7, pp. 417–424, 2012.

[37] T. Xu, H. Zhang, X. Huang, S. Zhang, and D. N. Metaxas, "Multimodal deep learning for cervical dysplasia diagnosis," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Lecture Notes in Computer Science, pp. 115–123, Springer, Cham, 2016.

[38] M. E. Plissiti, E. E. Tripoliti, A. Charchanti, O. Krikoni, and D. I. Fotiadis, "Automated detection of cell nuclei in pap stained cervical smear images using fuzzy clustering," *IFMBE Proceedings*, vol. 22, pp. 637–641, 2009.

[39] J. Shi, R. Wang, Y. Zheng, Z. Jiang, H. Zhang, and L. Yu, "Cervical cell classification with graph convolutional network," *Computer Methods and Programs in Biomedicine*, vol. 198, p. 105807, 2021.

[40] W. Książek, M. Hammad, P. Pławiak, U. R. Acharya, and R. Tadeusiewicz, "Development of novel ensemble model using stacking learning and evolutionary computation techniques for automated hepatocellular carcinoma detection," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1512–1524, 2020.

[41] H. Su, Y. Yu, Q. Du, and P. Du, "Ensemble learning for hyperspectral image classification using tangent collaborative representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3778–3790, 2020.

[42] H. Zhang, C. Chen, R. Gao et al., "Rapid identification of cervical adenocarcinoma and cervical squamous cell carcinoma tissue based on Raman spectroscopy combined with multiple machine learning algorithms," *Photodiagnosis and Photodynamic Therapy*, vol. 33, p. 102104, 2021.

[43] F. B. M. Suah, "Preparation and characterization of a novel co(II) optode based on polymer inclusion membrane," *Analytical Chemistry Research*, vol. 12, pp. 40–46, 2017.

[44] T. Xu, H. Zhang, C. Xin et al., "Multi-feature based benchmark for cervical dysplasia classification evaluation," *Analytical Chemistry Research*, vol. 63, pp. 468–475, 2017.

[45] B. J. Cho, Y. J. Choi, M. J. Lee et al., "Classification of cervical neoplasms on colposcopic photography using deep learning," *Scientific Reports*, vol. 10, no. 1, p. 13652, 2020.

Hindawi
BioMed Research International
Volume 2021, Article ID 5548517, 10 pages
https://doi.org/10.1155/2021/5548517

Research Article

Coronary Vessel Segmentation by Coarse-to-Fine Strategy Using U-nets

Le Nhi Lam Thuy,[1,2] **Tan Dat Trinh**,[1] **Le Hoang Anh**,[1] **Jin Young Kim**,[3] **Huynh Trung Hieu**,[2] **and Pham The Bao**[1]

[1]*Information Science Faculty, Sai Gon University, Vietnam*
[2]*Industrial University of Ho Chi Minh City, Vietnam*
[3]*Electronics and Computer Engineering Department, Chonnam National University, Republic of Korea*

Correspondence should be addressed to Pham The Bao; ptbao@sgu.edu.vn

Received 15 January 2021; Revised 4 March 2021; Accepted 23 March 2021; Published 9 April 2021

Academic Editor: Changming Sun

Each level of the coronary artery has different sizes and properties. The primary coronary arteries usually have high contrast to the background, while the secondary coronary arteries have low contrast to the background and thin structures. Furthermore, several small vessels are disconnected or broken up vascular segments. It is a challenging task to use a single model to segment all coronary artery sizes. To overcome this problem, we propose a novel segmenting method for coronary artery extraction from angiograms based on the primary and secondary coronary artery. Our method is a coarse-to-fine strategic approach for extracting coronary arteries in many different sizes. We construct the first U-net model to segment the main coronary artery extraction and build a new algorithm to determine the junctions of the main coronary artery with the secondary coronary artery. Using these junctions, we determine regions of the secondary coronary arteries (rectangular regions) for a secondary coronary artery-extracted segment with the second U-net model. The experiment result is 76.40% in terms of Dice coefficient on coronary X-ray datasets. The proposed approach presents its potential in coronary vessel segmentation.

1. Introduction

In all living mammal species, including humans, blood vessels inside the body are highly organized and complex, ensuring that blood flows unidirectionally on vessel branches. Localization, segmentation, and visualization of blood vessels from X-ray angiograms are highly necessary and useful in various medical diagnoses. Based on the blood vessel width, reflectivity, and abnormal branching, we can determine symptoms of vessel diseases such as stenosis, vascular malformation, and atherosclerosis. By using the X-ray angiogram, medical experts or doctors manually detect and delineate the blood vessels. However, this process is time-consuming and challenging in the cases of enormous number of X-ray angiograms and small and thin vessel structures. Hence, it is highly necessary to develop automatic and accurate blood vessel detection and segmentation methods from angiograms. Many related works conducted the coronary vessel segmentation based on the weak contrast between the coronary arteries and the background, strong overlapping shadows of the bones, nonuniform illumination in X-ray angiogram, small and thin vessel branches, complex shape of the vessel tree, and/or other body tissues [1, 2]. These factors can decrease the accuracy of segmentation results.

The improvements in coronary vessel enhancement and segmentation algorithms can be divided into six main categories, such as pattern recognition approaches, model-based approaches, tracking-based approaches, artificial intelligence-based approaches, neural network-based approaches, and miscellaneous tube-like object detection approaches [2]. Liao et al. [3] applied an enhanced multiscale approach to extract 2D coronary artery central lines from X-ray projection images. Authors introduced the 3D symbolic reconstruction based on an energy minimization

problem incorporating a soft epipolar line constraint and a smoothness term. The nonlinear anisotropic filtering [4] approach performs anisotropic smoothing without blurring the vessel edges on the local orientation. Hessian-based multiscale filtering [5–8] has been proposed for vessel enhancement. In this technique, an input image is filtered by the derivatives of a Gaussian at multiple scales. Then, the Hessian matrix is analyzed at each pixel in the filtered image to determine the structures' local shape. However, due to the second-order derivatives, the Hessian-based approaches are highly sensitive to noise. Furthermore, this approach led to suppressing junctions, as junctions are characterized similarly to the blob-like structures.

In [6], the authors proposed a filter model based on the regularized gradient vector correlation matrix to avoid the need for second-order derivatives. However, this technique faces the same limitations as Hessian-based filters in finding small and low-contrast vessels when dealing with angiography images, which are noisier and suffer from nonuniform illumination. Truc et al. [7] introduced a new framework for vessel enhancement by applying the directional information present in an image. The input images are first decomposed by a decimation-free directional filter bank (DDFB) into a set of directional images. Distinct appropriate enhancement filters are then used to enhance vessels in the respective directional images. Finally, the enhanced directional images are recombined to generate the output image with enhanced vessels. Although this approach is still noise-sensitive, it reveals the small vessel network and avoids junction suppression. Trinh et al. [8] introduced a hierarchical approach to extract coronary vessels from an X-ray angiogram. They applied the DDFB and Homographic Filtering (HF) since they are suitable for strengthening the vessels at different orientations and radii. To obtain the main and small coronary vessels in various sizes, they used a coarse-to-fine strategy for iterative segmentation based on the Otsu algorithm.

Recently, deep learning approaches have been applied for medical image segmentation and analysis [9–13]. These new powerful techniques based on convolutional neural networks (CNNs) lead to high performance in the field of medical imaging for segmentation without expert knowledge. Many studies confirm that deep learning models outperform traditional medical segmentation systems. In [10], authors developed a successful and well-known network based on the CNN, named as U-net, for biomedical image segmentation. The network architecture consists of two paths: encoder and decoder. The encoder is a contraction stack of convolutional layers used to capture the context of input images. After each convolutional layer, a rectified linear unit (ReLU), max pooling, and dropout layers are added. The decoder is an expansive path that is used to enable precise localization by using transposed convolutions. In the decoder, the final layer is used to map the feature vector to the binary prediction (i.e., vessel vs. nonvessel). The U-net requires the inputs as 2D image patches and returns the 2D segmentation probability map for each given patch. Milletari et al. [11] introduced a V-net architecture that adopts a volumetric CNN for prostate segmentation from MRI. Similar to U-net, V-net induced two paths. The first path (left path) of the V-net consists of a compression path. The second one (right path) decompresses the input image until its original size is reached. Holistic-net [12] was proposed for brain tumor segmentation. It is a combination of holistic CNNs and generalized Wasserstein Dice scores for multiclass segmentation. In [13], a graph neural network (GNN) is proposed to learn global vascular structures in medical images. The authors combined the GNN into a unified CNN architecture to learn not only local appearances but also the global structures of vessels.

Deep learning-based automated ventricle segmentation methods are summarized in the research [14]. Authors [15] developed a novel encoder-decoder deep network algorithm to exploit $2D + t$ sequential images' contextual information in a sliding window. The encoder extracts the temporal-spatial features. The skip connection layers subsequently fuse these features and deliver them to the corresponding decoder stages. The decoder employed the channel attention mechanism. In [16], the authors proposed a nested encoder-decoder architecture named T-Net. T-Net consists of several small encoder-decoders for each block constituting a convolutional network. They evaluated T-Net by segmenting only three main vessels in coronary angiography images and archive the Dice similarity coefficient score of 88.97%. In the research [17], the blood vessels are segmented from both the coronary angiogram and the retinal fundus images using a single VSSC Net after performing the image-specific preprocessing. The VSSC Net consists of two-vessel extraction layers with additional supervision on top of the base VGG-16 network. The VSSC Net attains average AUC values of 0.98205 across the target datasets. Authors [18] proposed a novel weakly supervised training framework to alleviate the annotator's burden by learning from noisy pseudo labels generated from automatic vessel enhancement instead of fully manual annotation. Their annotation-refining self-paced learning framework (AR-SPL) corrects the possible errors using suggestive annotation. Experiments confirm that their proposed framework largely reduced annotation cost and Dice score of 82.09%. Another study proposed an automated prostate MRI data segmentation using bicubic interpolation with improved 3D V-Net. Two clinical prostate-MRI data datasets were used to evaluate the model's effectiveness with the manual delineations available as the ground truth [19]. The segmentation result is 98.29% of average accuracy and 0.9765 of Dice metric.

With the supportive goal of interpreting pathophysiological processes and clinical decision-making, the study [20] developed a multiview recurrent aggregation network (MV-RAN) for the echocardiographic sequence's segmentation with the full cardiac cycle analysis. Experiments were conducted on spatial-temporal $(2D + t)$ datasets of multicenter and multiscanner clinical studies. Compared to other studies, the research [20] achieved results of 0.92 Dice score.

This study proposes a novel hierarchical approach to extract coronary vessels from X-ray coronary angiographic images. We use a coarse-to-fine strategy for iterative

segmentation based on the U-net model to segment the coronary vessels in various sizes as follows:

(i) We use U-net to segment the main and large blood vessels

(ii) We propose a new approach to extract junctions from vascular trees and detect small vessel regions based on the main information from extracted vessels

(iii) We apply the region-based U-net segmentation to locate and obtain the small vessels

2. Materials and Methods

In this section, we describe the proposed hierarchical approach in detail. As illustrated in Figure 1, our proposed framework includes tree main steps: preprocessing, extracting the large coronary vessels, and extracting the small ones.

The preprocessing procedure is applied to remove high-frequency noise and also enhance the contrast of X-ray coronary angiographic images. We first apply a Gaussian filter to smooth the vessel image. The Gaussian filter is low-pass filtering that is used to reduce high-frequency noise in order to make our vessel segmentation more accurate. In our study, we use Gaussian smoothing to detect false edges or artifacts (not small artery fragments) due to noises and reduce their effect on the input. In addition, a histogram equalization method [21] is applied to adjust the contrast of images. Figure 2 shows our preprocessing process.

In the next step, we apply a coarse-to-fine strategy for iterative segmentation. Particularly, we segment regions that include the main coronary vessels based on the high-contrast pixels. The main coronary vessels include features such as vascular tree and junctions. Subsequently, we use coarse information extracted in the previous step to detect the small vessels that often have low contrast and are affected by noises. We describe each step of the proposed technique in detail in the following sections.

2.1. Large Vessel Extraction Based on U-net. In this section, we describe a method to extract vessels by using U-net and the coarse-to-fine segmentation strategy. Figure 3 shows a block diagram of the vessel's extraction.

The U-net model is proposed for biomedical image segmentation [10]; as shown in Figure 4, the network architecture consists of encoder and decoder paths. The encoder is a contraction path that captures the context in the input image. The decoder is an expansive path that applies transposed convolutions to enable precise localization. In the decoder, the final layer maps the feature vector to the binary outputs such as vessel or nonvessel. The U-net receives the inputs as 2D image patches and returns the 2D segmentation probability map for each given patch.

The U-net uses the loss function as the cross-entropy function shown as follows:

$$J = -\sum_{x\in\Omega} w(x)\,\log\left(p_{l(x)}(x)\right), \tag{1}$$

where $p_{l(x)}$ is the soft-max function defined by $p_{l(x)} = \exp(a_l(x))/(\sum_{l'=1}^{L}\exp(a_{l'}(x)))$, where $a_l(x)$ is an activation in feature channel l at the pixel position $x \in \Omega$ with $\Omega \subset Z^2$, $l : \Omega \to \{1, \cdots, L\}$ is the true label of each pixel x, and L denotes the number of classes.

The weight map is computed as

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

where $w_c : \Omega \to R$ denotes the weight map to balance the class frequencies, $d_1 : \Omega \to R$ is the distance to the border of the nearest cell, and $d_2 : \Omega \to R$ denotes the distance to the border of the second nearest cell. In experiments, we set $w_0 = 10$ and $\sigma = 5$ pixels following the related research [10].

In this study, the original input images and their corresponding segmentation labeling (or ground truth segmentation) are used to train U-net for extracting the large vessels. For a test case, the input image is required for the U-net model and returns a 2D segmentation probability map. Figure 5 presents an example of U-net segmentation for large vessels. We can realize that the U-net model can obtain a good performance for large vessel segmentation because of its high contrast to the background. However, the model performance is limited for small vessels. Figure 5 illustrates that some small vessels are disconnected or broken up vascular segments due to their low contrast to the background and thin structures. Subsequently, to overcome this problem, we propose a coarse-to-fine algorithm-based U-net for detecting and extracting small and thin blood vessels.

2.2. Small Vessel Extraction Based on Coarse-to-Fine Algorithm-Based U-net. In the previous section, we represent the U-net approach to extract the main coronary vessels. However, it cannot reveal well the small vessels due to their blurring and low contrast compared with the background. To solve this problem, as shown in Figure 6, we use the information of the main extracted vessels and propose a new method to extract junctions on the vascular tree and extract small regions that included small vessel branches. Then, we apply a region-based U-net approach to segment small vessels based on a coarse-to-fine mechanism.

The branching geometry and junctions of the blood vessel tree are challenges in applying the coarse-to-fine U-net framework for vessel segmentation. The Zhang-Suen thinning algorithm [21] can be applied to extract the skeleton or central line of the main vessels. However, after segmentation, thin broken blood vessels may appear due to low contrast or low signal-to-noise ratio, leading to reduced performance. Therefore, we introduce an improved Zhang-Suen thinning algorithm to connect small broken blood vessels. We summarize this approach in Algorithm 1, and Figure 7 displays the result after applying this algorithm. Figure 8 presents the blood vessel's central line result based on an improved Zhang-Suen thinning algorithm.

The start, end, and junction nodes of the blood vessels are determined based on the central line of the large vessel segmentation result. In an X-ray angiogram, because of the huge

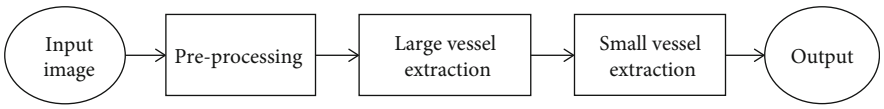


Figure 1: An illustration of the proposed method.

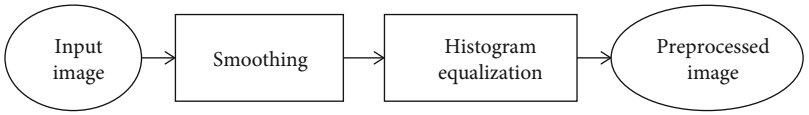


Figure 2: Preprocessing process.

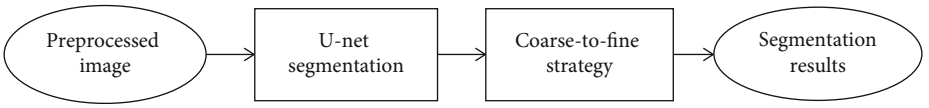


Figure 3: Vessel's extraction based coarse-to-fine segmentation.

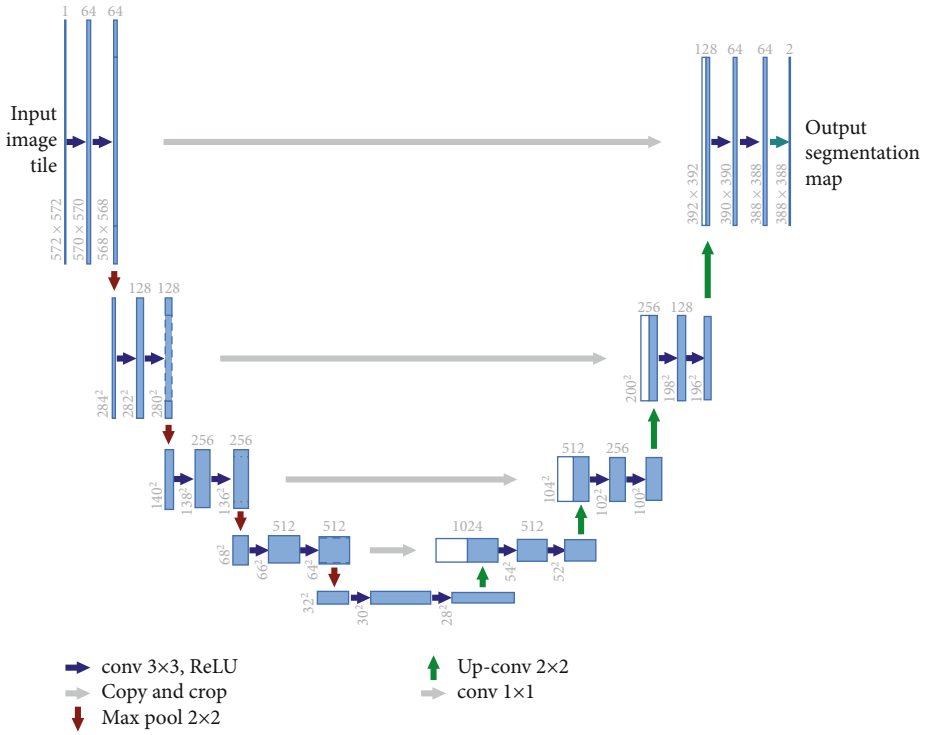
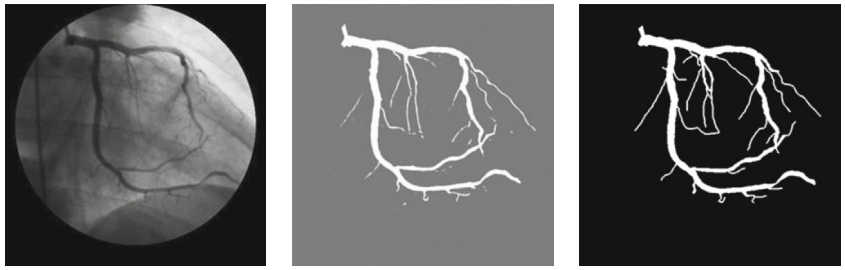


Figure 4: Illustration of the U-net architecture [10].

number of vessel branches, it is necessary to distinguish each blood vessel branch. Algorithm 2 describes a method to detect important nodes in the blood vessel. Given a central line image (skeleton binary image) of the large vessel segmentation result (output from Algorithm 1), object pixels (foreground) will have the value 1 (belonging to the blood vessel tree) and background pixels will have 0. For each pixel in the binary image, we classify each pixel (i, j) belonging to a particular label. Specifically, background pixels that have the value of 0 is classified into class 0 (or label 0). These background pixels are ignored while finding the important nodes.

Consider object pixels as the foreground, whether an object pixel has exactly two neighbour object pixels, this object pixel is considered a midpoint in the skeleton image (not the start, end, or junction points) and it is classified into class 1 and is ignored while finding the important nodes. Finally, an object pixel that has exactly one neighbour object pixel is considered start and end nodes and is classified into class 2; the object pixel has more than two neighbour object pixels, and it is a junction node and is classified into class 2. For each object pixel (i, j) in the skeleton image and having label 2, we find all neighbour object pixels of pixel (i, j) that were



(a) Input (b) U-net segmentation (c) Ground truth

Figure 5: An example of U-net segmentation result.

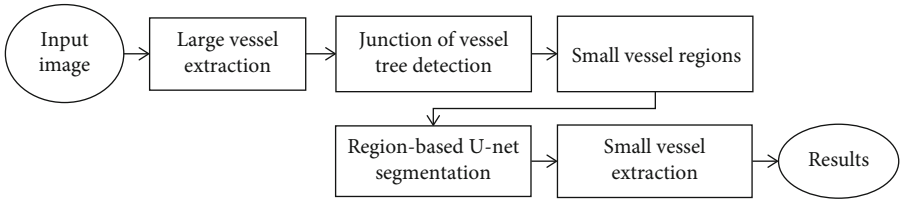
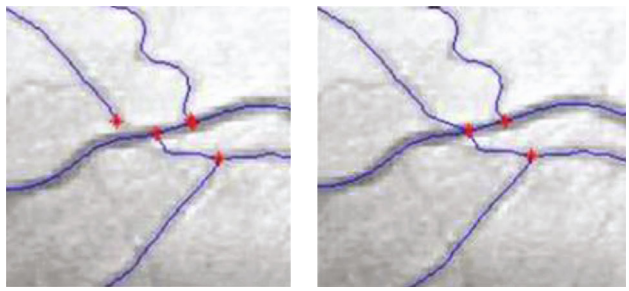


Figure 6: A block diagram for small vessel extraction.

Input: Binary image after applying large vessel extraction-based U-net
Output: Central line of vessels (output central line)
Step 1. Remove small regions less than γ pixels (γ ~116 pixels)
Step 2. Apply the baseline Zhang-Suen thinning algorithm to get the skeleton image, Γ
Step 3. Reconnect the broken segments in the Γ image
+ Find connected components in the Γ image
+ Find the largest connected component, LCC, in the Γ image
+ Initialize: output central line = LCC
+ For each remaining connected component (small component) in the Γ image, do
(i) Determine orientation (or direction) of small component to the LCC and connect each small connected component to the LCC.
(ii) Update: output central line = LCC

Algorithm 1: Vessel central line extraction from a binary image.

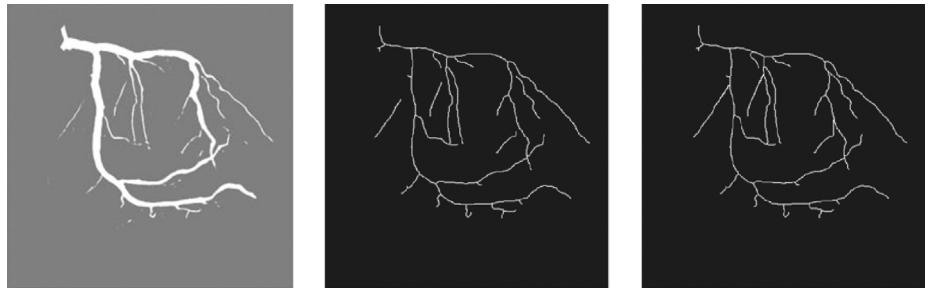


(a) (b)

Figure 7: An example of connecting the nearest the central lines: (a) input image and (b) image after connecting the nearest central lines (reproduced from Trinh et al. 2019 [under the Creative Commons Attribution License/public domain]).

classified into class 2 and then calculate their centroid point. The centroid points are considered the important nodes. Figure 9 demonstrates the determined nodes in the blood vessel tree.

Usually, the small vessels from an X-ray angiogram are blurring low-contrast images. It is difficult to extract large and small vessels simultaneously. For that reason, a local region-based segmentation approach should be used to extract the small ones. Based on the idea from local thresholding, we apply a region-based U-net to segment these small vessels. This approach helps reduce the effect of changing in grayscale values between the vessels and the background compared to the global approach. For each node in the blood vessel tree, we will construct a window between



Input image | Result of the baseline Zhang-Suen thinning algorithm | Result of the improve Zhang-Suen thinning algorithm

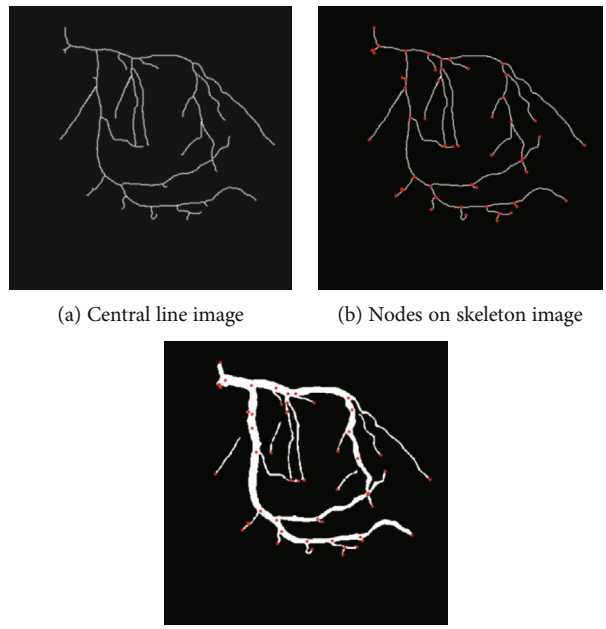
Figure 8: Results of the central lines of blood vessels based on the improved Zhang-Suen thinning algorithm. Input image. Result of the baseline Zhang-Suen thinning algorithm. Result of the improved Zhang-Suen thinning algorithm.

Input: The central line of the large vessel segmentation result (output central line)
Output: L: List of important nodes in the blood vessels
Step 1. Based on the image of the central line of a large vessel (binary image), we classify each pixel (i, j) into three classes (three labels) [0, 1, 2] as follows:

$$\text{label}(i,j) = \begin{cases} 0, \text{pixel}(i,j) \text{is background (have value of 0)}, \\ 1, \text{pixel}(i,j) \text{ is foreground and its two neighbour pixels are foreground}, \\ 2, \text{pixel}(i,j) \text{ is foreground and it has one or at least three neighbour pixels are foreground.} \end{cases}$$

Step 2. Find the important nodes (start, end, and junction nodes)
While $pixel(i, j) \in$ output central line (skeleton image) and $label(i, j) == 2$:
+ Find neighbour object pixels of pixel (i, j) that were classified into class 2 (label 2), and then calculate centroid point of them.
+ node ← centroid point
+ L.append(node) (add determined node into List L)
End

Algorithm 2: Detecting important nodes in the blood vessel tree.



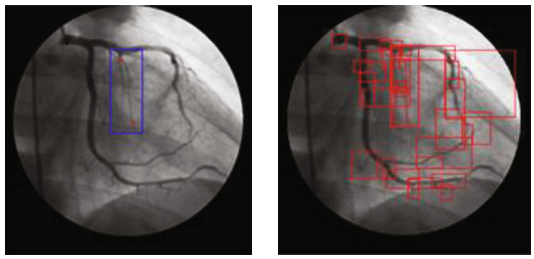
(a) Central line image (b) Nodes on skeleton image

(c) Nodes on main vessel segmentation

Figure 9: An example of nodes (starting, end, and junction points) in the blood vessel tree.

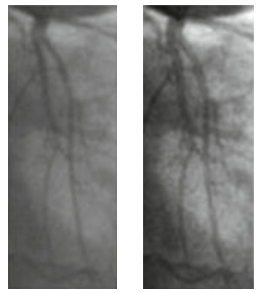
Input: L: List of important nodes in the blood vessels, the central line image (skeleton binary image), label of object pixels.
Output: B: List of rectangles including small blood vessels
Step 1. Find blood vessel segments, edges
  Init: edges = ∅
  Visit every node of list L:
    edge = ∅
    Repeat
      (i) Find neighbour object pixel (called as *nb*) of the current *node* (the pixel with label of 1 in the central line image),
      (ii) Update: *node* = *nb*,
      (iii) edge.append(*node*)
    Until *nb*∈ *L*
    edges.append(edge)
Step 2. Find the top-left and bottom-right coordinates of rectangle
  Init: *i* =0; *B* = ∅
  For edge ∈ edges:
      (i) Find two points, p1 and p2, so that p1 is the top-left point and p2 is the bottom-right point of a rectangle that includes the largest blood vessel region based on equations (3) and (4).
      (ii) B[i].append(p1, p2)
      (iii) *i* = *i*+1
  End

Algorithm 3: Small blood vessel region detection.



(a) A single region (b) Small vessel regions

Figure 10: An example of a region constructed between two nodes.



(a) (b)

Figure 11: An example of contrast enhancement in the region analysis: (a) input image and (b) result from contrast enhancement (reproduced from Trinh et al. 2019 [under the Creative Commons Attribution License/public domain]).

nodes *i* and *j*. The width (*w*) and height (*h*) of the window are described by

$$w = \left|\left(x_i - x_j\right)\right| + \text{bias}, \tag{3}$$

$$h = \left|\left(y_i - y_j\right)\right| + \text{bias}. \tag{4}$$

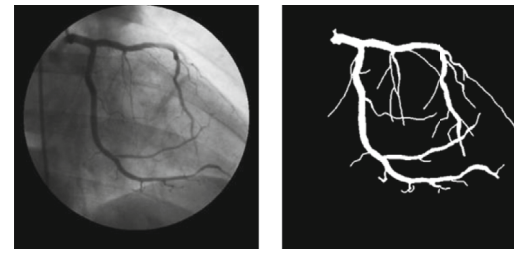
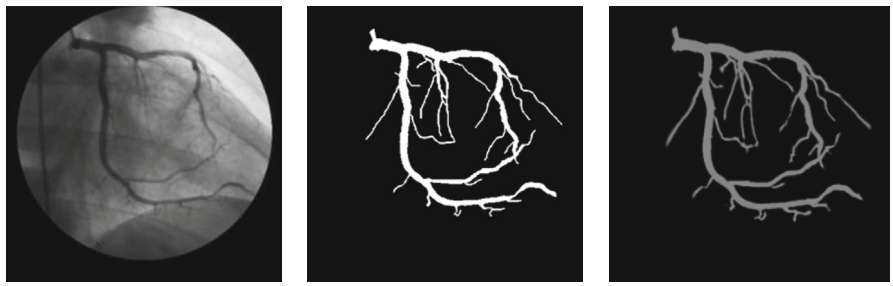


Figure 12: Ground truth of an X-ray angiogram image.

In our experiment, we select a bias of 20 pixels for obtaining small vessels near node *i*. The proposed approach focuses on determining regions of small blood vessels as summarized in Algorithm 3. Figure 10 describes an example of a local region that includes small vessels. In the local region, there exists a large vessel with high intensity and high contrast to the background compared to small vessels. Thus, we remove the effect of the large vessel in the window and then apply contrast adjustment based on image processing to areas that include the small vessels. Figure 11 presents a contrast enhancement based on image processing in the small region. Additionally, we apply the region-based U-net approach to segment these small vessels.

3. Results and Discussion

3.1. Dataset. All of the experiments were conducted on the X-ray angiogram database of the coronary vessel, which was collected and supported by local hospitals. The database contains 48 different vessel images corresponding to two categories: D1 and D2. The size of each image is 512 × 512 pixels, with 256 gray levels per pixel. The D1 dataset consists of 20 images that obtain a direct front view of the coronary vessels.



(a) (b) (c)

Figure 13: Segmentation result of the proposed approach: (a) input image; (b) ground truth; (c) segmented blood vessels.

The D2 dataset includes 28 images taken from four different angles of the coronary vessels [8].

Our dataset is divided into 40 images for training and 8 images for testing. During the training process, we use data augmentation methods to enhance the performance of the segmentation result. This method allows the network to become invariant and robust to certain transformations when the size of the training set is limited. For example, rotation, flip, and shear operators are usually used for convolutional neural networks and yield the desired invariance and robustness properties of the resulting network. In our experiment, the augmentation was applied using the ImageDataGenerator function implemented in Keras.

3.2. Experimental Environment. In the experiments, we use the software MIPAR of Sosa [22] to create ground truth in order to evaluate the performance of the segmentation algorithm. We compute the Dice similarity coefficient [23] between binary segmentation results and the ground truths to evaluate the accuracy of our system. Figure 12 shows a sample image and its ground truth. Our experiments are implemented on an Intel® Xeon® E5-2630, CPU @ 2.3GHz with 128 GB RAM, 4 GPU NVIDIA Geforce GTX 1080Ti - Vram 11 GB (CUDA 6.1). The average runtime of the proposed algorithm to be applied to each image is 66.67 ms. In the large blood vessel extraction procedure, we use the U-net model with 64 filters for the first convolutional layer, followed by the ReLU activation function, we set the learning rate of $1e-4$, and the sigmoid function is used as the final activation function. The training process of region-based U-net is similar to that of the original U-net model. The region-based U-net is a small version of the original U-net. Particularly, it is modified with a small number of filters of 16 for the first convolutional layer to deal with small input images (including small vessel regions). The small vessel regions from the same original training set are used to train a region-based U-net model.

In the small blood vessel extraction procedure, we use the U-net model with 16 filters for the first convolutional layer, followed by the ReLU activation function, we set the learning rate of $1e-4$, and the sigmoid function is applied for the final activation function.

4. Results

This research investigates the coronary vessel segmentation performance based on the coarse-to-fine strategy-based U-net for iterative segmentation comprising other approaches. Figure 13 illustrates a segmentation result of the proposed approach. We found through experimental analysis that our method segments the large coronary vessels significantly. The performance of size-independent coronary vessel segmentation attains 80.17%. Besides, our method reveals a large number of small and thin blood vessels. Finally, the proposed approach obtains the average of the performance of coronary vessel segmentation of 76.40%. We also compare the proposed approach's performance with the baseline U-net and the other techniques in [7, 8] on our database. Figure 14 shows a comparison of segmentation results. Table 1 describes a summary of the coronary vessel segmentation performance in terms of the Dice coefficient. The experimental results are described as mean ± standard deviation.

From Figure 14 and Table 1, we realize that our method using the hierarchical approach based on deep learning and coarse-to-fine strategy obtains better segmentation results and outperforms the standard approaches. In Figure 14, we can realize that the DFB-based segmentation [7] leads to more artifacts and fails to enhance small vessels compared to our approach correctly. Furthermore, it cannot detect the small vessels that have low intensity and large vessels with missing parts. Our method detects the large and small vessels at the same time; even in the case of existing large difference in intensity between the large vessels (which are high-contrast objects) and the small vessels (which are low-contrast objects), the DFB-based method cannot significantly extract small vessels. The proposed method in [8] can extract large blood vessels very well, but it leads to missed extraction of small and thin vessels and vessels with low contrast to the background. Experimental analysis indicated that the baseline U-net yields higher accuracy than the traditional DFB and Otsu approach. The U-net can obtain very high accuracy for the main vessels. However, it also leads to missed extraction of small ones. Our proposed method-based U-net and coarse-to-fine strategy-based segmentation provide the optimal performance.

Our method is proposed to overcome the problem by separately detecting the large vessels and small vessels based

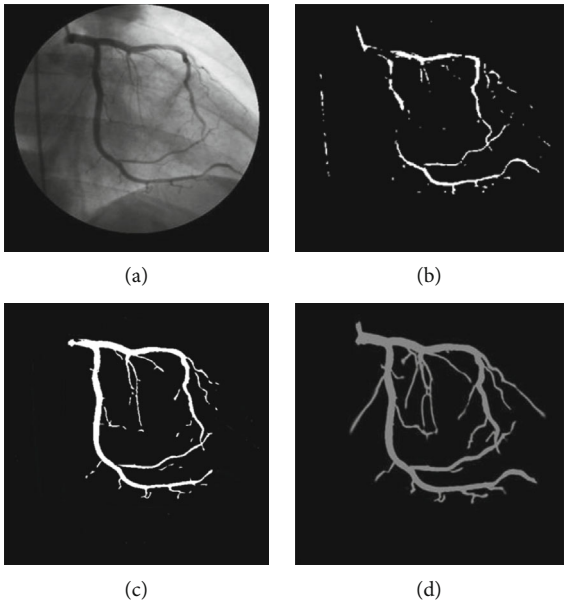


Figure 14: Comparison of the segmentation results: (a) input image; (b) result in [7]; (c) result in [8]; (d) result of our method.

Table 1: Performance comparison of the coronary vessel segmentation in terms of Dice coefficient.

Method	Dice coefficient (%)
DFB-based segmentation [7]	45.50 ± 1.31%
Coarse-to-fine-based DFB and Otsu [8]	71.34 ± 0.80%
Baseline U-net [10]	73.64 ± 1.32%
Proposed approach	76.40 ± 1.02%

on a hierarchical technique via the U-net model. Because we consider that the large vessels always have high contrast to the background than the small vessels, the U-net model is suitable for extracting them. The coarse-to-fine strategy-based segmentation guarantees that the method can correctly extract the large vessels. In the small vessel extraction stage, we first reduce the effect of large vessels and make a contrast enhancement on the region that includes small vessels. This deals with the low-contrast problem on small vessel regions. When the small vessel regions have increased the contrast, they were easily detected and segmented by U-net. This is significant to extract small vessels. The experimental results show that our method overcomes the limitations of the standard approaches, such as small vessel intensity and noise sensitivity. It also performs better on real angiography images.

However, most errors occurred while processing small vessels. These errors cause contrast enhancement based on an image processing technique and quality of small images. In particular, the traditional contrast enhancement approach has errors due to the background enhancement with fewer artifacts. In our cases, several small vessel images are affected by illumination and noises, such as low-light conditions and low contrast. The traditional contrast enhancement approach cannot deal with all problems leading to reducing the accuracy of our segmentation system. Figure 15 shows a small vessel image affected by illumination and contrast enhancement. Some small vessel branches are missing.

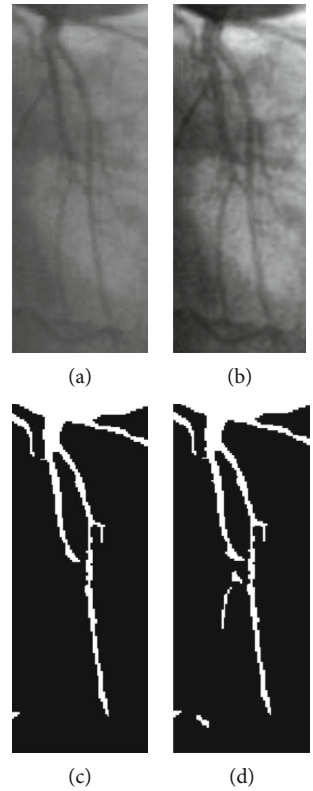


Figure 15: Under segmentation error of small vessel image due to the effect of illumination: (a) input image; (b) contrast enhancement image; (c) result using U-net; (d) result of our method.

Our research contains other limitations rather than the dependence on traditional methods for contrast enhancement. There is a limitation in the number of public datasets of X-ray angiograms of the coronary vessels. Researchers have limited access to X-ray angiograms of coronary vessel data. In addition, this research is a proof-of-concept study and limited by the size of the dataset. Our dataset is considered small for developing a completed medical image-based deep learning application. Medical image segmentation-based deep learning requires sufficient data to obtain higher accuracy than traditional systems.

5. Conclusions

We introduce an improved coronary vessel segmentation technique by a hierarchical approach based on the coarse-to-fine strategy for iterative segmentation using U-net architecture. Our method not only segments the main blood vessels but also locates and extracts the small and thin vessel branches. Through experiments results, it has been confirmed that our proposed method is effective and can enhance the performance of vessel segmentation. However, small vessel images are missing due to enhancing the background with fewer artifacts when these images are applied to contrast enhancement based on the traditional image processing technique. In the future, we intend to improve the

results for small and thin vessels by exploiting the superpixel-based deep learning approach to enhance the quality of small vessel image and explore other deep learning frameworks for coronary vessel segmentation and an extended method to deal with 3D images.

Data Availability

We use private data from the hospital.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. R. Fazlali, N. Karimi, S. M. R. Soroushmehr et al., "Vessel segmentation and catheter detection in X-ray angiograms using superpixels," *Medical & Biological Engineering & Computing*, vol. 56, no. 9, article 1793, pp. 1515–1530, 2018.

[2] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys*, vol. 36, no. 2, pp. 81–121, 2004.

[3] R. Liao, D. Luc, Y. Sun, and K. Kirchberg, "3-D reconstruction of the coronary artery tree from multiple views of a rotational X-ray angiography," *The International Journal of Cardiovascular Imaging*, vol. 26, no. 7, pp. 733–749, 2010.

[4] C. Yan, S. Hirano, and Y. Hata, "Extraction of blood vessel in CT angiography image aided by fuzzy logic," in *WCC 2000 - ICSP 2000. 2000 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000*, vol. 2, pp. 926–929, Beijing, China, 2000.

[5] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*, vol. 1496, pp. 130–137, 1998.

[6] G. Agam, S. G. Armato, and Changhua Wu, "Vessel tree reconstruction in thoracic CT scans with application to nodule detection," *IEEE Transactions on Medical Imaging*, vol. 24, no. 4, pp. 486–499, 2005.

[7] P. T. H. Truc, M. A. Khan, Y. K. Lee, S. Lee, and T. S. Kim, "Vessel enhancement filter using directional filter bank," *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 101–112, 2009.

[8] T. D. Trinh, T. B. Tran, L. N. L. Thuy, I. Shimizu, and J. Y. Kim, "Coronary vessel segmentation by coarse-to-fine strategy using Otsu algorithm and decimation-free directional filter bank," *Journal of IKEEE*, vol. 23, no. 2, pp. 557–570, 2019.

[9] N. Ghavami, Y. Hu, E. Gibson et al., "Automatic segmentation of prostate MRI using convolutional neural networks: investigating the impact of network architecture on the accuracy of volume measurement and MRI-ultrasound registration," *Medical Image Analysis*, vol. 58, article 101558, 2019.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference*, vol. 9351, pp. 234–241, Munich, Germany, 2015.

[11] F. Milletari, N. Navab, and S. A. Ahmadi, "V-net: fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, Stanford, CA, USA, 2016.

[12] L. Fidon, W. Li, L. C. Garcia-Peraza-Herrera et al., "Generalised Wasserstein Dice score for imbalanced multi-class segmentation using holistic convolutional networks," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers*, vol. 10670, pp. 64–76, Springer, Cham, 2018.

[13] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep vessel segmentation by learning graphical connectivity," *Medical Image Analysis*, vol. 58, article 101556, 2019.

[14] X. Zhou, Q. Ye, Y. Jiang et al., "Systematic and comprehensive automated ventricle segmentation on ventricle images of the elderly patients: a retrospective study," *Frontiers in Aging Neuroscience*, vol. 12, 2020.

[15] D. Hao, S. Ding, L. Qiu et al., "Sequential vessel segmentation via deep channel attention network," *Neural Networks*, vol. 128, pp. 172–187, 2020.

[16] T. J. Jun, J. Kweon, Y. H. Kim, and D. Kim, "T-Net: nested encoder-decoder architecture for the main vessel segmentation in coronary angiography," *Neural Networks*, vol. 128, pp. 216–233, 2020.

[17] P. M. Samuel and T. Veeramalai, "VSSC Net: vessel specific skip chain convolutional network for blood vessel segmentation," *Computer Methods and Programs in Biomedicine*, vol. 198, article 105769, 2021.

[18] J. Zhang, G. Wang, H. Xie et al., "Weakly supervised vessel segmentation in X-ray angiograms by self-paced learning from noisy labels with suggestive annotation," *Neurocomputing*, vol. 417, pp. 114–127, 2020.

[19] Y. Jin, G. Yang, Y. Fang et al., "3D PBV-Net: an automated prostate MRI data segmentation method," *Computers in Biology and Medicine*, vol. 128, article 104160, 2021.

[20] M. Li, C. Wang, H. Zhang, and G. Yang, "MV-RAN: multiview recurrent aggregation network for echocardiographic sequences segmentation and full cardiac cycle analysis," *Computers in Biology and Medicine*, vol. 120, article 103728, 2020.

[21] R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice-Hall, New Jersey, USA, 2002.

[22] J. Sosa, "MIPAR - premier image analysis & image segmentation software," 2021, https://www.mipar.us/.

[23] K. H. Zou, S. K. Warfield, A. Bharatha et al., "Statistical validation of image segmentation quality based on a spatial overlap index1," *Academic Radiology*, vol. 11, no. 2, pp. 178–189, 2004.

Hindawi
BioMed Research International
Volume 2021, Article ID 6681092, 14 pages
https://doi.org/10.1155/2021/6681092

Research Article

Hepatic Alveolar Echinococcosis: Predictive Biological Activity Based on Radiomics of MRI

Bo Ren,[1] **Jian Wang**,[1] **Zhoulin Miao**,[1] **Yuwei Xia**,[2] **Wenya Liu**,[1] **Tieliang Zhang**,[1] **and Aierken Aikebaier**[1]

[1]*Department of Imaging Center, The First Affiliated Hospital of Xinjiang Medical University, Li Yu Shan Road, No. 137 Urumqi City 830054, China*
[2]*Huiying Medical Technology Co., Ltd., Room A206, B2, Dongsheng Science and Technology Park, HaiDian District, Beijing City 100192, China*

Correspondence should be addressed to Wenya Liu; 13999202977@163.com

Received 31 December 2020; Revised 6 March 2021; Accepted 17 March 2021; Published 9 April 2021

Academic Editor: Changming Sun

Background. To evaluate the role of radiomics based on magnetic resonance imaging (MRI) in the biological activity of hepatic alveolar echinococcosis (HAE). *Methods*. In this study, 90 active and 46 inactive cases of HAE patients were analyzed retrospectively. All the subjects underwent MRI and positron emission tomography computed tomography (PET-CT) before surgery. A total of 1409 three-dimensional radiomics features were extracted from the T2-weighted MR images (T2WI). The inactive group in the training cohort was balanced via the synthetic minority oversampling technique (SMOTE) method. The least absolute shrinkage and selection operator (LASSO) regression method was used for feature selection. The machine learning (ML) classifiers were logistic regression (LR), multilayer perceptron (MLP), and support vector machine (SVM). We used a fivefold cross-validation strategy in the training cohorts. The classification performance of the radiomics signature was evaluated using receiver operating characteristic curve (ROC) analysis in the training and test cohorts. *Results*. The radiomics features were significantly associated with the biological activity, and 10 features were selected to construct the radiomics model. The best performance of the radiomics model for the biological activity prediction was obtained by MLP (AUC = 0.830 ± 0.053; accuracy = 0.817; sensitivity = 0.822; specificity = 0.811). *Conclusions*. We developed and validated a radiomics model as an adjunct tool to predict the HAE biological activity by combining T2WI images, which achieved results nearly equal to the PET-CT findings.

1. Introduction

Hepatic alveolar echinococcosis (HAE) is a parasitic disease caused by the larvae of *Echinococcus multilocularis* that parasitize the liver [1]. HAE causes lesions that are infiltrative and may spread to distant regions of the body, impairing health and may cause death. In its initial stages, HAE is idiopathic; thus, most patients are diagnosed late, rendering them unsuitable for radical resection surgery as they already have large hepatic lesions with vascular or biliary structure involvement [2]. The only choice of treatment for patients who are not candidates for radical resection surgery and those undergoing palliative resection involves antihydatid therapy with drugs such as albendazole [3]. However, there is no well-defined treatment period for HAE, which leads to long-term medication-associated complications in patients [4, 5]. Theoretically, the critical indicator for medication termination involves the absence of biological activity of HAE lesions; however, this criterion is not clinically feasible. As a result, assessing the biological activity of the HAE lesions is vital for the selection and design of treatment methods, including antiechinococcal chemotherapy for patients before and after surgery. It is against this background that determining the state of HAE lesions, whether active or inactive, is the primary goal of imaging procedures in clinical practice.

Although the proliferation and biological activity of HAE lesions can be evaluated by CT perfusion, energy CT imaging, and diffusion-weighted MRI [6–9], PET-CT is the most

prioritized method globally for assessing HAE lesions [10, 11]. However, compared with CT and PET-CT, MRI has the advantages of having no radiation and being noninvasive. Moreover, MRI shows better tissue contrast and shows the small vesicle structure of the lesions [6, 7]. Furthermore, MRI can detect small lesions in the early stage noninvasively and does not use radiation; therefore, MRI is used as the preferred imaging examination for HAE lesions in patients.

Research has shown that radiomics data analysis can provide vital quantitative imaging information to quantitatively and objectively analyze tumors and other lesions [12]. Accordingly, radiomics has been successfully applied in the diagnosis, treatment, and evaluation of multiple tumor types in the medical field [13–18]. Currently, radiomics research on HAE is in its early stages. This study is aimed at extracting high-throughput features through HAE lesion segmentation, dimensional reduction analysis, and training machine learning to establish prediction models for prognosis, diagnosis, and monitoring of HAE lesions in patients.

2. Materials and Methods

2.1. Data Collection. This was a retrospective study in a single institution, approved by the Medical Ethics Review Committee of the First Affiliated Hospital of Xinjiang Medical University, and exempted from informed consent. From January 2012 to June 2020, 156 patients with HAE were admitted and diagnosed at the First Affiliated Hospital of Xinjiang Medical University. In this study, the PET-CT findings were considered as the "gold standard" to assess whether the lesions have biological activity or not. After that, a predictive model based on MRI was constructed to predict the biological activity of HAE lesions as a basis for prognosis, diagnosis, and monitoring of HAE lesions in patients. HAE patients (confirmed by imaging and postoperative pathology) who underwent abdominal MRI scan and PET-CT examination (images were transferred to PACS), with no history of chronic liver disease, with no previous history of liver surgery, and with no primary solitary space-occupying lesion were included. Conversely, HAE patients whose MRI and PET-CT image quality were poor ($n = 2$); whose PET imaging results were lacking ($n = 15$); who had extensive fibrosis, nodules, or old lesions in the liver ($n = 1$); and who previously confirmed and were already treated by surgical intervention ($n = 2$) were excluded. As a result, 136 patients were enrolled in this study. According to the results of PET-CT, the patients were divided into the active group (90 cases) and the inactive group (46 cases).

2.2. Image Data Acquisition. In this study, MRI was performed using the Siemens 3.0 T (Skyra) or 1.5 T (Avanto) MR scanner with an 18/8-channel phased array body coil. All patients were asked to fast for about six hours before scanning, after which they underwent upper abdomen MRI examination in the supine position. All patients underwent MR imaging with T1-weighted, T2-weighted, and fat-suppressed T2-weighted image delineation. The MR imaging protocol was slightly adjusted due to different devices causing minor adjustments to the parameters. The MR scan sequences were as follows:

(1) 3.0 T

(a) T1WI: TR/TE = 400/8.0 ms; FOV = 320 mm × 320 mm; matrix = 320 × 192; NEX = 2.0; ST = 3.0 mm

(b) T2WI: TR/TE = 4000/125 ms; NEX = 4.0; slice thickness = 3.0 mm

(2) 1.5 T

(a) T1WI: TR/TE = 200/4.5 ms; matrix = 204 × 256

(b) T2WI: TR/TE = 3500/100 ms

After the acquisition of 3D data, the attenuation correction of the PET image was performed based on the CT image, and the corrected PET image was automatically fused with the CT image to obtain axial, coronal, sagittal, and PET-CT fusion images. Two nuclear medicine doctors with more than ten years of experience in the diagnosis of PET-CT examined the images. In case of disagreement, the negotiated results were considered. According to the information provided by the MRI images, the SUV value was measured at the corresponding position. For each patient, two independent examinations were performed and completed within a week.

2.3. Radiomics Workflow. Figure 1 illustrates the radiomics workflow adopted in this study. It included image collection; lesion segmentation; and radiomic feature extraction, selection of features, construction of models in the training cohorts, and evaluation of the performance of prediction models in the test cohorts.

2.4. Image Preprocessing and HAE Lesion Segmentation. T2WI data in DICOM format is uploaded to the Radcloud platform (version 3.1.0, http://radcloud.cn/, Huiying Medical Technology Co., Ltd., Beijing, China). As MR scanning with different field intensities is used, image preprocessing is required to obtain more robust radiomics features. Image preprocessing consists of two steps.

Step 1. We used the following formula to normalize the intensity of the image to minimize the change in MRI intensity collected by machines with different parameters (i is the original intensity; $F(i)$ shows the normalized intensity; μ_i is the mean value of the image intensity values; σ_i indicates the standard deviation of the image intensity values; and s is an optional zoom and is set to 1 by default). Normalization is for the whole image, not just the region of segmentation.

$$F(i) = \frac{s(i - \mu_i)}{\sigma_i}. \quad (1)$$

Step 2. In order to eliminate the intrinsic dependence of radiomics features on voxel size, the resampling method with a linear interpolation algorithm was used to normalize voxel size.

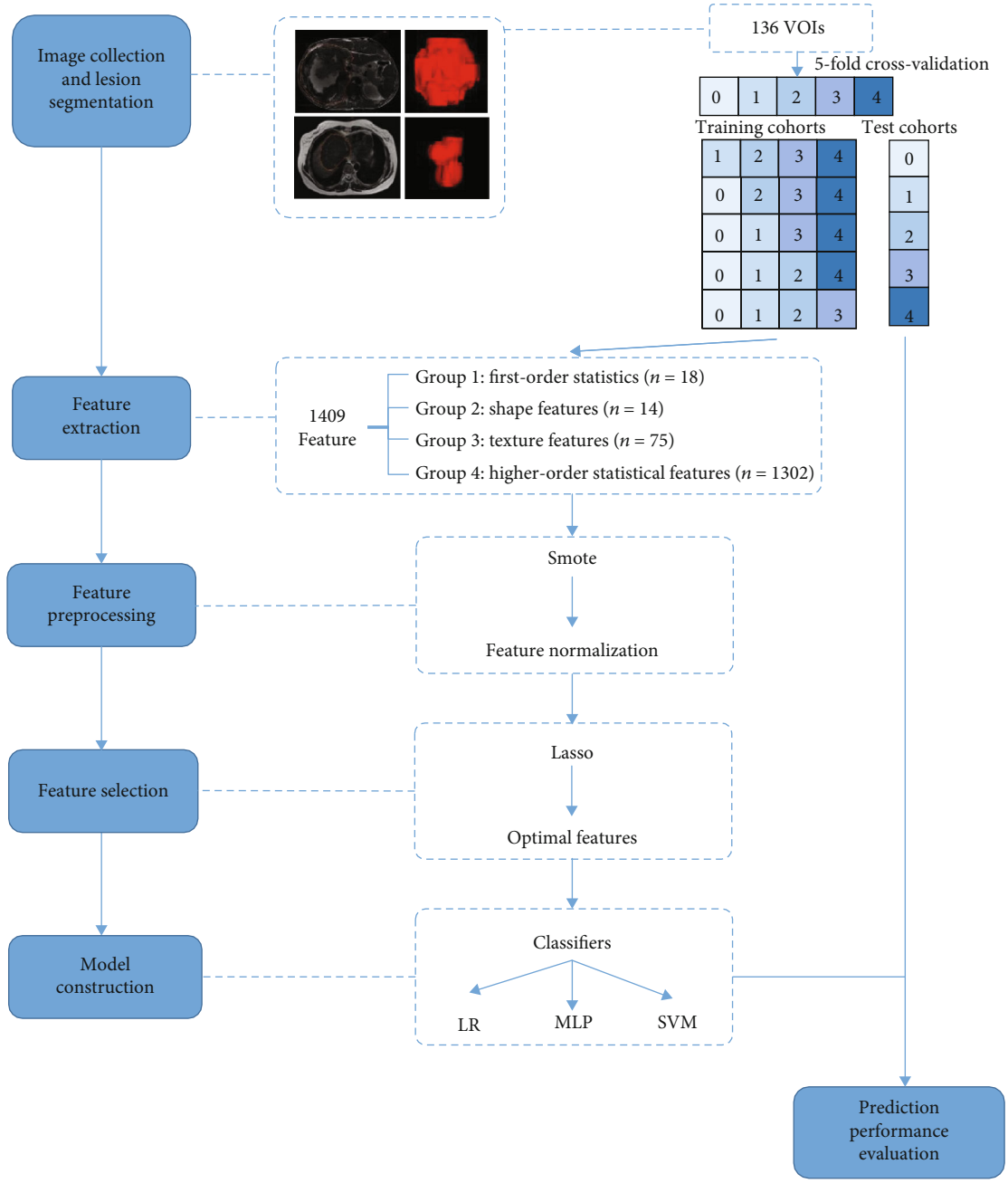
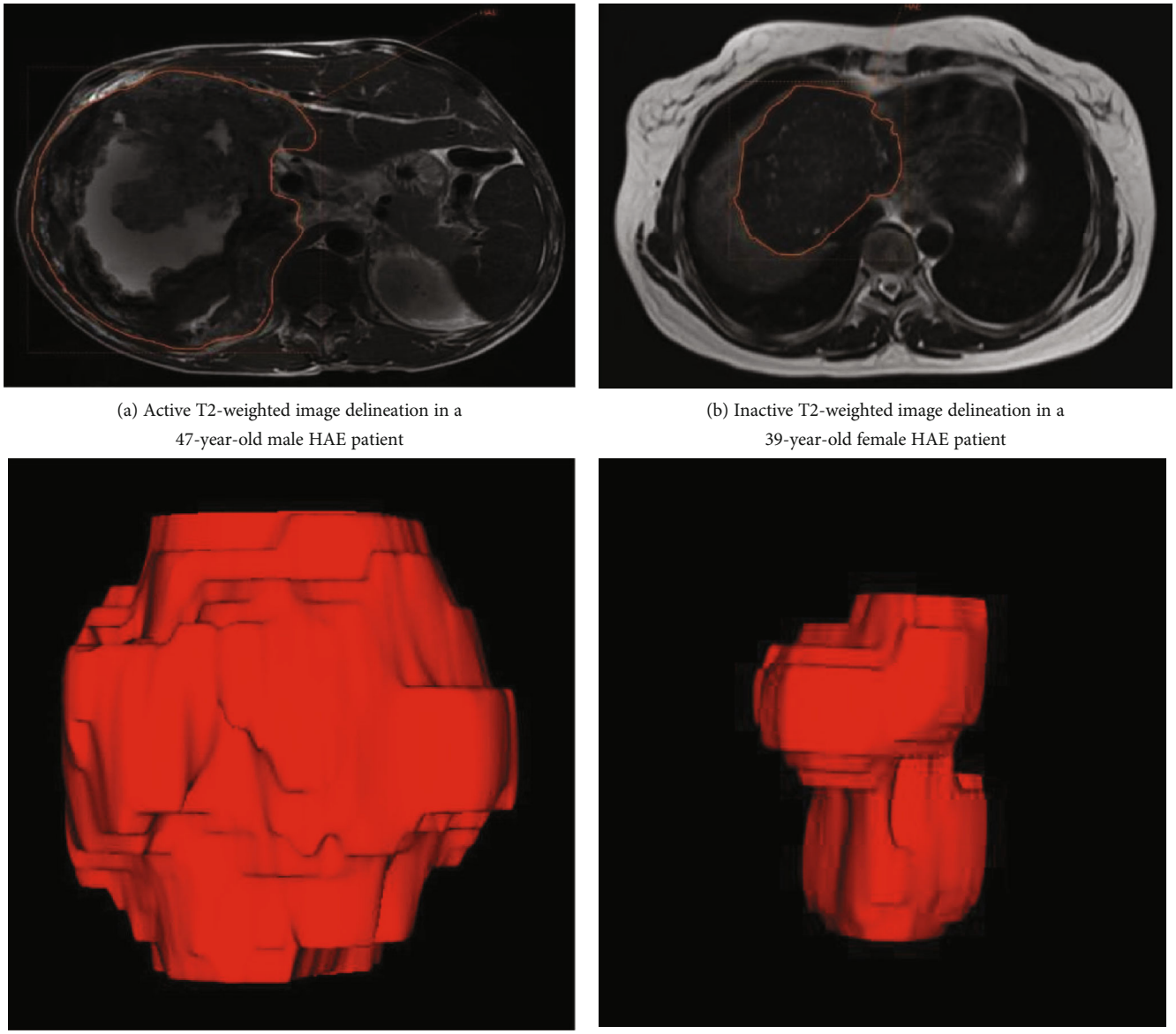


Figure 1: Illustration of the radiomics workflow adopted in this study. Note: SMOTE (synthetic minority oversampling technique); LASSO (least absolute shrinkage and selection operator); LR (logistic regression); MLP (multilayer perceptron); SVM (support vector machine).

As shown in Figure 2, to obtain the volume of interest (VOI) for further analysis, four radiologists manually delineated the region of interest (ROI) along the edge of the lesion, layer by layer, on each T2WI. All depicted regions of interest (ROI) in T2WI were strictly delineated with the same criteria and visually validated by the same expert (with 10 years of experience in abdomen MRI). Then, the 3D VOI of the lesion is generated automatically by computer interpolation.

2.5. Feature Extraction. In this study, the "pyradiomics" package (version 2.1.2, https://pyradiomics.readthedocs.io/) in Python was used to extract 1409 radiomics features in the VOI of each T2WI. The features can be divided into four categories: shape features, first-order gray histogram features, second-order texture features, and higher-order features based on filter transformation. Shape features ($n = 14$) reflect the three-dimensional size and shape of a given VOI, including mesh volume, surface area, surface area to volume ratio, sphericity, compactness and spherical disproportion, elongation, and flatness. First-order gray histogram features ($n = 18$) reflect the overall information of the histogram, including energy, minimum, 10th/90th percentile, maximum, mean, median, standard deviation, range, mean absolute deviation, and entropy. Second-



(a) Active T2-weighted image delineation in a 47-year-old male HAE patient

(b) Inactive T2-weighted image delineation in a 39-year-old female HAE patient

(c) A 47-year-old male HAE patient with active computer-generated 3D VOI

(d) 39-year-old female HAE patient without active computer-generated 3D VOI

Figure 2: A representation of the manual segmentation in the T2-weighted images.

order texture features include the gray-level cooccurrence matrix (GLCM, $n = 24$), gray-level size zone matrix (GLSZM, $n = 16$), gray-level dependence matrix (GLDM, $n = 14$), neighborhood gray-level dependence matrix (NGLDM, $n = 5$), and gray-level run-length matrix (GLRLM, $n = 16$). The second-order texture features can respond to the image pixels at a certain level of relative distribution from the side, so they expound the complexity and heterogeneity within the lesion. High-order filter transform features ($n = 1302$) also include the original image through the filter transform to get the intensity and texture feature, through the neighborhood grayscale difference matrix and gray areas such as the size of the matrix computation, using seven kinds of filters: logarithm filter, exponential filter, gradient filter, square filter, square root filter, local binary pattern (LBP) filter, and wavelet filter. Features are compliant with definitions as defined by the Imaging Biomarker Standardization Initiative (IBSI) [19].

2.6. Subsampling. The experiment took into account the imbalance between the active and inactive HAE groups, which does not satisfy the balanced endpoint hypothesis of most machine learning-based prediction models. To tackle this problem, we use the synthetic minority oversampling technique (SMOTE) for subsampling. It is important to note, however, that the synthesized new data appears only in the training cohorts and not in the test cohorts.

2.7. Feature Selection and Model Construction. Prior to the steps of feature selection, because the range of radiomics features of different properties varies greatly, the normalization of radiomics features ensures the convergence of the training model. At the same time, in order to avoid model overfitting, fivefold cross-validation was performed during the experiments. In each fold of the training cohorts, the least absolute shrinkage and selection operator (LASSO) feature selection algorithm was used to select the relevant features and calculate the correlation coefficient of the selected features. The ten most valuable features with the highest correlation coefficient are retained as the final feature subset.

We used multiple ML algorithms to test the impact of different machine learning models on the predictive performance, including logistic regression (LR), multilayer perceptrons (MLP), and support vector machine (SVM). We developed diagnostic classifiers based on quantitative image grouping features (by using the T2WI features selected from the training queue) and quantitative radiomics features (by using the T2WI features selected in a training cohort). The evaluation indicators included the receiver operating characteristic curve (ROC curve) with indices of area under the curve (AUC), 95% confidence level (95% CI, AUC), accuracy, sensitivity, and specificity.

2.8. PET-CT Observation Items and Evaluation Criteria. The standardized uptake value (SUV) was calculated automatically by semiquantitative analysis in the workstation to judge the FDG uptake of the lesions according to the SUV value. Prior to analysis for measurement of SUV values, the basic information of each case and the location, size, and Kodama classification of the lesions were registered in detail to ensure that the lesions measured on PET-CT images and those segmented on MRI were one lesion. Most of the HAE lesions showed elevated cyclic glucose metabolism on PET-CT, and no hyperglycemia was found in the lesions. If the SUV of the lesion is higher than that of the surrounding liver at the same plane, the lesion is judged to be active; if the SUV value of the lesion is lower than that of the surrounding liver, the lesion is not active.

2.9. Statistical Analysis. The study was performed using the programming language Python 3.6 (https://www.python.org/). The packages of "pyradiomics" (https://pyradiomics.readthedocs.io/), "scikitlearn" (https://scikit-learn.org/), and "matplotlib" (https://matplotlib. org/) were used for feature selection, model building, and plotting in this study. Another statistical analysis was performed with SPSS 17.0 and MedCalc15.2.2. P value < 0.05 was considered statistically significant. ROC curve analysis was used to evaluate the diagnostic performances of ML classifiers.

3. Results

3.1. Demographic Data of the HAE Patients. Among the 136 patients, there were 64 males (47%) and 72 females (53%), with an average age of 39 ± 13 years. They were comprised of the Kazakh, Uygur, Han, and Tibetan ethnic groups. There was no significant difference in gender, age, lesion location, lesion size, and other clinical characteristics between the observation group and the control group ($P > 0.05$; Table 1).

Table 1: Demographic data of the HAE patients.

Patient attributes	Active group	Inactive group	P value
n	90	46	
Age (mean ± SD, yr)	39 ± 13	38 ± 14	0.847
Gender			0.816
Male	43	21	
Female	47	25	
Location of lesions			0.264
Less than 3 liver segments	18	14	
3-6 liver segments	67	28	
More than 6 liver segments	5	4	
Lesion size (mm^3)	1388844.180	1357771.448	0.926

3.2. Feature Selection of Radiomics. In this study, 1409 features were obtained from each T2WI VOI image, and the data were divided into five groups of different training and test sets using a fivefold cross-validation method (the data set was divided into 5 parts, 4 of which were training sets and 1 was a test set). The dimensionality of each training set was analyzed by the LASSO method, which adopted a 10-fold intragroup cross-validation strategy, and the maximum iteration times of the model. To avoid model overfitting problems caused by a high-feature dimension, an alpha with the least mean square error in cross-validation was selected, and the corresponding correlation coefficient was calculated. The absolute value of the correlation coefficient was further determined by ranking the correlation coefficients of features for groups with more than 10 features after LASSO feature screening. The 10 characteristics of HAE lesions are shown in Table 2.

A bar graph was constructed using the sum of the best feature coefficients in the 5 groups of experiments with fivefold cross-validation, as shown in Figure 3. The results showed that optimal features were the first-order statistical features ($n = 2$) and texture features ($n = 1$) on the original image. Also, optimal first-order statistical features ($n = 16$) and texture features ($n = 29$) after wavelet transformation of the first-order statistical features of the maximum operator (wavelet-HLH_firstorder_Maximum) were obtained. The results suggested that cumulative maximum correlation coefficients can be used as biomarkers for effective radiomics to assess HAE characteristics.

3.3. Diagnostic Performance of the Radiomics Models. A variety of machine learning algorithms were used to train the model using a fivefold cross-validation procedure. Figures 4 and 5 show the ROC curves of the three machine learning classifier models. Tables 3 and 4 summarize the diagnostic performance and model cutoff values of the three machine learning classifier models. In general, all three machine

Table 2: 5-fold cross-validation for the best 5 group of best features.

Experimental group	0-fold	1-fold	2-fold	3-fold	4-fold
Feature name	wavelet-HLH_firstorder_Maximum	wavelet-HLL_glszm_SmallAreaEmphasis	wavelet-LHL_glcm_ClusterShade	squareroot_firstorder_InterquartileRange	wavelet-LHL_gldm_HighGrayLevelEmphasis
	wavelet-LLL_glrlm_ShortRunLowGrayLevelEmphasis	wavelet-HLH_firstorder_Maximum	original_firstorder_Minimum	wavelet-HLH_firstorder_Maximum	wavelet-HLH_firstorder_Maximum
	wavelet-HLL_glszm_SmallAreaEmphasis	wavelet-LHL_gldm_HighGrayLevelEmphasis	wavelet-HLH_firstorder_Maximum	wavelet-LHL_glcm_ClusterShade	squareroot_firstorder_InterquartileRange
	wavelet-LHH_firstorder_Skewness	squareroot_firstorder_RobustMeanAbsoluteDeviation	wavelet-LLL_glrlm_ShortRunLowGrayLevelEmphasis	wavelet-HHH_glrlm_LowGrayLevelRunEmphasis	squareroot_firstorder_Range
	wavelet-LHL_gldm_HighGrayLevelEmphasis	wavelet-HHH_glrlm_LowGrayLevelRunEmphasis	wavelet-HHH_glrlm_LowGrayLevelRunEmphasis	wavelet-HLL_gldm_DependenceVariance	wavelet-HLL_glszm_SmallAreaEmphasis
	squareroot_firstorder_InterquartileRange	squareroot_firstorder_InterquartileRange	wavelet-LHL_gldm_LargeDependenceLowGrayLevelEmphasis	wavelet-HLL_glcm_SumEntropy	wavelet-LLL_glrlm_ShortRunLowGrayLevelEmphasis
	wavelet-HHL_glcm_ClusterShade	wavelet-LLL_firstorder_InterquartileRange	wavelet-LHL_glszm_SmallAreaEmphasis	wavelet-HLL_glszm_SmallAreaEmphasis	wavelet-HHH_glrlm_LowGrayLevelRunEmphasis
	wavelet-LLH_glszm_SizeZoneNonUniformity	wavelet-LLL_firstorder_Range	wavelet-LHH_firstorder_Skewness	wavelet-HLL_glszm_SmallAreaHighGrayLevelEmphasis	wavelet-HHL_glszm_GrayLevelVariance
	original_ngtdm_Busyness	wavelet-HLL_gldm_DependenceVariance	wavelet-HLL_glszm_SmallAreaEmphasis		wavelet-LHH_glszm_SizeZoneNonUniformityNormalized
	original_firstorder_Minimum	wavelet-LHL_glszm_SmallAreaLowGrayLevelEmphasis	wavelet-HHH_glszm_SmallAreaLowGrayLevelEmphasis		wavelet-LHH_firstorder_Skewness

Note: glcm (gray-level cooccurrence matrix); glrlm (gray-level run-length matrix); glszm (gray-level-size zone matrix); gldm (gray-level dependence matrix); ngtdm (neighbouring gray-tone difference matrix).

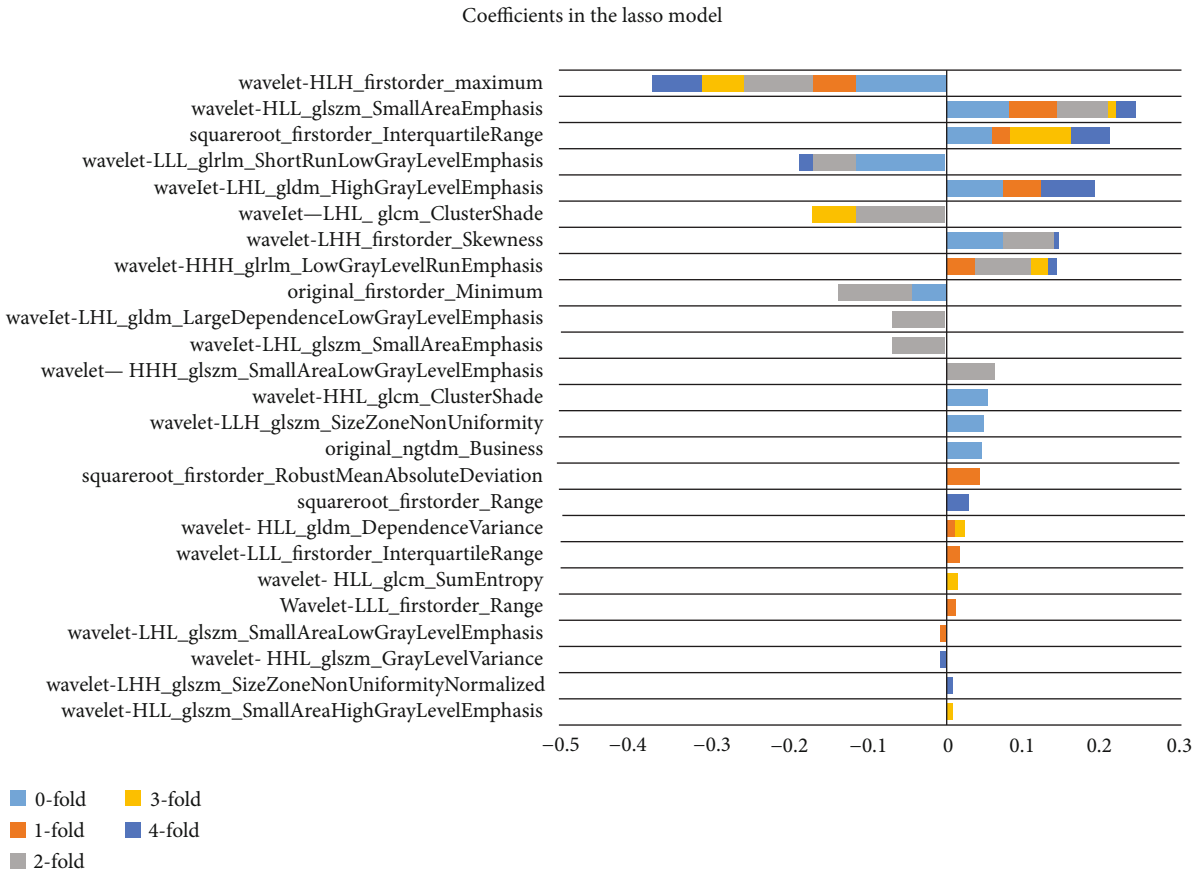


Figure 3: Cumulative graph of optimal feature coefficients of 5-fold cross-validation.

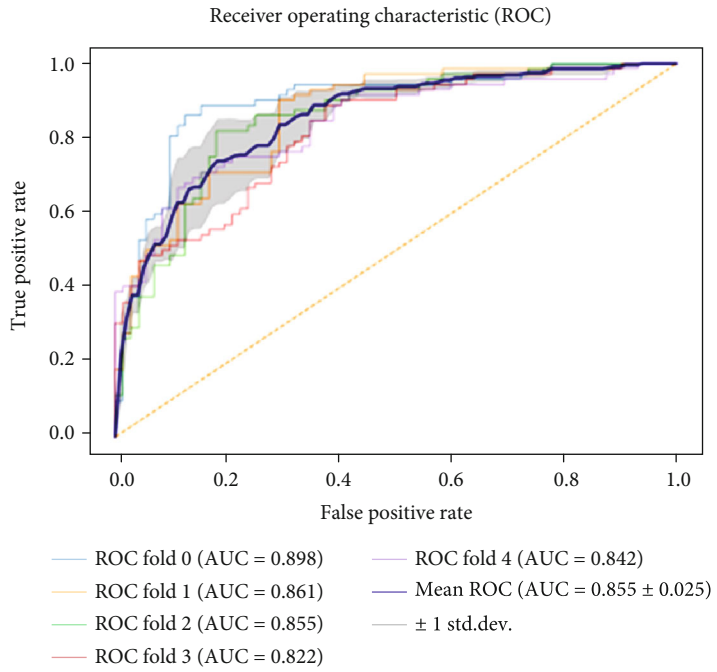
learning classifier models performed well. The average AUC of the test cohorts was higher than 0.800. The MLP had the best discrimination for HAE characteristic prediction. The mean AUC of the training cohorts was 0.925 ± 0.057, with a mean accuracy of 0.866. The mean sensitivity was 0.883, while the mean specificity was 0.889. Besides, the mean AUC of the test cohorts was 0.830 ± 0.053, with a mean accuracy of 0.817. The mean sensitivity was 0.822, and the mean specificity was 0.811.

4. Discussion

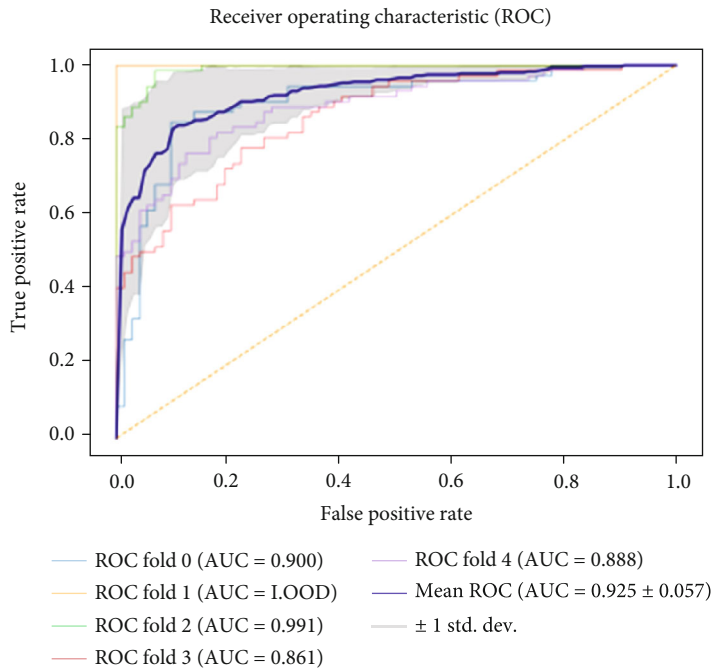
Hepatic alveolar echinococcosis (HAE) is a rare disease, often known as "worm cancer," affecting the liver [20]. Compared with cystic echinococcosis (CE), HAE is by far more severe in affected patients. Despite the marked healthcare improvements in the western agricultural and pastoral regions of China, and the rising national health examination rates, the early detection rates of HAE cases continue to rise at an alarming rate [21, 22]. In the past, radical resection surgery was the first choice for patients with HAE [23]; however, with early diagnosis of lesions and the progress of treatment, choosing the treatment with less trauma and fewer complications can better reflect the humanistic care for the patients [24–26]. Therefore, it is imperative to accurately evaluate and analyze the biological activity of the lesions for better medical care. However, conventional imaging examinations are not sufficient to accurately and quantitatively evaluate the disease.

Invasive diagnostic methods, such as biopsy examinations, only obtain a small part of the lesion tissue, which may not fully reflect all characteristics of the lesion, and therefore, offer insufficient information. To date, a robust, noninvasive, affordable, and accessible method of evaluating and monitoring HAE lesions has not been developed. The rapid development of artificial intelligence, especially radiomics, in the field of radiology, in recent years, has presented unprecedented opportunities for the assessment of HAE lesions. Radiomics, which performs the high-throughput information extraction, yields robust and valuable data more reliably than visual observation. However, the research on its potential application in the diagnosis and treatment of HAE lesions is still in its initial stages.

The main pathological manifestations of HAE lesions are liquefaction, necrosis, calcification, and solid areas. Microscopically, numerous small vesicles are observed at the edge of the solid areas of the lesions. Many inflammatory cells, eosinophils, necrotic areas, new capillaries, and other structures are evident around the vesicles. The small vesicles continue to proliferate and erode the surrounding normal liver tissue. There are granulomatous reactions around HAE lesions characterized by fibrous tissue hyperplasia;

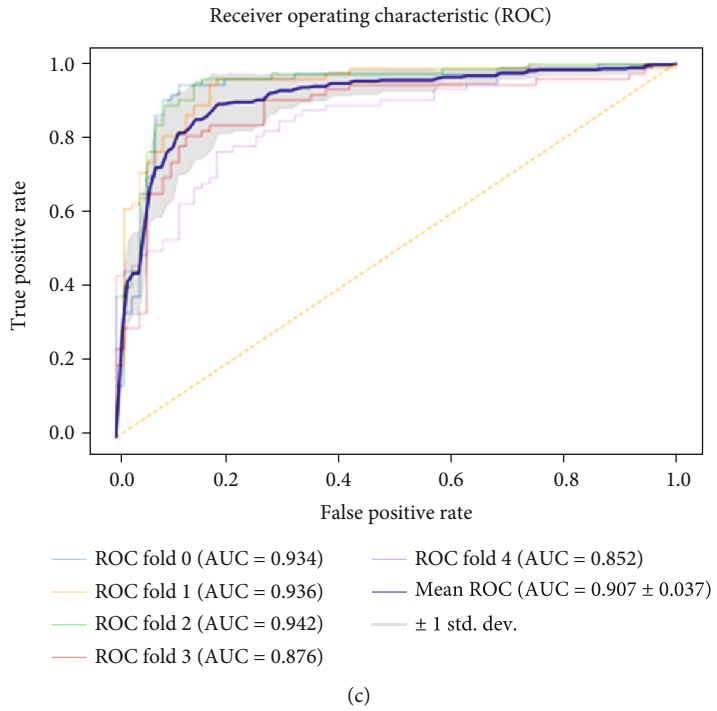


(a)



(b)

Figure 4: Continued.



(c)

Figure 4: The ROC curves of the LR, MLP, and SVM machine learning classifiers in the training cohorts: (a) LR, (b) MLP, and (c) SVM.

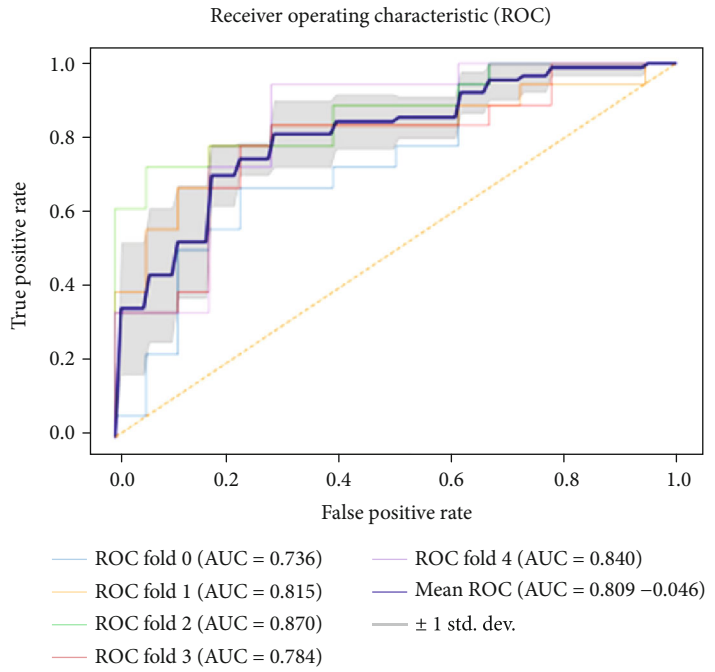
infiltration of eosinophils, lymphocytes, and foreign body giant cells; and observable formation of alveolar echinococcosis nodules. This area is regarded as the bioactive part of HAE lesions. Therefore, internal vascularization and fibrosis of HAE lesions appear alternately and constitute the basic pathological changes of alveolar echinococcosis lesions in different periods [27–29].

Compared with CT and ultrasound images, MRI images can more accurately display the active part of the lesions, which can significantly reduce errors of manual operation for early imaging procedures. Besides, PET-CT has the characteristics of both anatomical morphology and functional metabolic imaging. It detects the status of glucose metabolism in parasites and indirectly interferes with the proliferative activity of lesions. The 18F-FDG PET-CT reflects the metabolism of the lesions through the semiquantitative index (SUV) value to determine biological activity in HAE lesions. Therefore, the 18F-FDG PET-CT shows the active areas of lesions that cannot be detected by traditional imaging examinations [30–32].

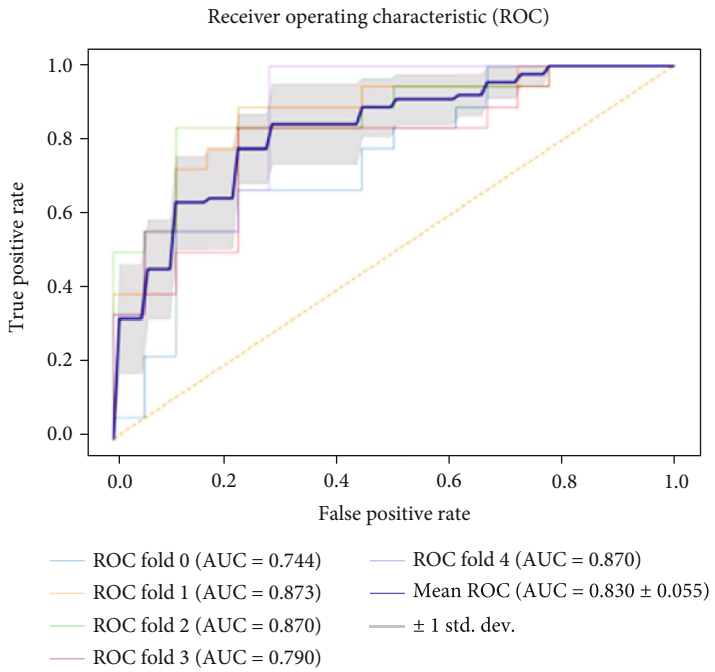
In this study, PET-CT showed that the FDG uptake pattern of HAE lesions was located in marginal areas, and most of them were semicircular and nodular, which was consistent with the distribution characteristics of small vesicles on MRI images. According to the study of Kodama et al. [33], in the early stage of HAE, a parasitic cyst manifests as a small vesicle structure. The formation of the germinal layer into a vesicle is among the two important larval development stages. Small vesicles structurally surround the granulation tissue, thus stimulating and mediating host immune responses [34]. The immune cells can absorb FDG, but the vesicles can not, thus indicating the biological activity of lesions. It could be better explained that 18F-FDG in PET-CT is mainly concentrated at the margin of the lesion rather than the small vesicles.

This study suggests that an imaging model based on the combination of radiomics features and machine learning methods might improve the accuracy of noninvasive diagnosis and serve as a valuable guide in clinical decision-making [35–37]. The construction of the HAE activity prediction model based on MRI radiomics features to evaluate the activity of HAE lesions does not use radiation, has high economic efficiency, and has high consistency with PET-CT, which would be an indispensable evaluation method for the diagnosis and treatment of HAE in the future.

In this study, conventional T2WI imaging features were extracted, and dimensionality reduction analysis was carried out by the LASSO regression algorithm to select the features that could best reflect the difference in HAE activity. The purpose of the LASSO method was to minimize the cost function and to obtain all features with nonzero coefficients, which would improve the interpretation and prediction accuracy of the model. The selected optimal feature subset contained a large number of first-order statistical features and texture features. The first-order statistical features reflect the internal voxel intensity of the lesions, and the texture features reflect the gray distribution characteristics in dimensional space, suggesting the heterogeneity of the lesions. Among them, the maximum intensity descriptor (wavelet-HLH_firstorder_Maximum, $P = 0.00167$, U test) appeared in five groups of experiments simultaneously showing the highest cumulative correlation. This indicates the heterogeneity of composition or distribution in HAE lesions by the maximum gray level intensity within the VOI and may be

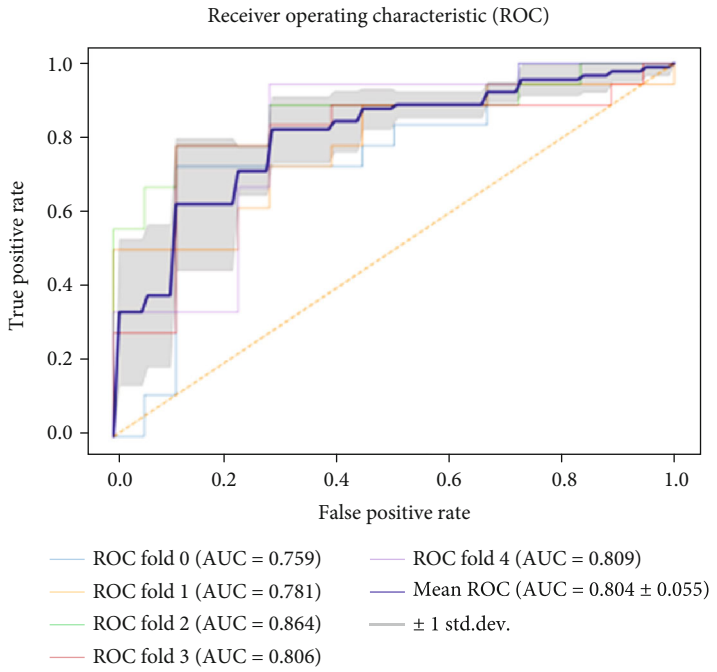


(a)



(b)

Figure 5: Continued.



(c)

Figure 5: The ROC curves of the LR, MLP, and SVM machine learning classifiers in the test cohorts: (a) LR, (b) MLP, and (c) SVM.

Table 3: Diagnostic performance of machine learning-based MRI radiomics classifiers to assess the bioactivity of HAE lesions in the training cohort.

	AUC	Accuracy	Sensitivity	Specificity	95% CI, AUC	Cutoff
LR						
0-fold	0.898	0.868	0.861	0.875	0.836-0.942	0.517
1-fold	0.861	0.806	0.903	0.708	0.793-0.913	0.436
2-fold	0.855	0.819	0.819	0.819	0.786-0.908	0.484
3-fold	0.822	0.757	0.889	0.625	0.750-0.881	0.392
4-fold	0.842	0.778	0.708	0.847	0.772-0.898	0.563
Mean	0.855 ± 0.025	0.806	0.836	0.775	0.750-0.942	
MLP						
0-fold	0.900	0.875	0.847	0.903	0.839-0.944	0.544
1-fold	1.000	1.000	1.000	1.000	0.975-1.000	0.834
2-fold	0.991	0.958	0.986	0.931	0.958-1.000	0.387
3-fold	0.861	0.778	0.778	0.778	0.793-0.913	0.527
4-fold	0.888	0.819	0.806	0.833	0.825-0.935	0.488
Mean	**0.925 ± 0.057**	**0.886**	**0.883**	**0.889**	**0.793-1.000**	
SVM						
0-fold	0.898	0.868	0.861	0.875	0.836-0.942	0.517
1-fold	0.861	0.806	0.903	0.708	0.793-0.913	0.436
2-fold	0.855	0.819	0.819	0.819	0.786-0.908	0.484
3-fold	0.822	0.757	0.889	0.625	0.750-0.881	0.392
4-fold	0.842	0.778	0.708	0.847	0.772-0.898	0.563
Mean	0.907 ± 0.037	0.806	0.836	0.775	0.750-0.942	

Table 4: Diagnostic performance of machine learning-based MRI radiomics classifiers to assesses bioactivity of HAE lesions in the test cohort.

	AUC	Accuracy	Sensitivity	Specificity	95% CI, AUC
LR	0.809 ± 0.046	0.794	0.778	**0.811**	0.565-0.959
MLP	**0.830 ± 0.053**	**0.817**	**0.822**	**0.811**	**0.571-0.960**
SVM	0.804 ± 0.035	0.794	0.778	**0.811**	0.565-0.959

used as an effective imaging biomarker to evaluate the activity of HAE lesions.

Considering that the performance of some classifiers may vary with different lesions, we employed three machine learning methods with different computing mechanisms to construct a biological activity prediction model of HAE lesions. Both LR and SVM are linear classification algorithms if the kernel function is not considered. However, SVM only considers the points near the local boundary line, while LR considers all. MLP is a generalization of a single-layer perceptron, which could solve the nonlinear problems that a single-layer perceptron could not solve [38]. In this study, the MLP training cohorts showed a promising AUC of 0.928. Generally, the three models performed well, and the average AUCs of the test cohorts were higher than 0.800. This also suggests that MRI images have higher tissue resolution and could reflect the internal heterogeneity of the lesions better.

Also, the results showed significantly improved model sensitivity and specificity after the data ratio of the active group and the inactive group was balanced by the SMOTE algorithm. The SMOTE algorithm is an enhanced sampling method. Computation for new synthetic sampling is based on Euclidian distance for variables, rather than a simple oversampling [39]. It has been shown that SMOTE is robust to the variation of unbalanced ratio with various classifiers.

Nevertheless, the present research has several limitations. First, it is a single center study with a small sample size; further expanding the sample size and carrying out a multicenter study to improve the effectiveness of the model is needed. Second, the retrospective nature of this study, the long period, the incomplete clinical data, the manual segmentation of lesions, subjectivity, the inevitable existence of selective bias, and the results of different personnel calibration may affect the establishment of the model. Third, the radiomics model of HAE lesions based on MRI features needs further discriminant analysis with intrahepatic neoplasia and tumor with poor blood supply. Finally, the diagnostic efficiency of the radiomics model of HAE activity needs to be further compared with Kodama classification and Graeter classification.

5. Conclusions

In conclusion, T2WI-based imaging features and machine learning models can evaluate the biological activity of HAE lesions, which is helpful for the selection and monitoring of clinical treatment methods.

Abbreviations

ML: Machine learning
MRI: Magnetic resonance imaging
HAE: Hepatic alveolar echinococcosis
PET-CT: Positron emission tomography computed tomography
T2WI: T2-weighted images
SMOTE: Synthetic minority oversampling technique
LASSO: Least absolute shrinkage and selection operator
LR: Logistic regression
MLP: Multilayer perceptron
SVM: Support vector machine
ROC: Receiver operating characteristic curve
AUC: Area under the curve.

Data Availability

The radiomics features data extracted from MR images are included within the supplementary information file. However, the image datasets in the current study are not publicly available due to patient privacy protection, but are available from the corresponding author on reasonable request.

Ethical Approval

This retrospective study was approved by the Medical Ethics Review Committee of the First Affiliated Hospital of Xinjiang Medical University, and exempted from informed consent.

Disclosure

The funding body contributed to the design of the study and analysis of data.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

BR and JW contributed equally to this study. BR and WL conceived of the present idea and wrote the manuscript. JW, ZM, TZ, and AA did some work in performing the experiments. YX analyzed the results. All authors read and approved the final manuscript for publication.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (81974263).

References

[1] D. P. McManus, W. Zhang, J. Li, and P. B. Bartley, "Echinococcosis," *The Lancet*, vol. 362, no. 9392, pp. 1295–1304, 2003.

[2] M. Bulakçı, M. G. Kartal, S. Yılmaz et al., "Multimodality imaging in diagnosis and management of alveolar echinococcosis: an update," *Diagnostic and interventional radiology*, vol. 22, no. 3, pp. 247–256, 2016.

[3] E. Brunetti, P. Kern, and D. A. Vuitton, "Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans," *Acta Tropica*, vol. 114, no. 1, pp. 1–16, 2010.

[4] A. Hemphill, B. Stadelmann, R. Rufener et al., "Treatment of echinococcosis: albendazole and mebendazole—what else?," *Parasite*, vol. 21, p. 70, 2014.

[5] N. M. Bakhtiar, A. Akbarzadeh, A. Casulli et al., "Therapeutic efficacy of nanocompounds in the treatment of cystic and alveolar echinococcoses: challenges and future prospects," *Parasitology Research*, vol. 118, no. 9, pp. 2455–2466, 2019.

[6] W. Liu, É. Delabrousse, O. Blagosklonov et al., "Innovation in hepatic alveolar echinococcosis imaging: best use of old tools, and necessary evaluation of new ones," *Parasite*, vol. 21, p. 74, 2014.

[7] J. Wang, Y. Xing, B. Ren, W. D. Xie, H. Wen, and W. Y. Liu, "Alveolar echinococcosis: correlation of imaging type with PNM stage and diameter of lesions," *Chinese Medical Journal*, vol. 124, no. 18, pp. 2824–2828, 2011.

[8] Y. Jiang, J. Li, J. Wang et al., "Assessment of vascularity in hepatic alveolar echinococcosis: comparison of quantified dual-energy CT with histopathologic parameters," *PLoS One*, vol. 11, no. 2, article e0149440, 2016.

[9] Y. Abudureheman, J. Wang, and W. Liu, "Comparison of intravoxel incoherent motion diffusion-weighted magnetic resonance (MR) imaging to T1 mapping in characterization of hepatic alveolar echinococcosis," *Medical science monitor: international medical journal of experimental and clinical research*, vol. 23, pp. 6019–6025, 2017.

[10] C. Caoduro, C. Porot, D. A. Vuitton et al., "The role of delayed 18F-FDG PET imaging in the follow-up of patients with alveolar echinococcosis," *Journal of nuclear medicine: official publication, Society of Nuclear Medicine*, vol. 54, no. 3, pp. 358–363, 2013.

[11] R. T. H. Leijenaar, G. Nalbantov, S. Carvalho et al., "The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis," *Scientific Reports*, vol. 5, no. 1, 2015.

[12] P. Lambin, E. Rios-Velazquez, R. Leijenaar et al., "Radiomics: extracting more information from medical images using advanced feature analysis," *European journal of cancer*, vol. 48, no. 4, pp. 441–446, 2012.

[13] M. Liu, X. Ma, F. Shen, Y. Xia, Y. Jia, and J. Lu, "MRI-based radiomics nomogram to predict synchronous liver metastasis in primary rectal cancer patients," *Cancer Medicine*, vol. 9, no. 14, pp. 5155–5163, 2020.

[14] E. K. Oikonomou, M. Siddique, and C. Antoniades, "Artificial intelligence in medical imaging: a radiomic guide to precision phenotyping of cardiovascular disease," *Cardiovascular Research*, vol. 116, no. 13, pp. 2040–2054, 2020.

[15] H. Sui, L. Liu, X. Li, P. Zuo, J. Cui, and Z. Mo, "CT-based radiomics features analysis for predicting the risk of anterior mediastinal lesions," *Journal of Thoracic Disease*, vol. 11, no. 5, pp. 1809–1818, 2019.

[16] P. Nie, G. Yang, Z. Wang et al., "A CT-based radiomics nomogram for differentiation of renal angiomyolipoma without visible fat from homogeneous clear cell renal cell carcinoma," *European Radiology*, vol. 30, no. 2, pp. 1274–1284, 2020.

[17] Y. Zhang, Y. Zhu, X. Shi et al., "Soft tissue sarcomas: preoperative predictive histopathological grading based on radiomics of MRI," *Academic Radiology*, vol. 26, no. 9, pp. 1262–1268, 2019.

[18] C. Liu, J. Ding, K. Spuhler et al., "Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI," *Journal of magnetic resonance imaging: JMRI*, vol. 49, no. 1, pp. 131–140, 2019.

[19] A. Zwanenburg, S. Leger, C. Vallières, and S. Löck, "Image biomarker standardisation initiative-feature definitions," 2016, https://arxiv.org/abs/1612.07003.

[20] H. Wen, L. Vuitton, T. Tuxun et al., "Echinococcosis: advances in the 21st century," *Clinical Microbiology Reviews*, vol. 32, no. 2, 2019.

[21] X. Feng, X. Qi, L. Yang et al., "Human cystic and alveolar echinococcosis in the Tibet Autonomous Region (TAR), China," *Journal of Helminthology*, vol. 89, no. 6, pp. 671–679, 2015.

[22] X. Wang, G. Dai, M. Li, W. Jia, Z. Guo, and J. Lu, "Prevalence of human alveolar echinococcosis in China: a systematic review and meta-analysis," *BMC Public Health*, vol. 20, no. 1, p. 1105, 2020.

[23] L. A. Salm, A. Lachenmayer, S. F. Perrodin, D. Candinas, and G. Beldi, "Surgical treatment strategies for hepatic alveolar echinococcosis," *Food and waterborne parasitology*, vol. 15, 2019.

[24] J. Schmidberger, J. Steinbach, P. Schlingeloff, W. Kratzer, and B. Grüner, "Surgery versus conservative drug therapy in alveolar echinococcosis patients in Germany - A health-related quality of life comparison," *Food and waterborne parasitology*, vol. 16, 2019.

[25] J. Fabbri, P. E. Pensel, C. M. Albani et al., "Albendazole solid dispersions against alveolar echinococcosis: a pharmacotechnical strategy to improve the efficacy of the drug," *Parasitology*, vol. 147, no. 9, pp. 1026–1031, 2020.

[26] G. Beldi, D. Vuitton, A. Lachenmayer et al., "Is ex vivo liver resection and autotransplantation a valid alternative treatment for end-stage hepatic alveolar echinococcosis in Europe?," *Journal of Hepatology*, vol. 70, no. 5, pp. 1030-1031, 2019.

[27] M. Reinehr, C. Micheloud, F. Grimm et al., "Pathology of echinococcosis," *The American Journal of Surgical Pathology*, vol. 44, no. 1, pp. 43–54, 2020.

[28] J. B. Taxy, W. E. Gibson, and M. W. Kaufman, "Echinococcosis: unexpected occurrence and the diagnostic contribution of routine histopathology," *The American Journal of Surgical Pathology*, vol. 41, no. 1, pp. 94–100, 2017.

[29] T. Graeter, W. Kratzer, S. Oeztuerk et al., "Proposal of a computed tomography classification for hepatic alveolar echinococcosis," *World Journal of Gastroenterology*, vol. 22, no. 13, pp. 3621–3631, 2016.

[30] Y. Qin, X. Li, Q. Zhang et al., "Analysis of the clinical value of 18F-FDG PET/CT in hepatic alveolar echinococcosis before and after autologous liver transplantation," *Experimental and Therapeutic Medicine*, vol. 11, no. 1, pp. 43–48, 2016.

[31] T. Graeter, N. Eberhardt, R. Shi et al., "Hepatic alveolar echinococcosis: correlation between computed tomography morphology and inflammatory activity in positron emission tomography," *Scientific Reports*, vol. 10, no. 1, p. 11808, 2020.

[32] F. Lötsch, F. Waneck, M. Groger et al., "FDG-PET/MRI imaging for the management of alveolar echinococcosis: initial clinical experience at a reference centre in Austria," *Tropical Medicine & International Health*, vol. 24, no. 6, pp. 663–670, 2019.

[33] Y. Kodama, N. Fujita, T. Shimizu et al., "Alveolar echinococcosis: MR findings in the liver," *Radiology*, vol. 228, no. 1, pp. 172–177, 2003.

[34] J. Zheng, J. Wang, J. Zhao, and X. Meng, "Diffusion-weighted MRI for the initial viability evaluation of parasites in hepatic alveolar echinococcosis: comparison with positron emission tomography," *Korean Journal of Radiology*, vol. 19, no. 1, pp. 40–46, 2018.

[35] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, no. 1, 2014.

[36] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[37] P. O. Alderson and R. M. Summers, "The evolving status of radiomics," *Journal of the National Cancer Institute*, vol. 112, no. 9, pp. 869-870, 2020.

[38] Z. X. Guo, W. K. Wong, and M. Li, "Sparsely connected neural network-based time series forecasting," *Information Sciences*, vol. 193, pp. 54–71, 2012.

[39] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

Hindawi
BioMed Research International
Volume 2021, Article ID 5562801, 19 pages
https://doi.org/10.1155/2021/5562801

Research Article

A Segmentation of Melanocytic Skin Lesions in Dermoscopic and Standard Images Using a Hybrid Two-Stage Approach

Yoo Na Hwang,[1] **Min Ji Seo**,[1] **and Sung Min Kim**[1,2]

[1]*Department of Medical Biotechnology, Dongguk University-Bio Medi Campus, 10326, Republic of Korea*
[2]*Department of Medical Device Industry, Dongguk University, 04620, Republic of Korea*

Correspondence should be addressed to Sung Min Kim; smkim@dongguk.edu

Received 12 January 2021; Revised 17 March 2021; Accepted 26 March 2021; Published 7 April 2021

Academic Editor: Lin Gu

The segmentation of a skin lesion is regarded as very challenging because of the low contrast between the lesion and the surrounding skin, the existence of various artifacts, and different imaging acquisition conditions. The purpose of this study is to segment melanocytic skin lesions in dermoscopic and standard images by using a hybrid model combining a new hierarchical K-means and level set approach, called HK-LS. Although the level set method is usually sensitive to initial estimation, it is widely used in biomedical image segmentation because it can segment more complex images and does not require a large number of manually labelled images. The preprocessing step is used for the proposed model to be less sensitive to intensity inhomogeneity. The proposed method was evaluated on medical skin images from two publicly available datasets including the PH^2 database and the Dermofit database. All skin lesions were segmented with high accuracies (>94%) and Dice coefficients (>0.91) of the ground truth on two databases. The quantitative experimental results reveal that the proposed method yielded significantly better results compared to other traditional level set models and has a certain advantage over the segmentation results of U-net in standard images. The proposed method had high clinical applicability for the segmentation of melanocytic skin lesions in dermoscopic and standard images.

1. Introduction

Melanoma is a dangerous skin cancer that mostly appears in pigmented cells (melanocytes) in the skin. It is a major cause of death associated with skin cancer [1]. Early diagnosis of melanoma is essential because early-stage detection and proper treatment increase the survival rate [2, 3]. Melanoma is mostly detected by expert dermatologists through visual inspection using the naked eye alone with a diagnostic accuracy of about 60% [4, 5].

Clinical images are normally obtained using digital cameras. However, the imaging conditions are frequently inconsistent because images are acquired from different distances or under variable illumination conditions. These may lead to problems when the size of the lesion is too small. Dermoscopy, a technique whereby a hand-held device is used to detect a mole and inspect the underlying skin, is better than unaided visual inspection and increases the sensitivity of detection by 10-30% [6]. Nevertheless, the within- and between-observer concordance is very low, even for expert clinicians [7]. An additional problem is related to the presence of intrinsic noise and artifacts, such as hair, blood vessels, air bubbles, and frames; variegated colors inside the lesion; and the lack of distinct boundaries to the surrounding skin [8]. These make it difficult to distinguish the skin lesion [9]. Thus, a growing interest has developed in the computational analysis of skin lesion images to assist clinicians in distinguishing early melanoma from benign lesions [10].

The first step in the computerized analysis of skin lesion images is the segmentation of the lesion. The segmentation of skin lesions from the surrounding skin is essential to provide important information for an accurate analysis of skin lesions and to extract important clinical features such as atypical pigment networks, blue-white areas, and globules

[11, 12]. Moreover, this step is the key process by which lesion diameters are quantified and the extent of border irregularities are evaluated. Effective methods have been proposed to improve the segmentation accuracy.

Active contour-based medical image segmentation, such as a level set, is a well-established approach [13]. It was first introduced by Osher and Sethian. Level set evolution, which is established on partial differential equations and dynamic implicit interfaces, has been widely used in the field of medical image segmentation. Silveira and Marquez [13], Nourmohamadi and Pourghassem [14], and Li et al. [15] used the level set method with clustering-based initial estimation models, such as the Otsu thresholding, weighting combination of fuzzy C-mean and K-means, and spatial fuzzy clustering. The level set method is an efficient way to identify low contrast boundaries [16]. Schmid [17] presented a color clustering-based technique with a modified version of fuzzy C-means clustering. Donadey et al. [18] also detected a border by using the intensity component of hue-saturation-intensity (HSI) space. However, traditional models such as the region-based active contour model often failed when applied to images containing inhomogeneities. These are very sensitive to parameter tuning [16]. Recently, machine learning algorithms, including deep learning architectures, such as Residual net [1] or U-net [9], have emerged as reliable segmentation methods for skin lesion images. However, these algorithms can deal with inhomogeneities but require postprocessing and a large training set [16]. Some cases still show a low performance of skin lesion segmentation due to very low contrast and hair artifacts in skin lesion images [8]. These make it hard to train effectively deep networks with a large number of parameters [1].

To tackle the abovementioned problems, a hybrid model which integrates unsupervised learning with a region-based active contour model is proposed in this study. The proposed method combined the hierarchical K-means clustering and level set methods. This model thus can be less sensitive to parameter controlling of the level set model and to intensity inhomogeneity. The rest of this study was organized as follows. Section 2 introduces the overall processes used in the segmentation: (a) preprocessing, (b) segmentation, and (c) performance evaluation. Sections 3 and 4 provide the experimental results and discussions, respectively. Finally, Section 5 concluded the paper and identified future directions.

2. Materials and Methods

To segment a melanocytic skin lesion accurately, the proposed method was implemented through four steps: image acquisition, preprocessing, a two-stage segmentation model, and postprocessing. The statistical significance of the suggested method was evaluated by the Jaccard index, the Dice coefficient, sensitivity, and other measures. Figure 1 shows an overall flowchart of the suggested approach for the segmentation of each skin lesion. The detailed procedures are described below.

2.1. Image Acquisition. This study used dermoscopic and standard images from the following two dermatology atlases:

(1) The PH^2 data [19] is a dataset that includes 200 dermoscopic images, including 40 malignant melanomas and 160 melanocytic nevus (80 common nevi and 80 atypical nevi) at 768×560 resolution, collected by a group of researchers from the Technical Universities of Porto and Lisbon in the Dermatology Service of Pedro Hispano Hospital. Each image has 8-bit red, green, and blue (RGB) channels.

(2) The Edinburgh Dermofit Image Library [20] is a dataset that includes high-quality skin lesion images (1,300 biopsy-proven cancers and moles) collected across 10 different classes, including 331 melanocytic nevus images and 76 malignant melanoma images. The images are snapshots of the skin lesions surrounded by normal skin captured using a Canon EOS 350D SLR camera with a pixel resolution of about 0.03 mm.

Figure 2 shows the sample images with different artifacts and aberrations. The skin images obtained from these atlases were annotated by expert dermatology resource providers. All images were allocated to diagnosis labels and binary segmentation masks that denote the lesion area. In the binary segmentation mask, the pixels outside the lesions were assigned pixel intensity values of 0 and pixels inside the lesion were assigned pixel intensity values of 255. 116 images of malignant melanoma and 491 images of melanocytic nevus were acquired from two different atlases (Table 1).

2.2. Preprocessing. Dermoscopic and standard images usually contain artifacts such as illumination variations, dermoscopic gel, air bubbles, and outlines (hair, skin lines, vignetting around the lesion, ruler markers, and blood vessels). These artifacts can attenuate the accuracy of border detection and increase computational time. As a result, there is a need for robust methods to attenuate artifacts. To do this, the first step of this study is to create an image that converts the image into a different color space and removes artifacts including hair, vignetting around the lesion, and ruler markings as shown in Figure 3.

All skin images are RGB-colored images, which are the combination of gray values from the individual R, G, and B channels [21]. This color space is not as sensitive as human vision. The segmentation of skin lesions on RGB-colored images is difficult because of the influence of the pixel intensity [10]. Specifically, a skin lesion is likely to show different visual colors due to various conditions, such as illumination variations and low contrast between the skin lesions and a surrounding skin region. The RGB-colored images were converted to International Commission on Illumination (CIE) $L * a * b$ color space to clearly detect the color differences between the skin lesion and the background skin. In the CIE $L * a * b$ color space, L indicates the luminance (lightness) and a and b are chromaticity coordinates. The a axis represents a complementary color of the green-red component, and the b axis represents a complementary color of the blue-yellow component [22]. After color space transforming, only both of the two channels (a and b) were

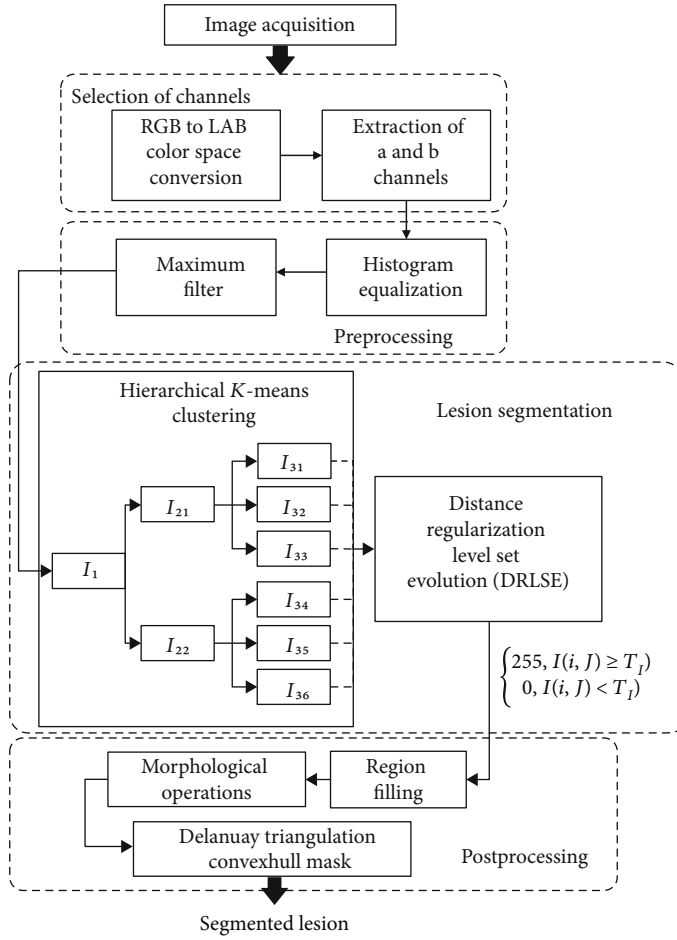


Figure 1: Overall flowchart of the proposed scheme for the segmentation of each skin lesion image in dermoscopic and standard images.

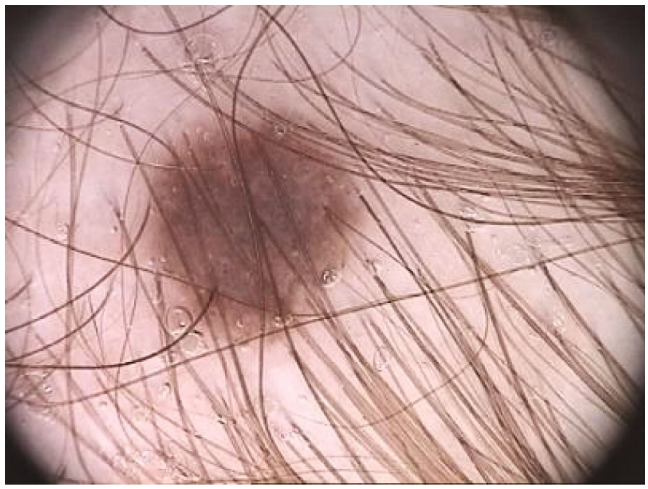
extracted and the lightness channel was excluded. The histogram equalization was applied to only two channels. Finally, we created a new 3-channel fusion image that reduces the illumination variations and skin color difference.

After the first step, maximum filters with a 5×5 kernel were also applied before the border detection to remove noise, such as hair and air bubbles. Vignetting around the image was removed by extracting the largest blobs in the binary image.

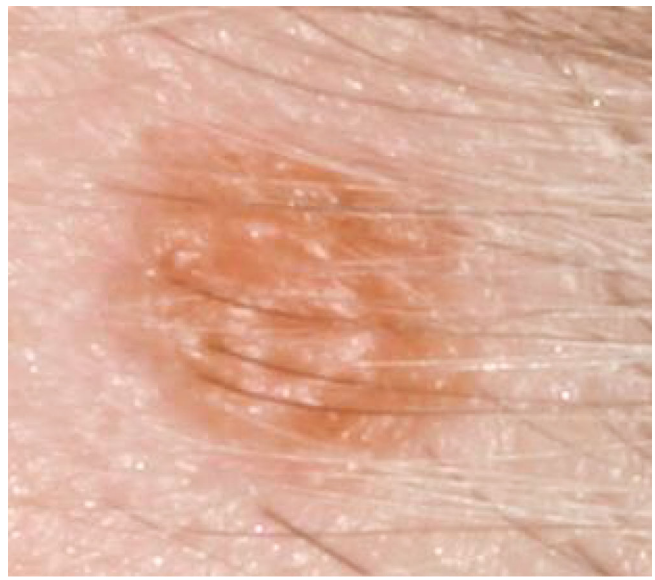
2.3. A Hybrid Two-Stage Segmentation Model. After preprocessing, a hybrid two-stage model was constructed for the segmentation of a melanocytic skin lesion. To obtain an initial contour mask of a melanocytic skin lesion area, the hybrid HK clustering was implemented first. Secondly, the Distance Regularized Level Set Evolution (DRLSE) was used to segment the fine border of the lesion. The detailed lesion segmentation step is described below.

2.3.1. Hybrid Hierarchical K-Means Clustering (HK Clustering). The basic concept of HK is to recursively split the dataset into a tree of clusters with predefined branches at each node. There are two approaches to hierarchical clustering. One is the top-down technique, and the other one is the bottom-up technique [23–25]. The top-down is more efficient than bottom-up because of the fast task and greedy attributes, meaning that it cannot cross the boundaries imposed by the top level [26, 27]. In other words, nearby points may end up in different clusters. The proposed method was a modified version of the top-down approach by Chen et al. [24]. At first, the data starts as one combined cluster. Next, the cluster splits into distinct parts of K_1 according to some degree of similarity (level 1). Finally, the clusters separate into distinct parts of K_2 again and again until the clusters only contain some small fixed number of points (level 2). Figure 4 shows a visualization of the hybrid HK clustering used in this study. K_n represents the number of clusters at the hierarchical level of n. The optimal number of clusters were set to K_1 of 2 at level 1 and K_2 of 3 at level 2 as shown in Figures 4(b) and 4(c). The number of iterations for each level of K-means was set to 20. The squared Euclidean distance measure was adopted for a similarity function.

2.3.2. A Fine Border Segmentation Based on DRLSE Model. To segment the fine border of the melanocytic skin lesion, the DRLSE, which is one of the level set evolution approaches, was employed. The traditional level set methods consider the front as the zero-level set of an embedded function on a track moving front, called the level set function (LSF) [28–31]. The objects were detected in a given image

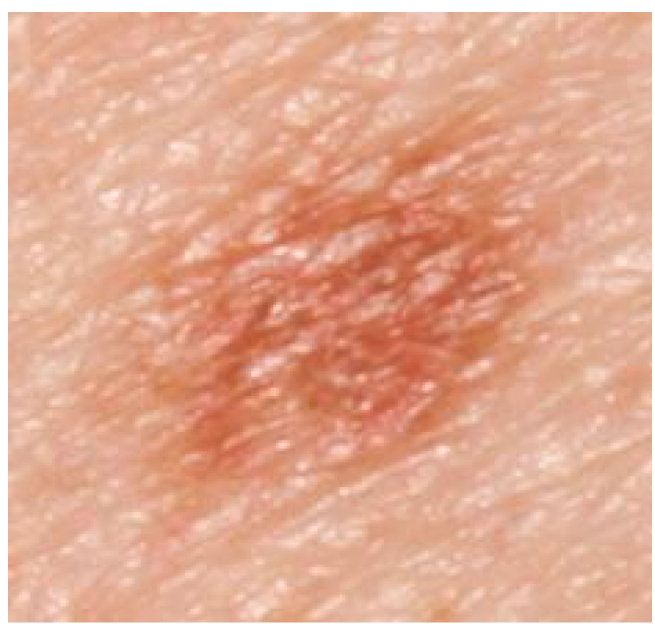


(a)



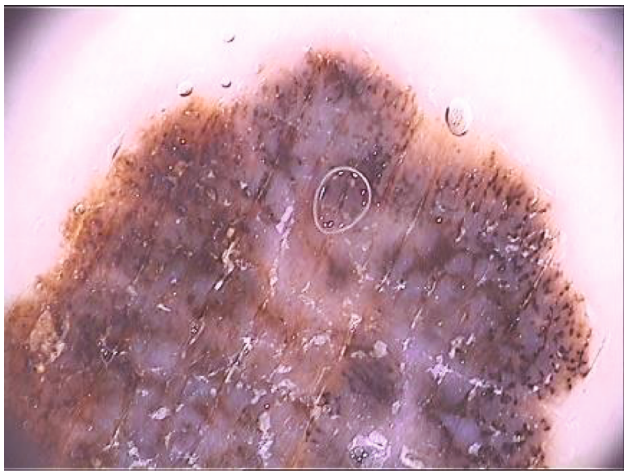
(b)

(c)

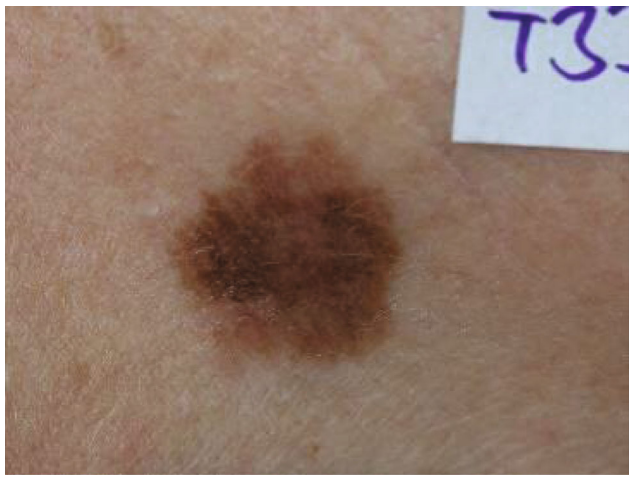


(d)

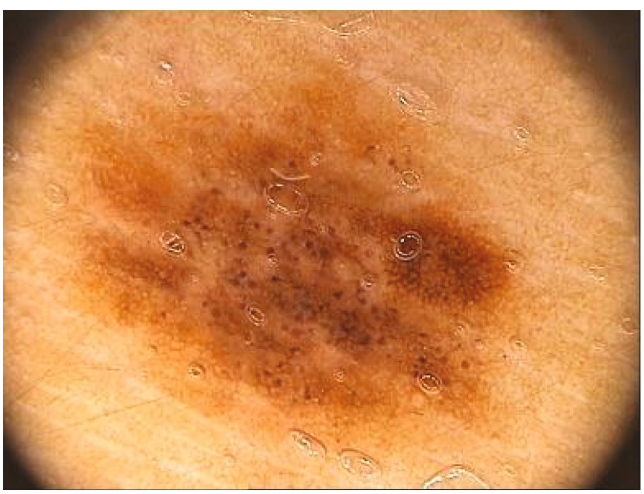
Figure 2: Continued.



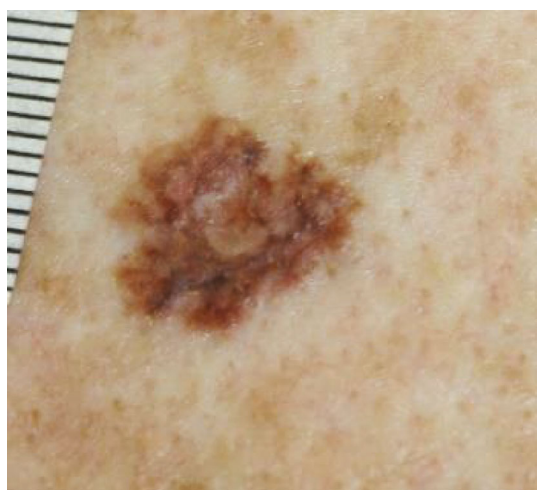
(e)



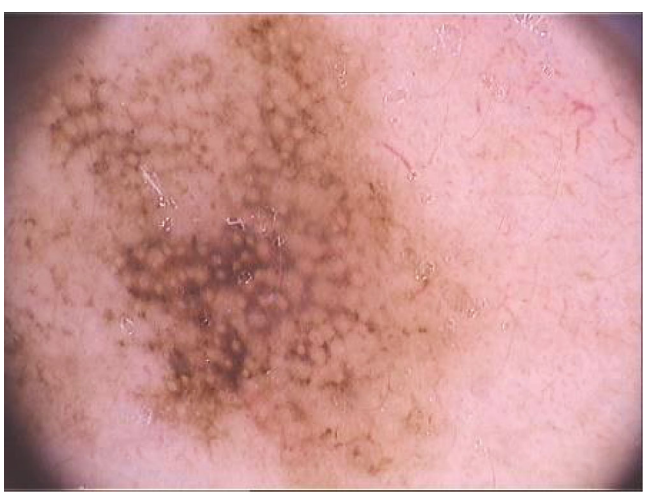
(f)



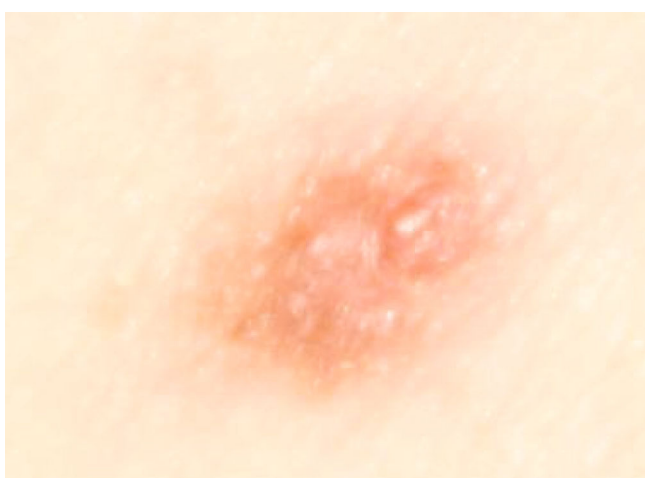
(g)



(h)



(i) The samples of dermoscopic images (PH^2) are shown in (a), (c), (e), (g), and (i)



(j) The samples of standard images (Dermofit) are shown in (b), (d), (f), (h), and (j)

Figure 2: Illustrative examples of dermoscopic images (a, c, e, g, and i) and standard images (b, d, f, h, and j).

Table 1: Dataset statistics.

Atlas (the number of images)	Skin lesion	The number of images
PH^2 data (200)	Malignant melanoma	40
	Nevus (common, melanocytic)	160
Dermofit (407)	Malignant melanoma	76
	Melanocytic nevus	331
Total (607)	Malignant melanoma	116
	Melanocytic nevus	491

by curve evolution [32]. To stop the curve evolution, the traditional level set method is influenced by the gradient of the given image by changing the LSF value. However, the LSF typically develops irregularities during its evolution in conventional level set formulations, which make an impact on numerical errors and eventually destroy the stability of the evolution [33]. Thus, to eliminate the need for reinitialization and avoid numerical errors, the DRLSE was employed to segment the fine border of the melanocytic skin lesions.

Each border of a skin lesion image can be regarded as the zero-level set of an LSF. Although the final segment result of the level set method is the zero-level set of the LSF, it is essential to maintain the LSF in a balanced state. This requirement can be satisfied by using signed distance functions with the unique property of $|\nabla\phi| = 1$, which is referred to as the signed distance property.

Given the LSF $\phi : \Omega \to \mathscr{R}$ in a rectangular domain, the energy function $E(\phi)$ is defined by

$$E(f) = \mu R_p(f) + E_{\text{ext}}(f), \tag{1}$$

where ϕ is the level set function, and $R_p(\phi)$ and $E_{\text{ext}}(\phi)$ indicate the level set regularization term and external energy function, respectively. $\mu > 0$ is a constant, and the level set regularization term $R_p(\phi)$ can be defined by

$$R_p(\phi) \triangleq \int_\Omega p(|\nabla\phi|)dx, \tag{2}$$

where p indicates the potential function ($p : [0,\infty] \to \mathscr{R}$). The energy $E_{\text{ext}}(\phi)$ is designed to achieve a minimum value when the zero-level set of the skin lesion is located at the desired position. Moreover, the edge indicator function g is stated by

$$g = \frac{1}{1 + |\nabla(G_\sigma * I)|^2}, \tag{3}$$

where I is the image $I(x, y)$ with a smoothing Gaussian kernel G_σ, and σ is the standard deviation. The edge indication function stops the level set evolution when the zero-level set of the skin lesion approaches the optimal position. The energy functional $E(\phi)$ is determined by

$$E(f) = \mu R_p(f) + \lambda A_g(f) + \alpha B_g(f), \tag{4}$$

where $\lambda > 0$ and α represent the coefficients of the energy functions $A_g(\phi)$ and $B_g(\phi)$, which can be written as follows:

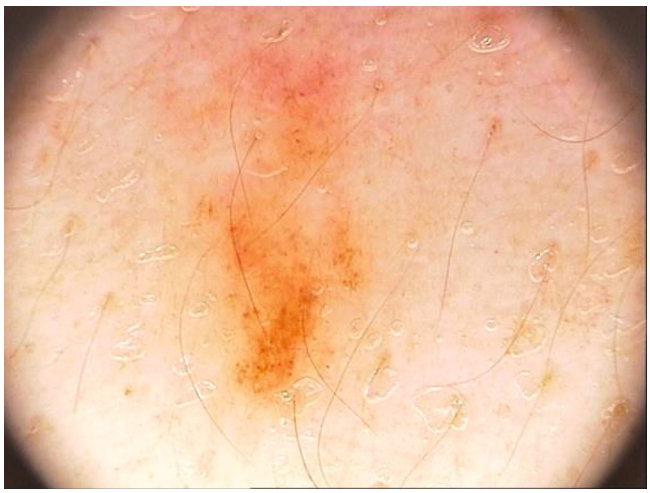
$$A_g(\phi) \triangleq \int_\Omega g\delta(\phi)|\nabla\phi|dx, \tag{5}$$

$$B_g(\phi) \triangleq \int_\Omega gH(-\phi)dx, \tag{6}$$

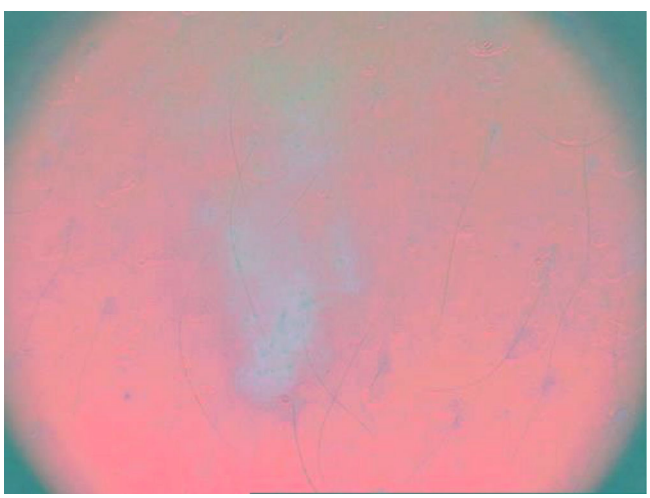
where δ and H represent the Dirac delta function and the Heaviside function, respectively. Since a signed distance function is used as the initial level set function (ϕ_0) in the standard level set and initialization should be done periodically to retain a stable evolution of zero level set function, the computational cost of these methods is high [34]. The level set evolution is derived as the gradient flow that minimizes an energy functional with a distance regularization term and an external energy that drives the motion of the zero-level set toward the desired location. The distance regularization term is defined by a potential function which includes a unique forward-and-backward (FAB) diffusion effect [33]. For instance, when the initial borders were located outside of the desired borders, alfa was set to a positive value to force the zero-level set to shrink toward the region of interest. In contrast, alfa was assigned a negative value to expand the borders when the initial borders were located on the inside. The detailed equation has been described previously [33].

The DRLSE parameters were set as follows: a constant controlling the gradient strength of the initial LSF(c_0) of 3, a coefficient of the weighted length term (λ) of 5, a width of the Dirac delta function (δ) of 1.5, a coefficient of the distance regularization term (μ) of 0.02, a time-step of 8, and a standard deviation of the Gaussian kernel (σ) of 1.5. The initial LSF (R_0) of this study was automatically detected by using the results of the HK clustering as shown in Figure 5. A set of if-then rules were applied to optimize the parameters at different conditions of images. An α, the coefficient of the weighted area term, was set to 3 or 5 regarding the size of the initial LSF. Double-well potential was used for a distance regularization term, and the iteration numbers were set to 600 and 1000 for the images of malignant and melanocytic nevi, respectively. A binary image was obtained with a threshold of 80. The area inside the fine border was filled in during the postprocessing step. A morphological erosion of the mask, using a square with a width of 5 pixels, and Delaunay triangulation were also carried out in the postprocessing step. Examples of the border segmentation results for the dermoscopic image (PH^2 dataset) and standard image (Dermofit dataset) are presented in Figure 6.

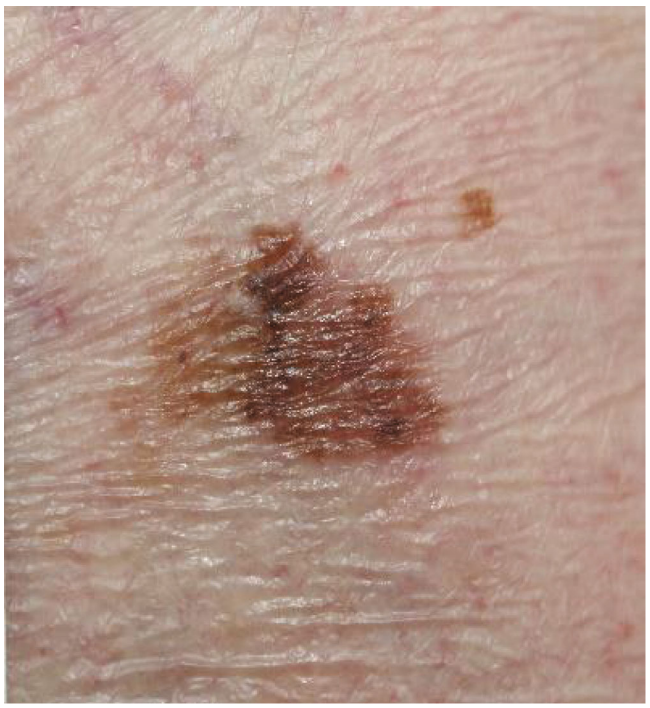
2.4. Performance Evaluation. The output of the proposed method was binarized with a lesion mask. The performance of the proposed method was evaluated on two different datasets of melanocytic skin lesion images from the PH^2 database [19] and the Dermofit database [20], which are publicly available on the ground truth data. To evaluate the proposed method, the well-known segmentation measures were calculated, including accuracy, specificity, sensitivity, Jaccard



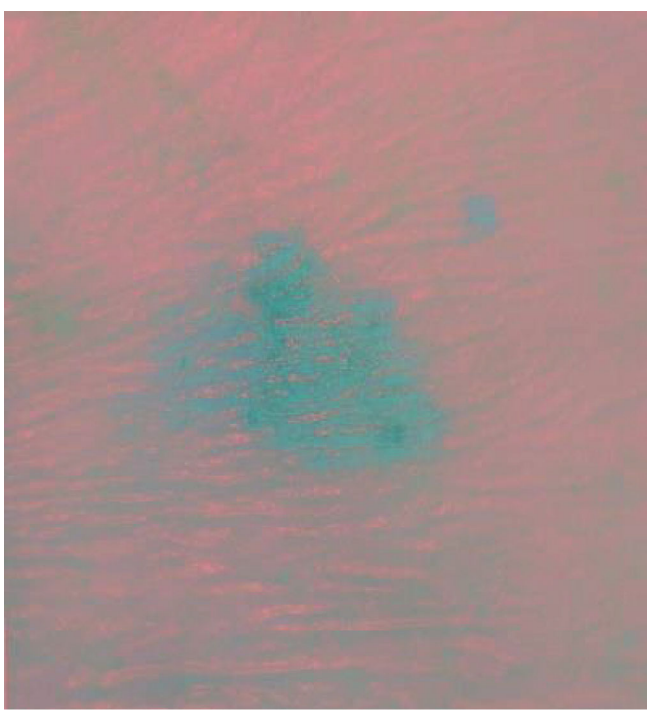
(a)



(b)



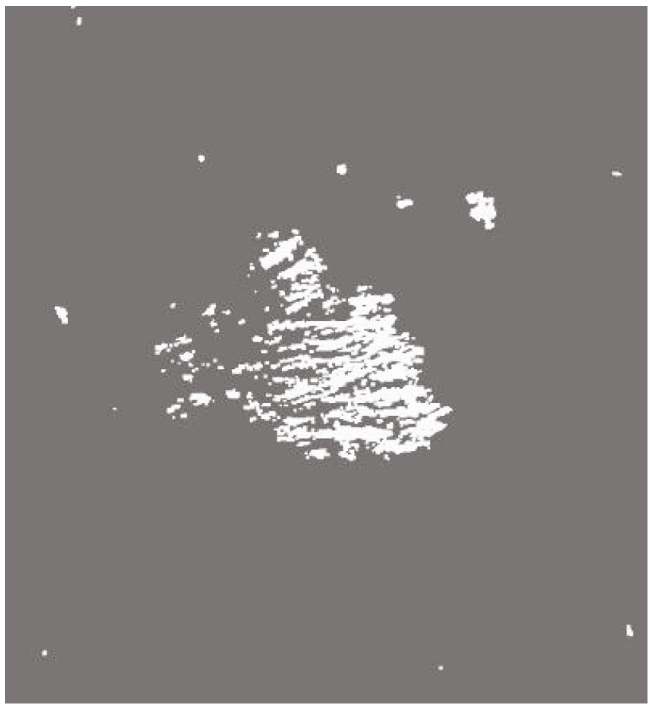
(f)



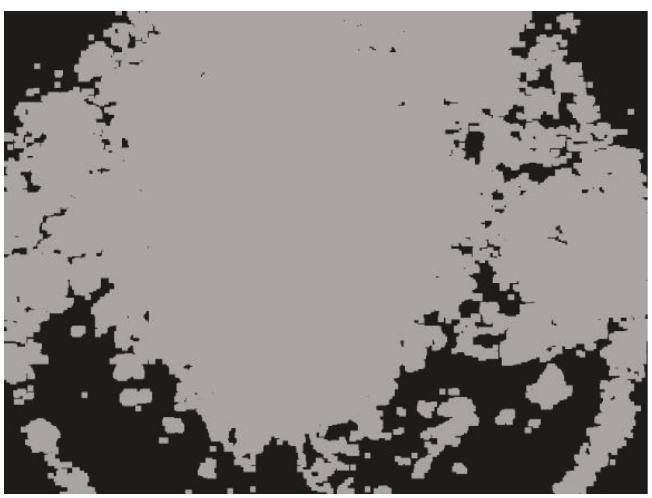
(g)

Figure 3: Continued.

(c)



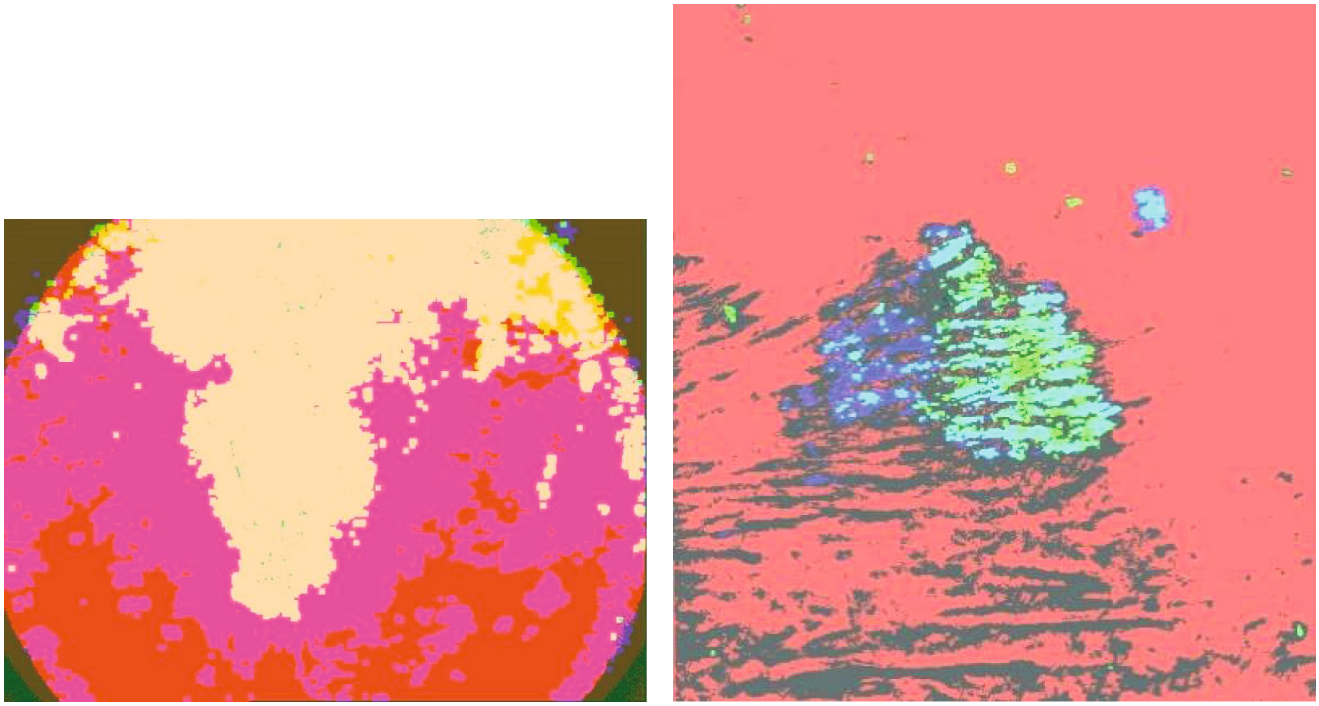
(h)



(d)

(i)

Figure 3: Continued.



(e) (j)

Figure 3: Example results of preprocessing on PH^2 (a, c, e, g, and i) and Dermofit (b, d, f, h, and j). The first row (a, b) contains original images. The second row (c, d) contains the images converted from RGB to CIE LAB color space. The 3rd row (e, f) contains histogram equalization images of channel *a* and (g, h) channel *b*. The 4th row (i, j) contains the final fusion images.

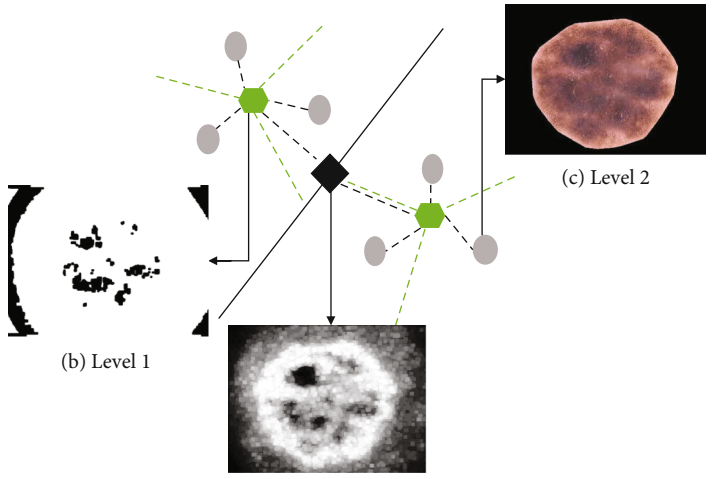


Figure 4: Visualization of the hybrid hierarchical K-means (HK) clustering method with $K_1 = 2$ and $K_2 = 3$ at level 1 and level 2, respectively. (a) Input image which was obtained after the preprocessing step. (b) Initial contour mask with $K_1 = 2$ at level 1. (c) Final initial contour mask with $K_2 = 3$ at level 2. K is the number of clusters at each hierarchical level.

index (JI), Dice coefficient (DC), F-measure, and Hausdorff distance (HD). Specifically, these measures were calculated from the following four error factors: true positive (TP), true negative (TN), false positive (FP), and false negative (FN)

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \quad (7)$$

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

$$\text{Specificity} = \frac{(\text{TN})}{(\text{TN} + \text{FP})}, \quad (9)$$

where TP represents the pixel numbers of a skin lesion correctly segmented as a skin lesion, TN represents the

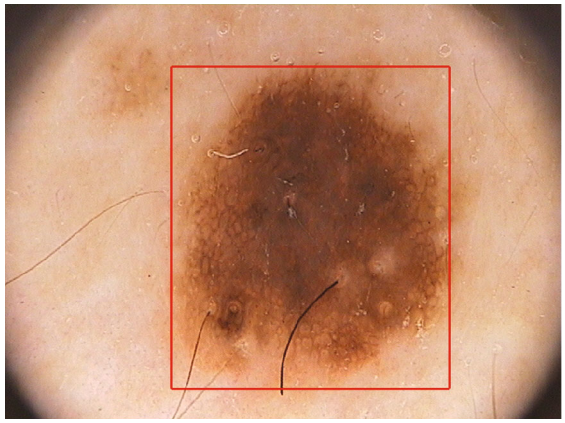


Figure 5: One example of the proposed method for the detection of a melanocytic skin lesion. The red rectangle indicates the boundary selection scheme, which includes the outline of a skin lesion.

pixel numbers of background skin correctly characterized as background, FP denotes the pixel numbers of background skin incorrectly characterized as a skin lesion, FN denotes the pixel numbers of a skin lesion incorrectly characterized as background skin. Accuracy was defined as the ability to segment all areas correctly. Sensitivity was the ability to segment skin lesions. Specificity was the ability to segment the background of the skin. F-measure is a statistical measure of a method's accuracy that considers both the recall and the precision of the method [35]. An F-measure value close to 1.0 indicated that the accuracy of the proposed approach was very high. HD was calculated to measure the resemblance of two sets of points [36]. It measures how far two subsets are from each other. The smaller the HD, the greater is their degree of similarity. Additionally, the Bland-Altman plots, known as the scatter plots of the difference against the mean between the area inside the automatic border and the area inside the manual border, were also used to visualize errors and potential bias in the border detection. Furthermore, linear regression was utilized to quantitatively compare the area inside the border drawn by the two measurements. These analyses were carried out using SPSS version 23 software (SPSS Inc., Chicago, IL, USA). A p value < 0.05 was considered to indicate statistical significance.

The algorithm was implemented on an Intel® Core™ i5-7500 CPU at 3.40 GHz with 16.00 GB RAM. All procedures were implemented with the MATLAB software package (R2018b, MathWorks Inc., Natick, MA, USA).

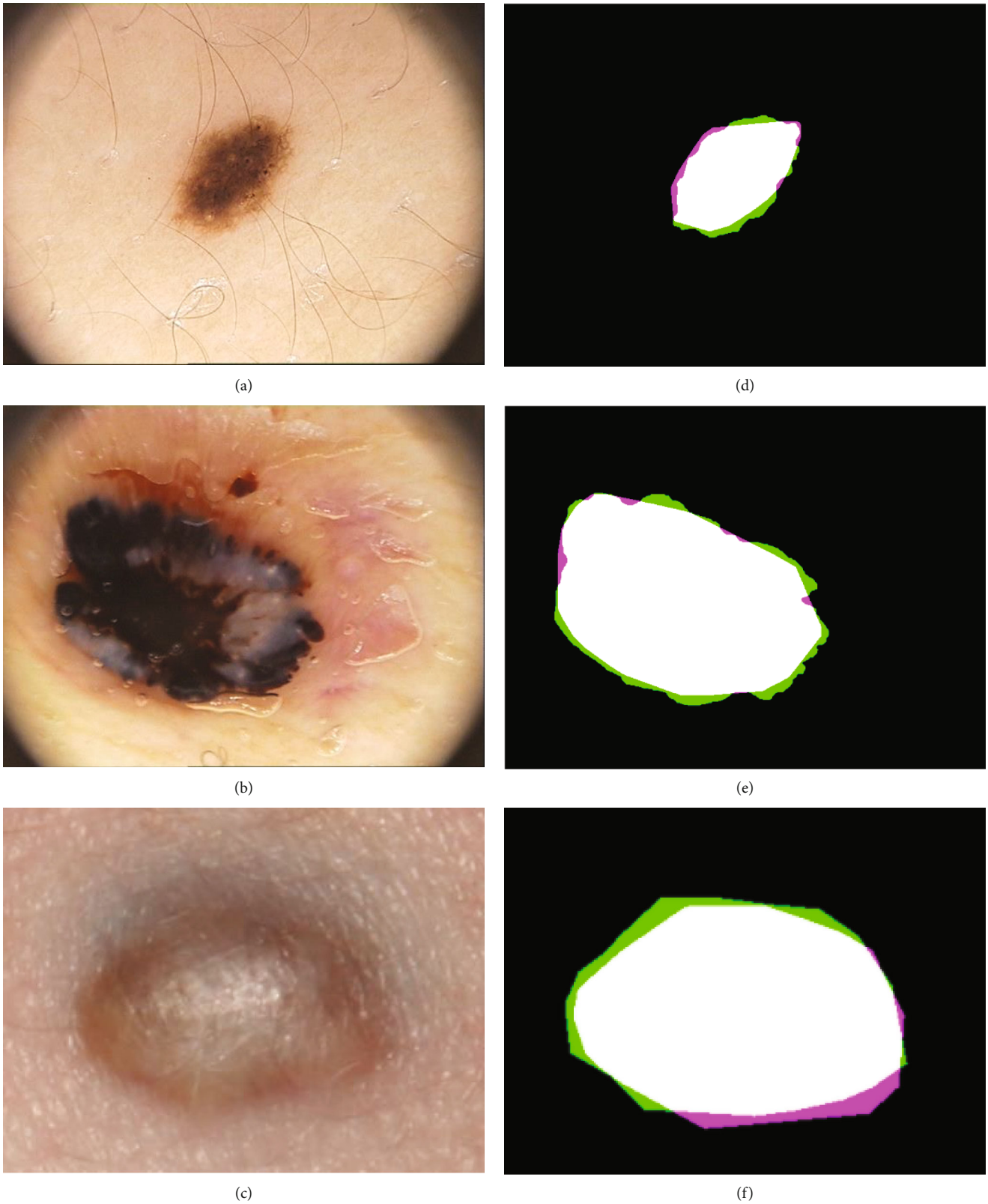
3. Results

3.1. Comparison Results of Accuracy and Run-Time for Different Numbers of Clusters at Each Level (K_1 and K_2). To obtain good segmentation results, the number of clusters for each level of HK clustering was experimentally determined. Figure 7 shows the mean accuracy and speed of the proposed method in different conditions of the number of clusters from set 1 to set 4 at each hierarchical level. The run-time performance was calculated by the total time taken from the preprocessing phase to the postprocessing phase. The run-time performance for each different condition had the following relationship: set 2 (19.2 seconds) < set 1 (19.28 seconds) < set 3 (19.72 seconds) < set 4 (20.4 seconds). This suggests that set 2 outperforms other conditions in terms of run-time performance. The experimental results showed that the optimal numbers of clusters were 2 and 3 at level 1 and level 2, respectively, which achieved an accuracy of 94.6% and a speed of 19.2 seconds.

3.2. Quantitative Evaluation of the Proposed Two-Stage Segmentation Approach in Dermoscopic and Standard Images. The performance of the segmentation-based level set scheme depends on an initial contour mask [15, 23]. Thus, the initial segmentation is a key step to increasing sensitivity. Our method was evaluated for two different datasets as shown in Table 2. The mean accuracy for each of the two atlases was greater than 90%. The F-measure for each of the two datasets was high (>0.91), and a very small difference of 0.02 was found between the two atlases. Small average HDs of 0.07 ± 0.02 and 0.09 ± 0.03 were obtained for each dataset. Our method achieved higher performance in the PH^2 database for all evaluation parameters, including sensitivity, specificity, and accuracy, than in the Dermofit database. All evaluation parameters showed promising results of over 90%, except for the Jaccard index which was 0.826 and 0.833 for the Dermofit and PH^2 data, respectively.

3.3. Comparison of Results of Segmentation between Different Disease Classes (Melanocytic Nevus and Malignant Melanoma). The proposed segmentation model was compared in two different disease classes, melanocytic nevus (common nevi, atypical nevi, and melanocytic nevi) and malignant melanoma. Table 3 shows the segmentation results that were obtained by processing the melanocytic nevus images and melanoma images. Our method obtained good accuracy for 607 skin lesion images, including an accuracy of 93.4% for 331 images of melanocytic nevus images and 95.6% for the 76 melanoma images in the Dermofit dataset. In the PH^2 dataset, the proposed method achieved an accuracy of 95.6% for the 160 melanocytic nevus images and 90.8% for the 40 melanoma images. Of note, the proposed method obtained a higher sensitivity of 92.6% for the melanocytic nevus images compared to a sensitivity of 86.4% for the melanoma images in the Dermofit dataset. Moreover, the F-measures showed 0.921 and 0.887 for the melanocytic nevus and melanoma images, respectively. In contrast, our method achieved a higher sensitivity of 92.5 for melanoma images compared to 91.7% for the melanocytic nevus images in the PH^2 dataset. The F-measures became 0.920 and 0.907 for the melanoma and melanocytic nevus images, respectively.

3.4. The Bland-Altman Plots and Linear Regression Analysis for the Area inside Each Border Detected Manually and by the Proposed Method. The mean values of the differences in

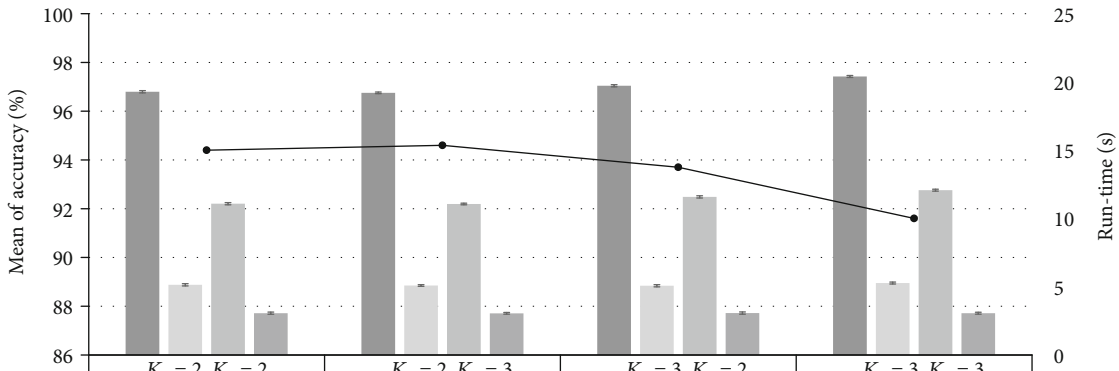


(a) (d)

(b) (e)

(c) (f)

Figure 6: Examples of lesion segmentation results from the hierarchical K-means level set scheme for the PH^2 and Dermofit atlases. (a) IMD144 dermoscopic image from the PH^2 dataset. (c) IMD168 dermoscopic image from the PH^2 dataset. (e) D105 dermoscopic image from the Dermofit dataset. (b, d, and f) Error evaluations: white pixels show true positives (TP), black pixels show true negatives (TN), pink pixels show false positives (FP), and green pixels show false negatives (FN).



	$K_1 = 2, K_2 = 2$	$K_1 = 2, K_2 = 3$	$K_1 = 3, K_2 = 2$	$K_1 = 3, K_2 = 3$
Total run-time (s)	19.28	19.2	19.72	20.4
Preprocessing (s)	5.14	5.09	5.07	5.27
HK-LS (s)	11.08	11.06	11.58	12.07
Postprocessing (s)	3.06	3.05	3.07	3.06
Mean of accuracy (%)	94.4	94.6	93.7	91.6

Total run-time (s)
Preprocessing (s)

Figure 7: Comparison of the mean accuracy and computation time (s) for different conditions of K_1 and K_2 at each level (p value < 0.05). K indicates the number of clusters at each hierarchical level.

Table 2: Quantitative evaluation of the segmentation for the Dermofit and PH^2 atlases in terms of accuracy, sensitivity, specificity, Jaccard index, Dice coefficient, F-measure, and Hausdorff distance.

Group	Jaccard index	Dice coefficient	Sensitivity	Specificity	Accuracy	F-measure	Hausdorff distance
Dermofit	0.826 ± 0.08	0.912 ± 0.07	0.919 ± 0.08	0.944 ± 0.06	0.942 ± 0.05	0.912 ± 0.07	0.07 ± 0.02
PH^2 data	0.833 ± 0.09	0.914 ± 0.05	0.923 ± 0.08	0.964 ± 0.05	0.946 ± 0.03	0.914 ± 0.05	0.09 ± 0.03

Table 3: Comparative results of segmentation between melanocytic nevus and melanoma images.

Group	Class	Jaccard index	Dice coefficient	Sensitivity	Specificity	Accuracy	F-measure	Hausdorff distance
Dermofit	Nevus	0.858 ± 0.08	0.921 ± 0.08	0.926 ± 0.09	0.936 ± 0.7	0.934 ± 0.04	0.921 ± 0.08	0.067 ± 0.02
	Melanoma	0.813 ± 0.08	0.887 ± 0.07	0.864 ± 0.07	0.971 ± 0.05	0.956 ± 0.06	0.887 ± 0.07	0.097 ± 0.038
PH^2 data	Nevus	0.823 ± 0.09	0.907 ± 0.06	0.917 ± 0.09	0.978 ± 0.03	0.956 ± 0.02	0.907 ± 0.06	0.078 ± 0.03
	Melanoma	0.855 ± 0.06	0.920 ± 0.04	0.925 ± 0.05	0.845 ± 0.07	0.908 ± 0.04	0.920 ± 0.04	0.101 ± 0.05

the Bland-Altman plots detected by the ground truth and the proposed approach are illustrated in Figures 8 and 9. In the PH^2 database, the average differences between the areas inside the borders detected by the ground truth and our method were 82.077 ± 15951.228 and −2702.371 ± 1498.615 for melanoma and melanocytic nevus images, respectively. In the Dermofit database, the average differences were 5025.59 ± 27,250.079 and −2279.233 ± 5734.517 for melanoma and melanocytic nevus images, respectively. All results showed differences close to 0, which were generally included within the limits of the agreement range.

The linear regression analysis shown in Figure 10 reports a high correlation (>0.97 and >0.96 for the Dermofit database and the PH^2 database, respectively) between the areas inside the automated extracted borders and contours of the ground truth. These results showed that the proposed segmentation method strongly correlated with the segmentation ground truth datasets.

3.5. Comparison of Segmentation Performance with Other Automated Segmentation Methods. The proposed method was compared with traditional segmentation methods in the same dataset. Results of the comparison between traditional classifiers and the proposed method are summarized in Table 4. Traditional classifiers showed relatively poorer results for melanocytic lesion segmentation compared to the proposed method. Specifically, the Otsu thresholding method showed the lowest segmentation accuracies in the Dermofit and PH^2 data (68.3% and 65.2%, respectively). The proposed method achieved a higher specificity of 94.4% than that of K-means

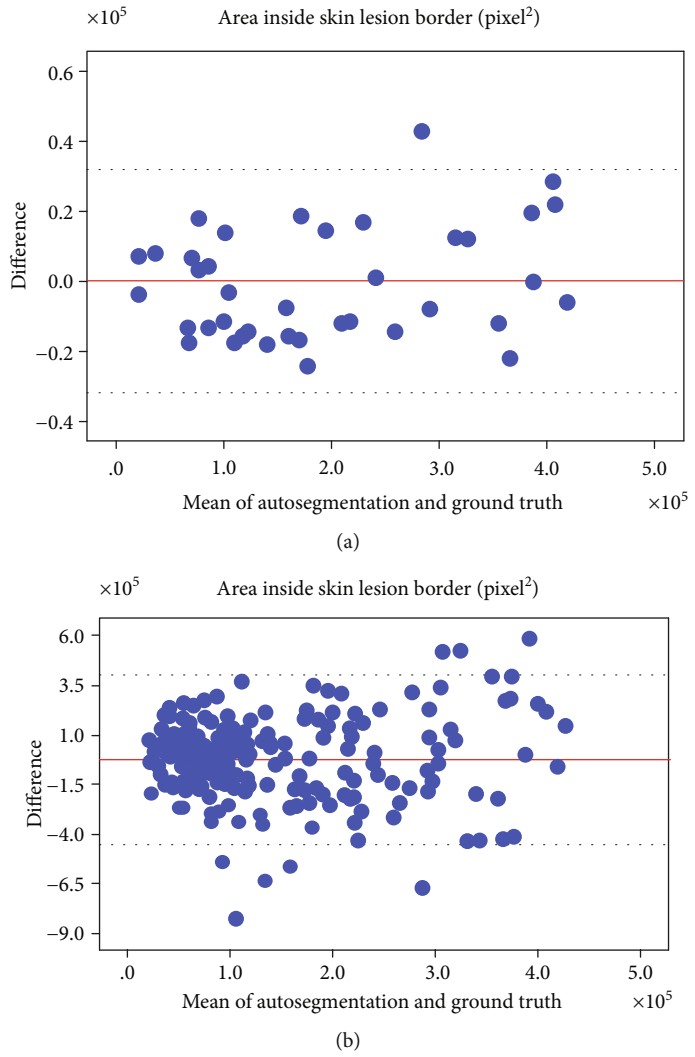


FIGURE 8: The Bland-Altman plots between the ground truth and automated segmentation of skin images obtained from the PH^2 database. (a) Area inside the skin lesion border of the melanoma image and (b) the melanocytic nevus image (unit: $pixels^2$).

clustering implemented on the same color space (CIE $L * a * b$). In addition, Pennisi et al. [37] segmented melanoma lesion images in the PH^2 database using ASLM with Delaunay triangulation. They showed the accuracy of 89.7% in the PH^2 data. In contrast, the overall accuracy of the proposed method was also better than that of the other techniques when the same dataset was used. These results demonstrated the feasibility of the proposed method for skin image segmentation.

In the comparative results between U-net [40] and our method for the PH^2 dataset, although U-net performed better according to the Jaccard index of (0.87 ± 0.19) and Dice coefficient (0.93 ± 0.13) compared to our method, U-net produced a much larger standard deviation than that of the proposed method (Table 5). Moreover, our method had better segmentation results for the Dermofit dataset compared to that of U-net [41]. These results confirm its effectiveness for melanocytic skin lesion segmentation in standard images compared to U-net.

4. Discussion

The segmentation of skin lesions in dermoscopic and standard images is crucial for quantifying the clinical diagnostic factors of melanoma lesions. The segmentation accuracy can greatly affect the next diagnostic procedure [45]. One issue with the level set model is its sensitivity to the initial contours. Recently, machine learning algorithms, such as U-net, have emerged as reliable segmentation methods for skin lesion images. However, the limited training dataset is a challenging task for skin lesion segmentation. The important challenge in machine learning algorithms is that these models require a large training set to reduce overfitting. Some cases still show a low performance due to low contrast and hair artifacts. Current state-of-the-art research using machine learning algorithms is sometimes required on postprocessing techniques, such as level sets [46]. Another challenge in machine learning such as CNN is that, when a network goes deeper, it is

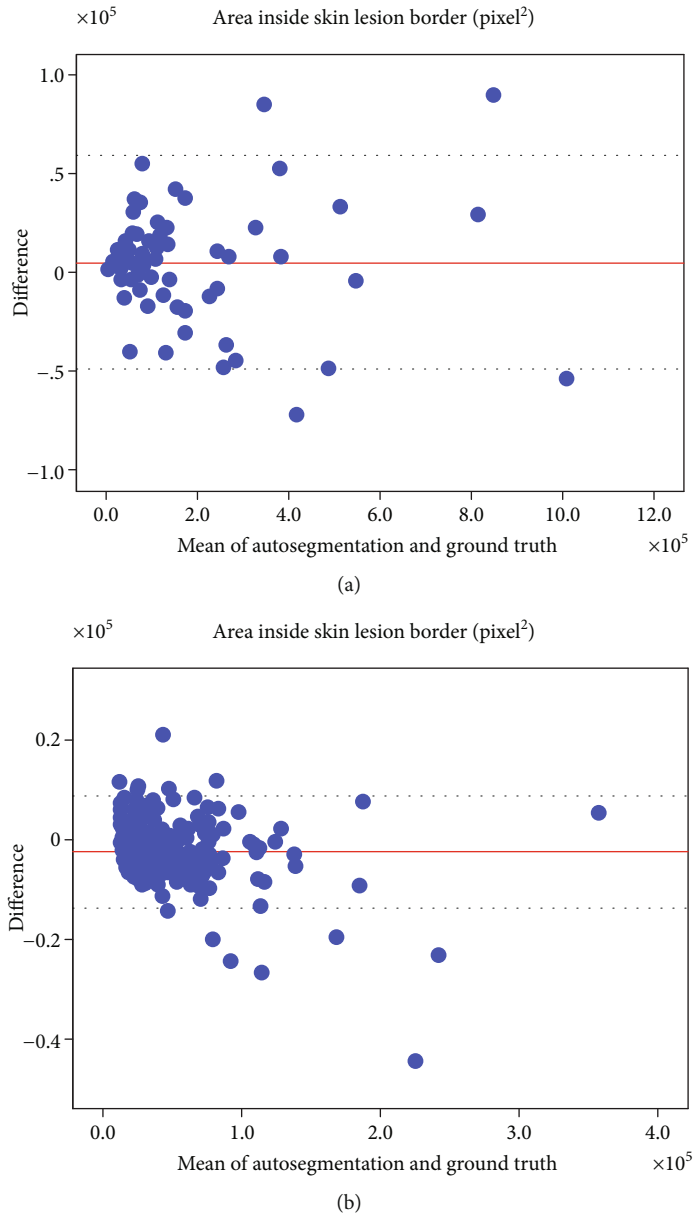
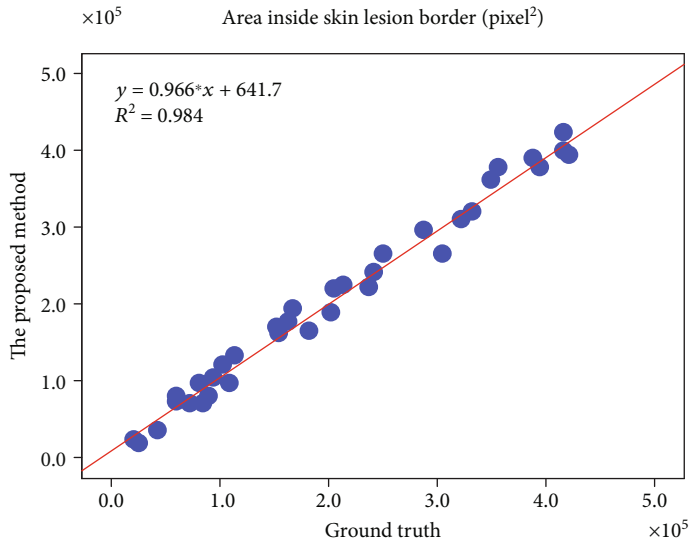


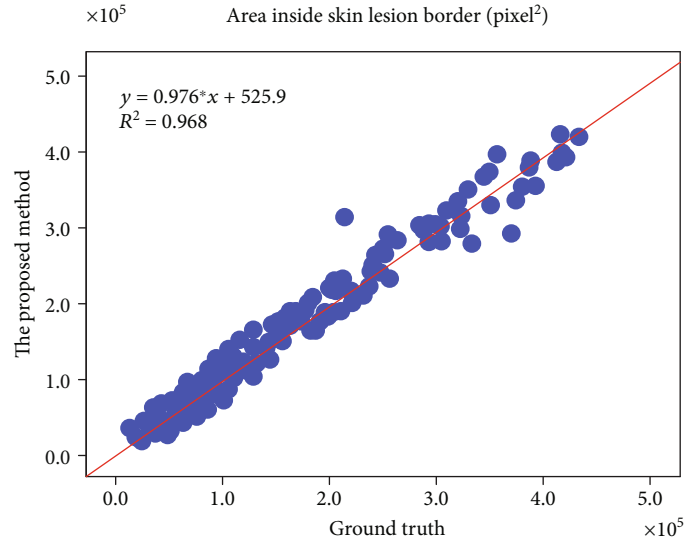
Figure 9: The Bland-Altman plots between the ground truth and automated segmentation of skin images obtained from the Dermofit database. (a) Area inside the skin lesion border of the melanoma image and (b) the melanocytic nevus image (unit: pixels^2).

difficult to tune the parameters of the early layers [8]. To tackle these problems, the purpose of this study was to propose a new two-stage segmentation model which integrates the Distance Regularization Level Set Evolution and the hierarchical K-means clustering. The proposed method that combines two different methods has the advantage of improving the final result of the image segmentation process, such as accurately defining the initial contours, and finding the approximate location of the lesion. The quantitative experimental results revealed that the proposed method yielded significantly better results compared to other traditional level set models, and has a certain advantage over the segmentation results of U-net in standard images.

The contribution of this paper can be summarized in the following aspects. Firstly, the proposed model integrates hierarchical K-means clustering with DRLSE. Some studies have attempted to use a mono-K-mean clustering-based level set evolution model with unsatisfactory results [15, 28]. However, this study showed the reliable accuracy of the segmentation of skin lesions under intrinsic noise and artifacts. To the best of our knowledge, no such studies for skin lesion segmentation have been reported previously. Secondly, the controlling parameters of level set segmentation are now derived from the results of the simple decision tree approach by using a set of if-then rules. Thirdly, the experimental results indicate that a new gray-scale image by using only the color components of a and b from CIE $L * a * b$ color



(a)



(b)

Figure 10: Continued.

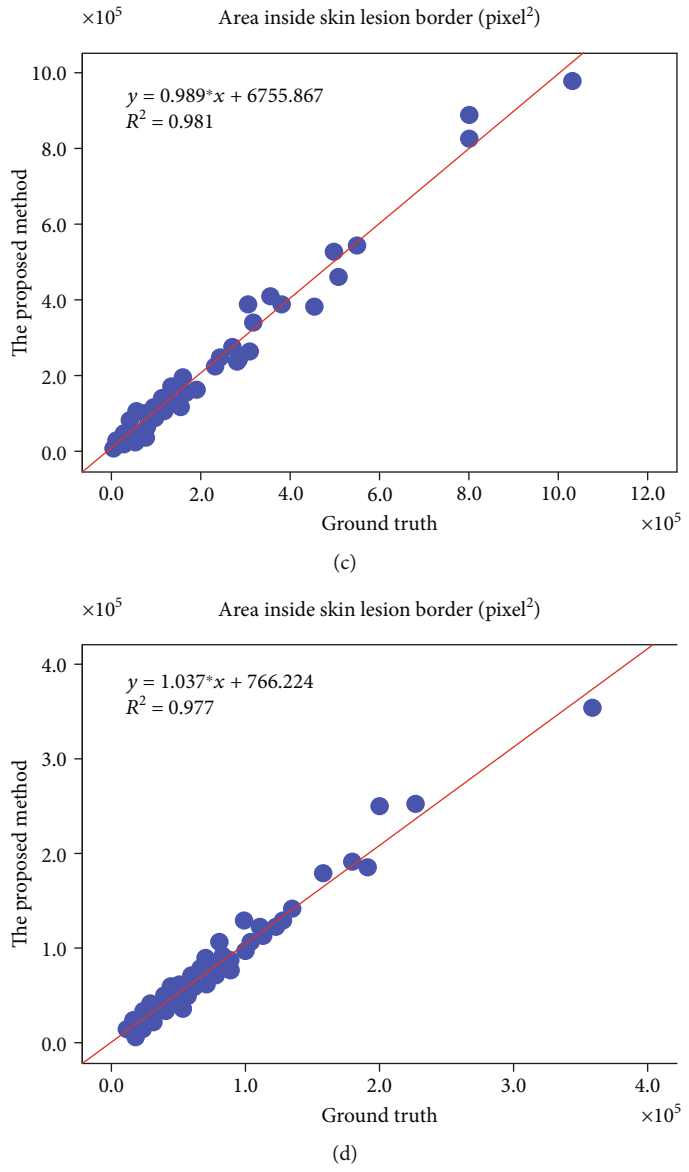


Figure 10: Linear regression analysis between the manual and automated segmentations of skin images obtained from the PH^2 and Dermofit databases. (a) Area inside the skin lesion border of melanoma images from the PH^2 database, (b) melanocytic nevus images from the PH^2 database, (c) melanoma images from the Dermofit database, and (d) melanocytic nevi images from the Dermofit database (unit: pixels^2).

Table 4: Comparative results of segmentation between the existing and proposed methods for images from the PH^2 and Dermofit atlases.

Methods	Dermofit			PH^2 data		
	SEN	SPE	ACC	SEN	SPE	ACC
Otsu with RGB (MATLAB 2018b)	0.611	0.723	0.683	0.522	0.706	0.652
Level set with RGB (MATLAB 2018b)	0.712	0.878	0.805	0.719	0.800	0.784
FC-LS with RGB [28]	0.873	0.926	0.918	0.891	0.914	0.904
Adaptive thresholding with YIQ [38]	0.618	0.980	0.937	0.703	0.949	0.879
K-means with CIELAB [21]	0.809	0.789	0.824	0.869	0.953	0.932
Local binary pattern clustering [39]	0.787	0.923	0.704	0.884	0.948	0.859
Proposed method (HK-LS with CIELAB)	0.919	0.944	0.942	0.923	0.964	0.946

*FC-LS: fuzzy C-mean thresholding-based level set; HK-LS: hierarchical *K*-means clustering-based level set; SEN: sensitivity; SPE: specificity; ACC: accuracy.

Table 5: Comparative Jaccard index and Dice coefficient results for segmentation of images from the PH^2 and Dermofit atlases by U-net and the proposed method.

Group	Dermofit		PH^2 data	
	Jaccard index	Dice coefficient	Jaccard index	Dice coefficient
U-net [40, 41]	0.781	0.887	0.87 ± 0.19	0.93 ± 0.13
U-net with illumination-based transformation [42]	0.774 ± 0.006	0.867 ± 0.004	0.756 ± 0.009	0.853 ± 0.007
Mutual bootstrapping DCNN [43]	—	—	0.894	0.942
FCN-16s [44]			0.802	0.881
Proposed method	0.826 ± 0.008	0.912 ± 0.07	0.833 ± 0.09	0.914 ± 0.05

space makes it less sensitive to illumination artifacts. Finally, we also evaluated the proposed method on two different datasets including the PH^2 database (dermoscopic image repository) and the Dermofit database (standard image repository). All skin lesions were segmented with high accuracy (>94%) and high correlation (>0.96) of the ground truth in the two databases. The segmentation results outperformed other initial estimation methods for level set models of melanoma and nonmelanoma images with various artifacts.

One of the main concerns of existing image segmentation methods resides mainly in the noise and artifacts of dermoscopic and standard images [47]. Moreover, another factor that complicates the lesion segmentation is the low contrast of the lesion boundaries [48]. Our designated model improved the segmentation performance in most cases, especially the proportion of true positive results. Experimental results show that this approach is insensitive to the low contrast between background around the lesion and skin lesion pixels. The main difference between the proposed method and other models was that only the color channels of CIE $L*a*b$ were used to constitute a new gray-scale image for the initial contour mask. Unlike the RGB and CMYK color spaces, the CIE $L*a*b$ is designed to approximate human vision. This color space is approximately perceptually uniform because the similarities between the perceived and the measured color are proportional [11]. Additionally, CIE $L*a*b$ color space is known to be less sensitive to artifacts from digital cameras and scanner images [21].

When our method and other segmentation methods are compared, especially with a classifier such as U-net, our model can get better segmentation results in standard images. The latest deep learning segmentation approaches such as U-net have been applied to segment melanoma lesions because these algorithms can handle complex patterns, but the limited quality training dataset and degradation problems are often limitations [1]. In addition, data augmentation, such as flipping, rotating, shifting, scaling, and changing the contrast of the original image, is usually required when the classifier is trained on medical images [49]. However, it is easier to lose the important features of melanocytic skin lesions in data augmentation because the proportional size of the skin lesion on the images is very small [50].

Although the proposed model achieved admirable segmentation accuracy in most of the images in the two independent atlases, there were cases where the proposed model revealed the need for further improvement. The challenge in the proposed method is the increase of the run-time when the size of the image is large, compared with deep learning approaches. The proposed method can be further improved for the more effective segmentation pipeline, in terms of average run-time.

5. Conclusions

The segmentation of the skin lesions is regarded as very challenging because of the low contrast between the lesion and the surrounding skin, the existence of various artifacts, and different imaging acquisition conditions. The traditional model such as the region-based active contour model has often failed when applied to images containing inhomogeneities. These are very sensitive to parameter tuning. The appropriate initialization and optimal configuration of controlling parameters in the presence of various artifacts are important to obtain the accurate performance of the level set segmentation. The important challenge in machine learning algorithms is that these models require a large training set to reduce overfitting. Current state-of-the-art research using machine learning algorithms is usually required on postprocessing techniques, such as level sets. The contribution of this study is to propose a new two-stage segmentation model in dermoscopic and standard images. This method integrates a new hierarchical K-means and level set approach. For the initial estimation of the level set function, the hybrid hierarchical K-means clustering was carried out. After initial segmentation by the hybrid HK clustering, DRLSE was implemented to achieve fine border segmentation. Moreover, only the color channels of a and b from CIE $L*a*b$ were used by this model to obtain robust image segmentation results in the presence of noise and artifacts. The generalization ability of the proposed model was validated by the independent testing of two publicly available databases. The experimental results showed the superior performance of the proposed method compared to other traditional level set models, and a certain advantage over the segmentation results of U-net in standard images. Additionally, the linear regression analysis demonstrated a good correlation of >0.98 and >0.96 with the proposed method for melanoma and melanocytic nevus images. The proposed model gives accurate segmentation results and requires a small dataset because our model is not sensitive to parameter tuning. Our experimental results revealed that integrating

hierarchical K-means clustering and DRLSE had high clinical applicability even in the presence of various artifacts and small datasets. The proposed model may facilitate the combination of machine learning and level set models in skin lesion images.

Data Availability

The datasets that were used in this study are openly available in the PH^2 database (https://www.fc.up.pt/addi/ph2%20database.html) [19] and the Dermofit image library (https://licensing.edinburgh-innovations.ed.ac.uk) [20].

Conflicts of Interest

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

Acknowledgments

This work was supported by the Dongguk University Research Fund of 2020 (S-2020-G0001-00047).

References

[1] 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.

[2] K. A. Freedberg, A. C. Geller, D. R. Miller, R. A. Lew, and H. K. Koh, "Screening for malignant melanoma: a cost-effectiveness analysis," *Journal of the American Academy of Dermatology*, vol. 41, no. 5, pp. 738–745, 1999.

[3] C. M. Balch, A. C. Buzaid, S.-J. Soong et al., "Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma," *Journal of Clinical Oncology*, vol. 19, no. 16, pp. 3635–3648, 2001.

[4] E. Flores and J. Scharcanski, "Segmentation of melanocytic skin lesions using feature learning and dictionaries," *Expert Systems with Applications*, vol. 56, no. 1, pp. 300–309, 2016.

[5] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *The Lancet Oncology*, vol. 3, no. 3, pp. 159–165, 2002.

[6] M. Silveira, J. C. Nascimento, J. S. Marques et al., "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 35–45, 2009.

[7] P. Carli, V. de Giorgi, L. Naldi, and G. Dosi, "Reliability and inter-observer agreement of dermoscopic diagnosis of melanoma and melanocytic naevi," *Dermoscopy Panel. European journal of cancer prevention: the official journal of the European Cancer Prevention Organisation (ECP)*, vol. 7, no. 5, pp. 397–402, 1998.

[8] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.

[9] B. S. Lin, K. Michael, S. Kalra, and H. R. Tizhoosh, "Skin lesion segmentation: U-nets versus clustering," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–7, Honolulu, HI, USA, 2017.

[10] A. Masood and A. A. Al-Jumaily, "Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms," *International journal of biomedical imaging*, vol. 2013, Article ID 323268, 22 pages, 2013.

[11] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Transactions on Medical Imaging*, vol. 20, no. 3, pp. 233–239, 2001.

[12] M. Emre Celebi, Y. Alp Aslandogan, W. V. Stoecker, H. Iyatomi, H. Oka, and X. Chen, "Unsupervised border detection in dermoscopy images," *Skin Research and Technology*, vol. 13, no. 4, pp. 454–462, 2007.

[13] M. Silveira and J. S. Marques, "Level set segmentation of dermoscopy images," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 173–176, Paris, France, 2008.

[14] M. Nourmohamadi and H. Pourghassem, "Dermoscopy image segmentation using a modified level set algorithm," in *2012 Fourth International Conference on Computational Intelligence and Communication Networks*, pp. 286–290, Mathura, India, 2012.

[15] B. N. Li, C. K. Chui, S. Chang, and S. H. Ong, "Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation," *Computers in Biology and Medicine*, vol. 41, no. 1, pp. 1–10, 2011.

[16] R. B. Oliveira, M. E. Filho, Z. Ma, J. P. Papa, A. S. Pereira, and J. M. R. S. Tavares, "Computational methods for the image segmentation of pigmented skin lesions: a review," *Computer Methods and Programs in Biomedicine*, vol. 131, pp. 127–141, 2016.

[17] P. Schmid, "Segmentation of digitized dermatoscopic images by two-dimensional color clustering," *IEEE Transactions on Medical Imaging*, vol. 18, no. 2, pp. 164–171, 1999.

[18] T. Donadey, C. Serruys, A. Giron et al., "Boundary detection of black skin tumors using an adaptive radial-based approach," in *Medical Imaging 2000: Image Processing*, vol. 3979, pp. 810–816, International Society for Optics and Photonics, 2000.

[19] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH2—a dermoscopic image database for research and benchmarking," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5437–5440, Osaka, Japan, 2013.

[20] L. Ballerini, R. B. Fisher, B. Aldridge, and J. Rees, "A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions," *Color Medical Image Analysis*, vol. 6, pp. 63–86, 2013.

[21] A. Agarwal, A. Issac, M. K. Dutta, K. Riha, and V. Uher, "Automated skin lesion segmentation using K-means clustering from digital dermoscopic images," in *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 743–748, Barcelona, Spain, 2017.

[22] R. Kaur, S. Gupta, and P. S. Sandhu, "Optimization color quantization in L ∗ A ∗ B ∗ color space using particle swarm optimization," in *Proceedings of International conference on Intelligent Computational Systems*, pp. 1–4, Bangkok, Thailand, 2011.

[23] A. Masood and A. A. Al-Jumaily, "Fuzzy C mean thresholding based level set for automated segmentation of skin lesions," *Journal of Signal and Information Processing*, vol. 4, no. 3, pp. 201–206, 2013.

[24] T.-S. Chen, T.-H. Tsai, Y.-T. Chen et al., "A combined K -means and hierarchical clustering method for improving the

clustering efficiency of microarray," in *2005 International symposium on intelligent signal processing and communication systems*, pp. 405–408, Hong Kong, China, 2005.

[25] H. Chipman and R. Tibshirani, "Hybrid hierarchical clustering with applications to microarray data," *Biostatistics*, vol. 7, no. 2, pp. 286–301, 2006.

[26] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pp. 2161–2168, New York, NY, USA, 2006.

[27] K. Arai and A. R. Barakbah, "Heirarchical K-means: an algorithm for centroid initialization for centroids initialization for K-means," *Reports of the Faculty of Science and Engineering*, vol. 36, no. 1, pp. 25–31, 2007.

[28] A. Masood, A. A. Al Jumaily, A. N. Hoshyar, and O. Masood, "Automated segmentation of skin lesions: modified fuzzy C mean thresholding based level set method," in *IEEE INMIC*, pp. 201–206, Lahore, Pakistan, 2013.

[29] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.

[30] J. A. Sethian and P. Smereka, "Level set methods for fluid interfaces," *Annual Review of Fluid Mechanics*, vol. 35, no. 1, pp. 341–372, 2003.

[31] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 271–293, 2002.

[32] A. El-Baz and J. S. Suri, *Level set method in medical imaging segmentation*, CRC Press, 2019.

[33] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3243–3254, 2010.

[34] P. R. Bai, Q. Y. Liu, L. Li, S. H. Teng, J. Li, and M. Y. Cao, "A novel region-based level set method initialized with mean shift clustering for automated medical image segmentation," *Computers in Biology and Medicine*, vol. 43, no. 11, pp. 1827–1832, 2013.

[35] C. Van Rijsbergen, "Information Retrieval," in *London available on internet*, vol. 30butterworths, 2nd edition, 1979.

[36] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[37] A. Pennisi, D. D. Bloisi, D. Nardi, A. R. Giampetruzzi, C. Mondino, and A. Facchiano, "Skin lesion image segmentation using Delaunay triangulation for melanoma detection," *Computerized Medical Imaging and Graphics*, vol. 52, pp. 89–103, 2016.

[38] A. Gupta, A. Issac, M. K. Dutta, and H.-H. Hsu, "Adaptive thresholding for skin lesion segmentation using statistical parameters," in *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 616–620, Taipei, 2017.

[39] P. M. M. Pereira, R. Fonseca-Pinto, R. P. Paiva et al., "Dermoscopic skin lesion image segmentation based on local binary pattern clustering: comparative study," *Biomedical Signal Processing and Control*, vol. 59, pp. 1–12, 2020.

[40] Z. Al Nazi and T. A. Abir, "Automatic Skin Lesion Segmentation and Melanoma Detection: Transfer Learning Approach with U-Net and DCNN-SVM," in *Proceedings of International Joint Conference on Computational Intelligence*, pp. 371–381, Singapore, 2020.

[41] 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 2015*, pp. 234–241, Munich, Germany, 2015.

[42] K. Abhishek, G. Hamarneh, and M. S. Drew, "Illumination-based transformations improve skin lesion segmentation in dermoscopic images," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognision Workshops (CVPRW)*, pp. 3132–3141, Seattle, WA, USA, 2020.

[43] Y. Xie, J. Zhang, Y. Xia, and C. Shen, "A mutual bootstrapping model for automated skin lesion segmentation and classification," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2482–2493, 2020.

[44] K. Zafar, S. O. Gilani, A. Waris et al., "Skin lesion segmentation from dermoscopic images using convolutional neural network," *Sensors*, vol. 20, no. 1601, pp. 1–14, 2020.

[45] M. H. Jafari, N. Karimi, E. Nasr-Esfahani et al., "Skin lesion segmentation in clinical images using deep learning," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 337–342, Cancun, Mexico, 2016.

[46] Y. Yang, C. Feng, and R. Wang, "Automatic segmentation model combining U-Net and level set method for medical images," *Expert Systems with Applications*, vol. 153, pp. 1–9, 2020.

[47] H. Zare, M. T. B. Toossi, M. E. Celebi, T. Mendonca, and J. S. Marques, "Early detection of melanoma in dermoscopy of skin lesion images by computer vision based system," in *Dermoscopy Image Analysis*, pp. 345–384, CRC Press, 2015.

[48] C. Barata, M. E. Celebi, and J. S. Marques, "Improving dermoscopy image classification using color constancy," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 1146–1152, 2015.

[49] F. Pollastri, F. Bolelli, R. P. Palacios, and C. Grana, "Improving skin lesion segmentation with generative adversarial networks," in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), IEEE*, pp. 442-443, Karlstad, Sweden, 2018.

[50] M. E. Celebi, H. A. Kingravi, B. Uddin et al., "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 362–373, 2007.

Hindawi
BioMed Research International
Volume 2021, Article ID 5519144, 10 pages
https://doi.org/10.1155/2021/5519144

Research Article

CT Texture Analysis for Preoperative Identification of Lymphoma from Other Types of Primary Small Bowel Malignancies

Shunli Liu,[1] Chuanyu Zhang,[1] Ruiqing Liu,[2] Shaoke Li,[1] Fenglei Xu,[1] Xuejun Liu,[1] Zhiming Li,[1] Yabin Hu,[1] Yaqiong Ge,[3] Jiao Chen ,[4] and Zaixian Zhang [1]

[1]*Department of Radiology, The Affiliated Hospital of Qingdao University, Qingdao, Shandong, China*
[2]*Department of Gastrointestinal Surgery, The Affiliated Hospital of Qingdao University, Qingdao, Shandong, China*
[3]*GE Healthcare China, Shanghai, China*
[4]*Department of Radiology, Yantai Yuhuangding Hospital, Yantai, Shandong, China*

Correspondence should be addressed to Jiao Chen; chenjiaocn@126.com and Zaixian Zhang; befate@126.com

Received 12 January 2021; Revised 18 March 2021; Accepted 22 March 2021; Published 5 April 2021

Academic Editor: Lin Gu

Objectives. To explore the application of computed tomography (CT) texture analysis in differentiating lymphomas from other malignancies of the small bowel. *Methods*. Arterial and venous CT images of 87 patients with small bowel malignancies were retrospectively analyzed. The subjective radiological features were evaluated by the two radiologists with a consensus agreement. The region of interest (ROI) was manually delineated along the edge of the lesion on the largest slice, and a total of 402 quantified features were extracted automatically from AK software. The inter- and intrareader reproducibility was evaluated to select highly reproductive features. The univariate analysis and minimum redundancy maximum relevance (mRMR) algorithm were applied to select the feature subsets with high correlation and low redundancy. The multivariate logistic regression analysis based on texture features and radiological features was employed to construct predictive models for identification of small bowel lymphoma. The diagnostic performance of multivariate models was evaluated using receiver operating characteristic (ROC) curve analysis. *Results*. The clinical data (age, melena, and abdominal pain) and radiological features (location, shape, margin, dilated lumen, intussusception, enhancement level, adjacent peritoneum, and locoregional lymph node) differed significantly between the nonlymphoma group and lymphoma group ($p < 0.05$). The areas under the ROC curve of the clinical model, arterial texture model, and venous texture model were 0.93, 0.92, and 0.87, respectively. *Conclusion*. The arterial texture model showed a great diagnostic value and fitted performance in preoperatively discriminating lymphoma from nonlymphoma of the small bowel.

1. Introduction

Primary small bowel malignancies (PSBM) are relatively rare, representing less than 3% of all gastrointestinal tract malignancies [1]. Early detection and differential diagnosis of small bowel tumors can be challenging for both clinicians and radiologists because of the nonspecificity of clinical signs and tumor deep location [2].

Although primary small bowel lymphoma (PSBL) accounts for a small proportion of PSBM (approximately 15%), the incidence has been increasing [3]. Surgery is considered to be the first-line treatment for most small bowel tumors, especially malignant and borderline tumors; however, chemotherapy seems to have a survival benefit over surgery alone in treating PSBL [4]. Due to the different management, it is of great significance to improve the imaging approach for preoperative identification of PSBL and other types of PSBM.

Clinically, several modalities are useful to investigate suspected small bowel tumors including capsule endoscopy, computed tomography (CT), or magnetic resonance (MR) enterography or enteroclysis. However, CT imaging is

currently recognized as the mostly used and valuable tool for evaluating small bowel masses. In clinical work, the ability of dynamic contrast-enhanced CT imaging to detect and formulate differential diagnosis of intestinal tumors has been well known. Previous studies have investigated the capacity of conventional multiphasic CT imaging and the diagnostic ability of the CT attenuation values for distinguishing different pathological types of small bowel tumors [5–8]. In addition, Yang et al. have reported the value of dual-energy spectral CT imaging and iodine quantification in the preoperative differentiation between PSBL and small bowel adenocarcinoma [9]. However, there are a number of overlaps between CT findings of PSBL and other small bowel tumors. Therefore, conventional CT imaging mainly focuses on subjective and qualitative features and provides limited quantitative parameters for differential diagnosis of small bowel tumors [6].

Recently, the advances in CT texture analysis have improved the processing capacity for tumor heterogeneity at imaging and provide indirect information of the tumor microenvironment within a certain range of quantitative parameters [10]. CT texture analysis reflects the distribution and relationship of pixels in CT images, which could reveal the subtle differences that cannot be recognized by the human eyes and make up for the shortcomings of conventional CT imaging.

The potential use of CT texture analysis has been widely studied in tumor research and demonstrates to be valuable for prognostic prediction in pathological features, overall survival, and treatment response of multiple tumor types [11–14]. Several studies have declared that CT texture analysis may potentially serve as biomarkers for preoperative risk stratification of small bowel gastrointestinal stromal tumors (GIST) [15, 16]. To the best of our knowledge, the diagnostic ability of CT texture analysis has not yet been fully studied in differentiating lymphoma from other types of PSBM.

We aimed to establish preoperative prediction models based on arterial and venous CT images for discriminating patients with PSBL from those with other types of PSBM. We also compared the predictive efficacy of texture models with that of subjective radiological features.

2. Materials and Methods

2.1. Patients. This retrospective study was approved by the local ethics committee, and the requirement for informed consent was waived.

From January 2013 to December 2019, a total of 87 patients with a diagnosis of PSBM who underwent contrast-enhanced CT examination at our hospital were identified and included in this study (Figure 1).

The inclusion criteria were as follows: (1) a pathological confirmation of small bowel malignant tumors based on histological examinations of biopsied or resected tissues and (2) availability of contrast-enhanced CT examination before treatment.

The exclusion criteria were as follows: (1) a history of other primary malignancies ($n = 1$); (2) loss of contrast-enhanced CT images ($n = 4$); (3) located in the duodenal papilla owing to an extremely rare incidence of lymphomas in the ampulla region ($n = 53$) [17]; (4) poor visualization of the lesion due to peristaltic motion, insufficient distention, or obvious artifacts ($n = 5$); and (5) pathological types with less than 3 cases ($n = 6$).

The clinical data, including gender, age, histologic type, and clinical symptoms (melena, abdominal pain, and intestinal obstruction), was obtained and recorded from the electronic medical record system. Patient and tumor characteristics are summarized in Tables 1 and 2.

2.2. Image Acquisition and Analysis. CT examinations were performed on a multidetector row scanner (SOMATOM Definition Flash, Siemens Medical Systems; iCT 256, Philips Healthcare; or Optima CT670, GE Healthcare). All patients were requested to fast for at least six hours before the procedure. All patients underwent abdominal CT protocol or CT enterography. For abdominal CT protocol, patients received 600–1000 mL water orally prior to the examination. For CT enterography, patients were encouraged to drink 1000–2000 mL 20% mannitol for over 40–60 min prior to the CT scanning.

All patients were in the supine position, and the scan covered the entire abdomen. The patients were trained to hold their breath during CT scanning. Intravenous 1.0 mL/kg contrast medium (iohexol injection, 300 mg/mL, Beilu Pharmaceutical Co. Ltd., Beijing, China) was injected at a flow rate of 3.0–3.5 mL/s using a power injector (Ulrich CT Plus 150, Ulrich Medical), followed by a saline flush (20 mL). Arterial phase scanning and venous phase scanning were performed at 30 seconds and 70 seconds, respectively, after initiation of contrast material injection.

The CT scanning parameters were as follows: automatic tube current and tube voltage 120 kV, detector collimation 64×0.6 or 128×0.625 mm, matrix 512×512, slice thickness 5 mm, slice interval 5 mm, and reconstructed section thickness 1.25 or 2 mm.

2.3. Image Interpretation and Segmentation. Transverse reconstructed CT images were reviewed and interpreted by two abdominal radiologists (reader 1 and reader 2, with 3 and 10 years of working experience, respectively) without knowledge of the clinical data of patients. The radiological features derived from subjective CT interpretation were evaluated by the two radiologists with a consensus agreement. The observed contents were recorded, as follows: (1) location (duodenum, jejunum, or ileum), (2) shape (regular or irregular), (3) margin (well defined or ill defined), (4) lumen dilation (positive or negative), (5) intussusception (positive or negative), (6) enhancement pattern (homogeneous or heterogeneous), (7) enhancement level (mild, moderate, or high), (8) adjacent peritoneum (clear or unclear), and (9) locoregional lymph node (enlarged or non-enlarged). The interpretation criteria were mainly based on clinical experience or previous studies [5, 18]. Lumen dilation was considered to be present if intralesion dilation of the lumen was observed on at least two planes. For the enhancement pattern, masses with intratumoral low-enhancing or nonenhancing areas were considered as heterogeneous enhancement. The high

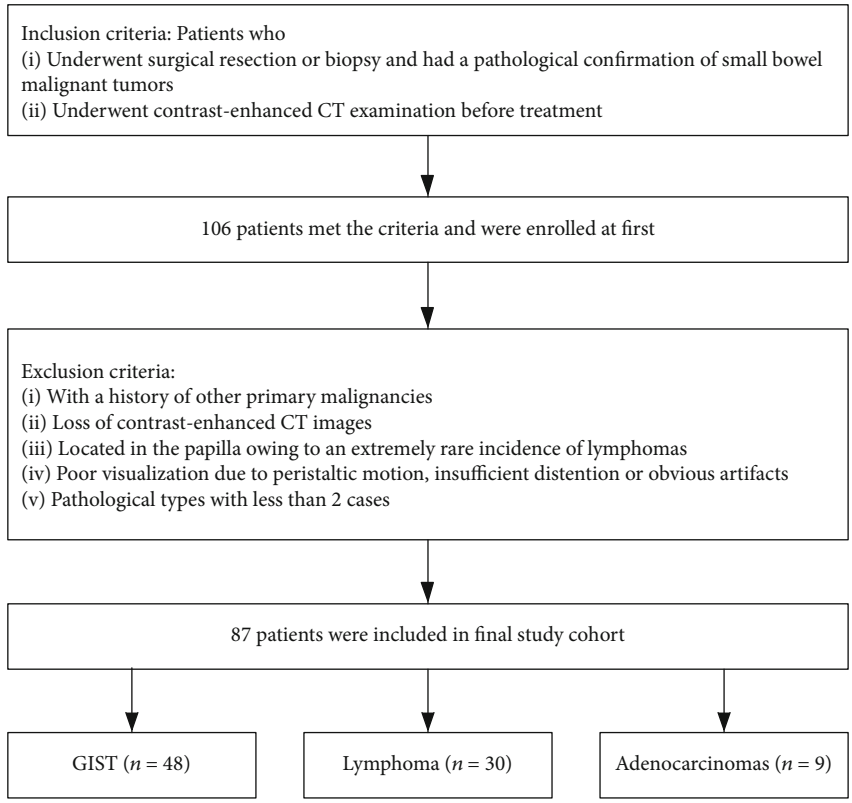


Figure 1: Flowchart of the patient inclusion and exclusion. Data in parentheses are the numbers of patients.

Table 1: Clinicopathological characteristics of patients with primary small bowel malignancies.

Feature	$n = 87$ (percentage)
Gender	
Male	47 (54.0%)
Female	40 (46.0%)
Age	
<50 years	18 (20.7%)
≥50 years	69 (79.3%)
Location	
Duodenum	4 (4.6%)
Jejunum	32 (36.8%)
Ileum	51 (58.6%)
Histologic type	
GIST	48 (55.2%)
Lymphoma	30 (34.5%)
Adenocarcinomas	9 (10.3%)

enhancement level was defined as solid components with the CT attenuation greater than 90 HU in arterial CT images. Lesions with CT attenuation less than 60 HU in arterial CT images were considered to be mild enhancement. An enlarged lymph node was considered to be present when a lymph node was greater than 10 mm in a short-axis diameter.

The arterial and venous reconstructed CT images were segmented using ITK-SNAP 3.8.0 (http://www.itksnap.org,USA).The region of interest (ROI) was manually delineated along the edge of the lesion on the largest slice by reader 1, excluding bowel lumen and blood vessels. To evaluate the reproducibility of feature extraction, 30 cases of CT images were randomly selected for calculating inter- and intraclass correlation coefficients (ICCs). Reader 2 independently drew the ROIs of the 30 cases. Reader 1 repeated the segmentations three months later. Texture features with an ICC greater than 0.75 suggested good agreement [19].

2.4. Feature Extraction and Selection. The original images were normalized before feature extraction, and texture features were automatically calculated and extracted by using AK software (Analysis Kit 1.0.3; GE Healthcare, China). A total of 402 quantified features were extracted from the delineated ROIs, including 42 histograms, 15 form factor features, 180 gray level run-length matrix (GLRLM) features with an offset of 1/4/7, 154 gray level cooccurrence matrix (GLCM) features with an offset of 1/4/7, and 11 grey level size zone matrix (GLSZM) features. The image analysis and feature extraction were performed separately for the arterial phase and venous phase based on CT images, following the same procedure.

We followed a four-step procedure to identify robust and predictive texture features. First, the texture features with both inter- and intrareader ICCs > 0.75 were retained for further procedure (Supplementary materials Part 1). Second,

Table 2: The univariate analysis of clinical data and radiological features between the nonlymphoma group and the lymphoma group in patients with primary small bowel malignancies.

Feature	Nonlymphoma ($n = 57$)	Lymphoma ($n = 30$)	p value	FDR-adjusted p values
Gender			0.206	0.050
Male	28	19		
Female	29	11		
Age			0.008	0.029
<50	7	11		
≥50	50	19		
Melena			0.001	0.011
Negative	29	26		
Positive	28	4		
Abdominal pain			0.008	0.029
Negative	28	6		
Positive	29	24		
Intestinal obstruction			0.177	0.046
Negative	53	30		
Positive	4	0		
Location			0.037	0.039
Duodenum	3	1		
Jejunum	26	6		
Ileum	28	23		
Shape			0.002	0.018
Regular	27	4		
Irregular	30	26		
Margin			0.001	0.011
Clear	32	6		
Unclear	25	24		
Dilated lumen			0.002	0.018
Negative	49	17		
Positive	8	13		
Intussusception			0.017	0.036
Negative	56	25		
Positive	1	5		
Enhancement pattern			0.053	0.043
Homogeneous	20	17		
Heterogeneous	37	13		
Enhancement level			<0.001	0.004
Mild	3	12		
Moderate	15	11		
High	39	7		
Adjacent peritoneum			0.002	0.018
Clear	42	12		
Unclear	15	18		
Locoregional lymph node			<0.001	0.004
Nonenlarged	46	7		
Enlarged	11	23		

FDR: false discovery rate.

Mann–Whitney U test was performed, exploring whether the features were significantly different between two groups. Then, univariate logistic regression was applied to select related features (with $p < 0.05$). Finally, we used the minimum redundancy maximum relevance (mRMR) algorithm to select the feature subsets; 8 features with high correlation and low redundancy were retained.

2.5. Model Construction and Evaluation. Two texture models based on the arterial phase and venous phase were constructed, by using the multivariable logistic regression to filter the independent features and construct the multivariable model, through backward stepwise selection with the likelihood ratio test to select the most predictive feature subset. The radiomics score of each patient was calculated.

Meanwhile, the clinical and radiologic features with $p < 0.1$ in univariate logistic regression were selected to develop a clinical model using the multivariate logistic regression analysis.

The 100-fold leave group out crossvalidation (LGOCV) was performed to verify that the multivariate models were valuable in discriminating one group from anther group and the result was not due to overfitting. The data analysis workflow is shown in Figure 2.

In order to assess the reliability of CT texture analysis, the support vector machine (SVM) also has been applied further for comparative analysis.

2.6. Statistical Analyses. The differences of continuous variables were analyzed by the Mann–Whitney U test, and the chi-square test or Fisher's exact test was used for categorical variables. The diagnostic performance of multivariate models was evaluated using ROC analysis and area under the ROC curve (AUC). Diagnostic sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) were also calculated. The corresponding definition/equation was presented in Supplementary materials Part 2. All these statistical analyses were performed with R statistical software 3.5.1 (R Foundation for Statistical Computing). A two-tailed p value of less than 0.05 was considered statistically significant. To control false-positive rates in multiple testing, the false discovery rate-adjusted p values were also calculated during multiple univariate analysis [20].

3. Results

3.1. Patient Characteristics. A total of 87 lesions from 87 patients were enrolled in the retrospective study for analysis: 48 gastrointestinal stromal tumors (GIST) (2 duodenal, 20 jejunal, and 26 ileal), 30 lymphomas (1 duodenal, 6 jejunal, and 23 ileal), and 9 adenocarcinomas (1 duodenal, 6 jejunal, and 2 ileal). Of these patients, there were 47 males and 40 females, and the mean age was 58 years (age range, 4–80 years).

3.2. Clinical and Radiological Feature Evaluation. Our data showed that controlling the false discovery rate in multiple chi-square tests found as many significant results as without false discovery rate adjustment (Table 2). For the clinical data, no significant difference was found between the nonlymphoma group and lymphoma group with regard to the gender ($p = 0.206$) and intestinal obstruction ($p = 0.177$). There was a significant difference between the nonlymphoma group and lymphoma group in the age ($p = 0.008$), melena ($p = 0.001$), and abdominal pain ($p = 0.008$). For radiological features, the location, shape, margin, dilated lumen, intussusception, enhancement level, adjacent peritoneum, and locoregional lymph node differed significantly between the nonlymphoma group and lymphoma group ($p \leq 0.001$–0.037). However, there was no significant difference between the nonlymphoma group and lymphoma group in the enhancement pattern ($p = 0.053$).

The univariate logistic regression analysis of clinical data and radiological features is shown in Supplementary Table 1. Multivariate logistic regression analysis revealed that the margin, locoregional lymph node, enhancement level, and enhancement pattern were independent indicators to distinguish the nonlymphoma from the lymphoma of the small bowel (Table 3). The ill-defined margin, homogeneous enhancement, mild or moderate enhancement, and enlarged lymph node were apt to be a PSBL rather than other types of PSBM. The sensitivity, specificity, accuracy, and AUC of the clinical model were 93.3%, 79.0%, 83.9%, and 0.93 (95% CI 0.87–0.98), respectively (Table 4).

3.3. Texture Feature Evaluation. After dimension reduction, 8 texture features of the arterial phase, 8 texture features of the venous phase, and 8 texture features of the dual phases were selected. There were significant differences of all those texture parameters between small bowel lymphoma and nonlymphoma (Supplementary Tables 2 and 3). The diagnostic performance of the selected texture features is presented in Supplementary Tables 4 and 5.

By using multivariate logistic regression analysis, the arterial texture model was generated using 6 selected features from arterial CT images (Supplementary Table 6). The sensitivity, specificity, accuracy, and AUC were 83.3%, 89.3%, 87.2%, and 0.92 (95% CI 0.87–0.98), respectively (Table 4). Similarly, the venous texture model was developed using 6 selected features (Supplementary Table 7). The sensitivity, specificity, accuracy, and AUC were 73.3%, 85.7%, 81.4%, and 0.87 (95% CI 0.79–0.94), respectively (Table 4). The corresponding RAD scores are shown in Supplementary materials Part 3.

As seen in Table 5, the crossvalidation trials showed that the average accuracy, sensitivity, and specificity values of the three multivariate models were all relatively stable in the training and validation sets. The predictive performance of the arterial texture model was slightly better than that of the venous model and clinical model. The appeared times of selected features during the 100-time crossvalidations are shown in Figure 3. Most texture features and all radiological features appeared more than 50 times during the 100-time trials, which means that these features had high stability and diagnostic values [21].

The diagnostic value of the SVM classification algorithm was presented in Supplementary Table 8. The predictive performance of SVM classifications was slightly better than that of the logistic regression model.

4. Discussion

The detection and differential diagnosis of small bowel neoplasms has been attracting the attention of researchers. Our

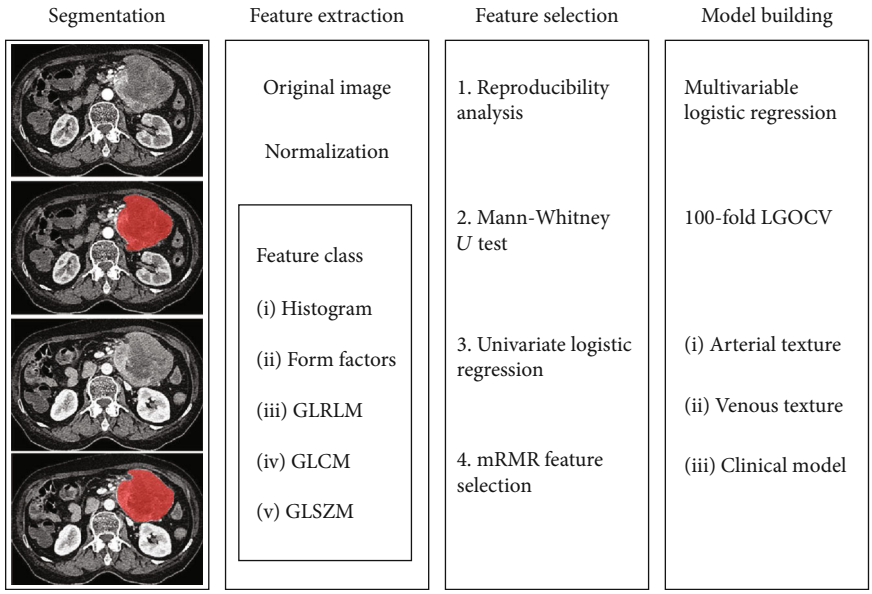


Figure 2: Flowchart of texture analysis. Main steps are tumor segmentation, feature extraction and selection, model construction, and validation. GLRLM: gray level run-length matrix; GLCM: gray level cooccurrence matrix; GLSZM: grey level size zone matrix; mRMR: minimum redundancy maximum relevance; LGOCV: leave group out crossvalidation.

Table 3: The multivariate logistic regression analysis of the clinical data and radiological features.

	Log OR	SE	OR	p value
Margin	3.265	1.156	26.179	0.005
Locoregional lymph node	2.984	0.825	19.766	<0.001
Enhancement level	−2.148	0.604	0.117	<0.001
Enhancement pattern	−2.17	0.906	0.114	0.017

Log OR: Logarithm of odds ratio; SE: standard error; OR: odds ratio.

study developed and validated two CT-based texture models and one radiological model as a novel approach to preoperatively differentiate the PSBL from other types of PSBM, which has never been reported previously.

Most previous studies mainly focused on evaluating the morphological findings and enhancement pattern of small bowel neoplasms for differential diagnosis [5, 6]. Our data showed that the clinical features (age, melena, and abdominal pain) and radiological features (location, shape, margin, dilated lumen, intussusception, enhancement level, adjacent peritoneum, and locoregional lymph node) differed significantly between the nonlymphoma group and lymphoma group. As reported previously, lymphoma of the small bowel tended to be more homogenous and show less contrast enhancement compared with other small bowel malignancies [22]. Our findings were relatively consistent with conventional consensus.

In the present study, 2 multivariate texture models, based on arterial and venous phases, were built to aid in preoperatively discriminating PSBL from other types of PSBM. The arterial logistic model showed a better diagnostic value than the venous logistic model; however, the predictive performance of the venous SVM classifier performed slightly better than that of the arterial SVM classifier. According to previous studies, the arterial phase can reflect the blood supply and functional capillary density and the venous phase may reflect more dysfunctional neovessels and represent distribution of contrast media in interstitial spaces [23]. Our findings indicated that the arterial and venous texture features both played a role in reflecting the heterogeneity of small intestine neoplasms.

Our data also suggested that the clinical model has a similar diagnostic ability compared to the arterial texture model. The clinical model is a multivariate logistic regression model enrolling 4 radiological features, including the margin, locoregional lymph node, enhancement level, and enhancement pattern. The results indicated that the ill-defined margin, enlarged lymph node, mild or moderate enhancement, and homogeneous enhancement derived from subjective CT interpretation were independent indicators of a diagnosis of small bowel lymphoma. Attentively, the definition of the enhancement level and enhancement pattern was based on the CT attenuation of small bowel neoplasms in arterial CT images.

Therefore, our findings indicated that the routine radiological features and texture analysis based on arterial CT imaging both held great diagnostic value in distinguishing nonlymphoma and lymphoma of the small bowel. Shinya et al. also reported that the CT attenuation of the arterial phase performed better than that of the venous phase in discriminating small bowel GIST from lymphoma, with an accuracy of 78.6% and 75.0%, respectively [6]. The potential explanation is that the blood supply distribution of small bowel lymphoma might be less robust and more homogeneous than that of GIST and adenocarcinoma of the small bowel.

The results of 100-fold crossvalidation showed that 5 of the selected 6 features in the arterial model appeared more than 50 times during the 100 times crossvalidations and all

Table 4: The diagnostic performance of the clinical model and two texture models.

	Arterial texture	Venous texture	Clinical model
AUC (95% CI)	0.92 (0.87-0.98)	0.87 (0.79-0.94)	0.93 (0.86-0.98)
Accuracy	0.872	0.814	0.839
Sensitivity	0.833	0.733	0.933
Specificity	0.893	0.857	0.790
PPV	0.804	0.730	0.700
NPV	0.910	0.859	0.957

PPV: positive predictive value; NPV: negative predictive value.

Table 5: The crossvalidation of three multivariate models.

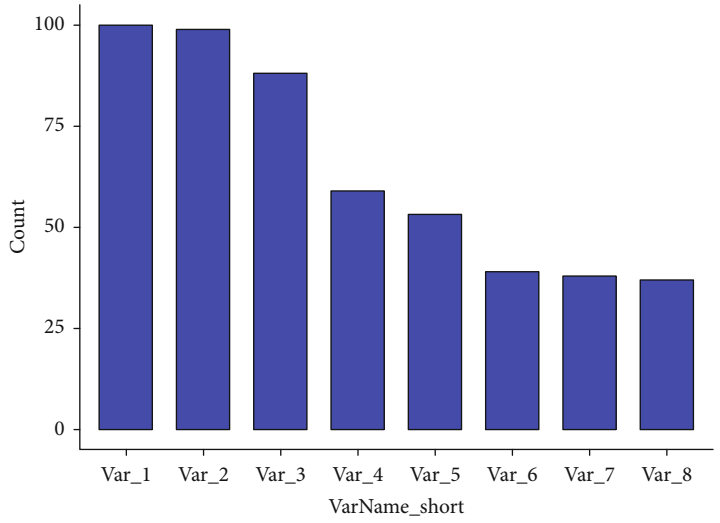
	Arterial texture		Venous texture		Clinical model	
	Training	Test	Training	Test	Training	Test
Accuracy	0.900	0.831	0.838	0.762	0.854	0.828
Sensitivity	0.877	0.829	0.840	0.780	0.818	0.801
Specificity	0.941	0.833	0.832	0.727	0.920	0.883

6 texture features of the venous model appeared more than 50 times, indicating that the two models were reliable and not overfitting. The arterial texture model included 1 GLRLM feature and 5 GLCM features. GLRLM_RunLengthNonuniformity, GLCM_ClusterShade, and GLCM_Correlation had great value and stability in identifying small bowel lymphoma. RunLengthNonuniformity describes the similarity of run lengths throughout the image, with a higher value indicating more heterogeneity among run lengths in the image. Cluster shade is a measure of the skewness and uniformity of the GLCM, with a higher value implying greater asymmetry about the mean. Correlation is a value between 0 (uncorrelated) and 1 (perfectly correlated) showing the linear dependency of gray level values to their respective pixels in the GLCM. In the venous texture analysis, 1 histogram feature, 2 GLRLM features, 2 GLCM features, and 1 GLSZM feature formed the multivariable model. GLRLM_RunLengthNonuniformity and GLCM_Correlation also had high stability and diagnostic values in the venous texture model. Besides, GLRLM_HighGreyLevelRunEmphasis and GLSZM_GreyLevelNonuniformity held a certain value in the venous texture model. HighGreyLevelRunEmphasis reflects the distribution of the higher gray level values, with a higher value indicating a greater concentration of high gray level values in the image. GreyLevelNonuniformity evaluates the variability of gray level intensity values in the image, with a lower value indicating more homogeneity in intensity values. Our data suggested that there was significant difference in the heterogeneity of small bowel lymphoma and nonlymphoma and those texture parameters could help to discriminate primary lymphoma from other small bowel malignancies.

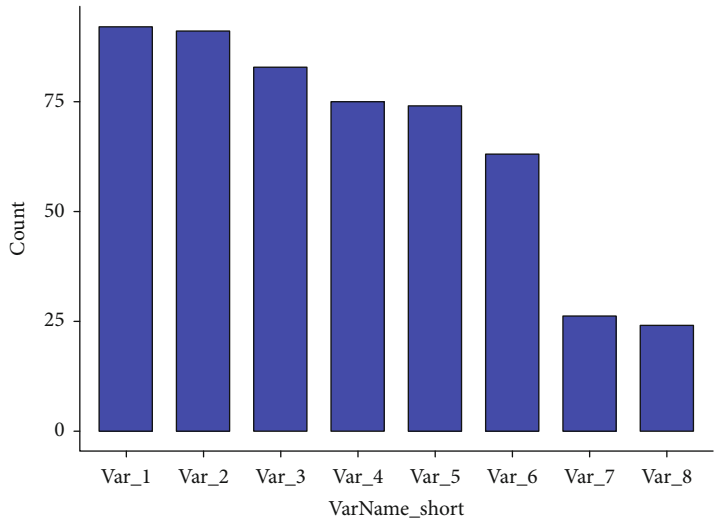
The crossvalidation trials showed that the diagnostic efficiency of three multivariate models were all relatively valuable and stable in the training and validation sets. However, the radiological features are a type of subjective findings with individual experience, leading potential diagnostic variation in the imaging analysis. CT texture analysis is an objective tool, and the overall predictive value of our arterial texture model was slightly better than that of the clinical model during crossvalidation trials. The predictive performance of SVM classifications was slightly better than that of the logistic regression model. Hence, CT texture analysis might provide more objective and valuable information in preoperatively discriminating lymphoma from nonlymphoma of the small bowel.

CT texture analysis has been widely investigated in risk grade prediction and prognosis assessment of GIST [15, 16, 24]. The role of CT texture analysis in discriminating lymphoma from nonlymphoma of the small bowel has never been reported previously, which might be due to the rare incidence of PSBL. Several studies have proved the potential promise of CT texture or radiomics analysis in identifying gastric lymphoma. Ma et al. reported that venous CT radiomics analysis had a potential to accurately differentiate Borrmann type IV gastric cancer from primary gastric lymphoma [25]. In another study, Ba-Ssalamah et al. found that CT texture features proved to be highly successful in distinguishing between gastric adenocarcinoma and lymphoma and GIST and lymphoma, with low misclassification [26]. CT texture analysis over small bowel lymphoma still deserves further investigation.

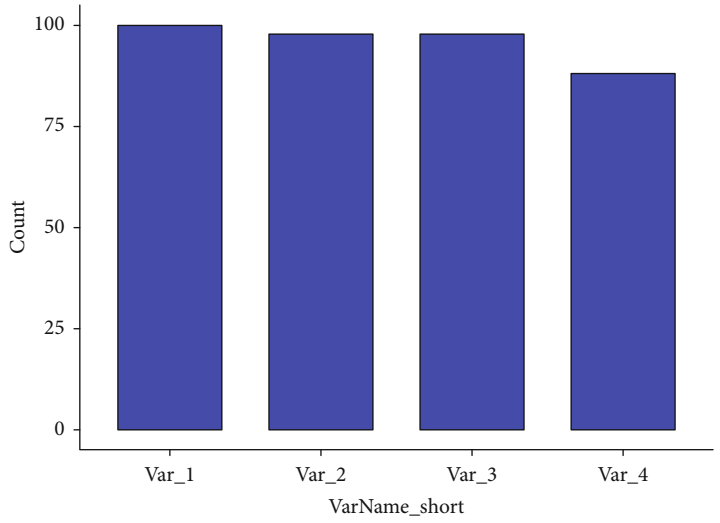
Our study had several limitations. First, it was a retrospective single-center study and the sample size was relatively small and imbalanced. We will further explore the effect of imbalanced data using oversampling or sample weighting method such as SMOTE on the texture analysis model in our future work [27]. Second, bias in patient selection was unavoidable in this retrospective study and no external validation was set owing to small sample size. Hence, the texture model deserves prospective and external validation to confirm its practicability. Third, texture analysis based on delayed-phase CT images was not performed in this study due to no available reconstructed images. The diagnostic efficiency of delayed texture features in small bowel malignant



(a)



(b)



(c)

Figure 3: The appeared times of selected features for 100-fold leave group out crossvalidation (LGOCV) in the arterial texture model (a), venous texture model (b), and clinical model (c). The concrete details are shown in Supplementary materials Part 4.

tumors requires further investigation. Finally, the number of patients with small bowel adenocarcinoma was relatively small. Therefore, those patients were grouped together with patients with GIST, as the nonlymphoma group. Further studies with a more precise classification and lager sample size will be needed.

In conclusion, the arterial texture model showed a great diagnostic value and fitted performance in differentiating PSBL from other types of PSBM, which could provide objective information to screen those patients with suspicion of PSBL.

Abbreviations

CT: Computed tomography
PSBL: Primary small bowel lymphoma
PSBM: Primary small bowel malignancies
GIST: Gastrointestinal stromal tumor
ROI: Regions of interest
ROC: Receiver operating characteristic
AUC: Area under the curve
mRMR: Minimum redundancy maximum relevance
LGOCV: Leave group out crossvalidation
PPV: Positive predictive value
NPV: Negative predictive value.

Data Availability

The data underlying the findings of our study are publicly available wherever possible. Please send an email to Zaixian Zhang (email: befate@126.com) if required.

Additional Points

Key points. Texture features extracted from arterial CT images outperformed those from venous CT images in identifying small bowel lymphoma. The clinical model and arterial texture models held a similar value in discriminating lymphoma from nonlymphoma of the small bowel. The texture models and clinical model showed relatively stable performance during crossvalidation trials.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

The inter- and intrareader correlation coefficients (ICCs) are shown in Supplementary materials Part 1. The definition/equation of accuracy, sensitivity, specificity, positive predictive value, and negative predictive value was presented in Supplementary materials Part 2. The arterial and venous texture rad-scores are shown in Supplementary materials Part 3. The appeared times of selected features for 100-fold leave-group-out crossvalidation (LGOCV) in the arterial texture model, venous texture model, and clinical model are shown in Supplementary materials Part 4. The univariate logistic regression analysis of clinical data and radiological features is shown in Supplementary Table 1. The statistical description of the selected arterial and venous texture features is presented in Supplementary Tables 2 and 3. The diagnostic performance of the selected arterial and venous texture features is presented in Supplementary Tables 4 and 5. The multivariate logistic regression analysis of the selected arterial and venous texture features is shown in Supplementary Tables 6 and 7. The diagnostic performance of the SVM classifier and logistic regression model is seen in Supplementary Table 8. *(Supplementary Materials)*

References

[1] D. Y. Cheung and M. G. Choi, "Current advance in small bowel tumors," *Clinical Endoscopy*, vol. 44, no. 1, pp. 13–21, 2011.

[2] E. A. Williams and A. W. Bowman, "Multimodality imaging of small bowel neoplasms," *Abdominal Radiology*, vol. 44, no. 6, pp. 2089–2103, 2019.

[3] K. Y. Bilimoria, D. J. Bentrem, J. D. Wayne, C. Y. Ko, C. L. Bennett, and M. S. Talamonti, "Small bowel cancer in the United States: changes in epidemiology, treatment, and survival over the last 20 years," *Annals of Surgery*, vol. 249, no. 1, pp. 63–71, 2009.

[4] S. Shenoy, "Primary small-bowel malignancy: update in tumor biology, markers, and management strategies," *Journal of Gastrointestinal Cancer*, vol. 45, no. 4, pp. 421–430, 2014.

[5] M. Boudiaf, A. Jaff, P. Soyer, Y. Bouhnik, L. Hamzi, and R. Rymer, "Small-bowel diseases: prospective evaluation of multi-detector row helical CT enteroclysis in 107 consecutive patients," *Radiology*, vol. 233, no. 2, pp. 338–344, 2004.

[6] T. Shinya, R. Inai, T. Tanaka et al., "Small bowel neoplasms: enhancement patterns and differentiation using post-contrast multiphasic multidetector CT," *Abdominal Radiology*, vol. 42, no. 3, pp. 794–801, 2017.

[7] A. Dohan, M. Boudiaf, X. Dray et al., "Detection of small-bowel tumours with CT enteroclysis using carbon dioxide and virtual enteroscopy: a preliminary study," *European Radiology*, vol. 28, no. 1, pp. 206–213, 2018.

[8] L. M. Minordi, C. Binda, F. Scaldaferri et al., "Primary neoplasms of the small bowel at CT: a pictorial essay for the clinician," *European Review for Medical and Pharmacological Sciences*, vol. 22, no. 3, pp. 598–608, 2018.

[9] C. B. Yang, N. Yu, Y. J. Jian et al., "Spectral CT imaging in the differential diagnosis of small bowel adenocarcinoma from primary small intestinal lymphoma," *Academic Radiology*, vol. 26, no. 7, pp. 878–884, 2019.

[10] M. G. Lubner, A. D. Smith, K. Sandrasegaran, D. V. Sahani, and P. J. Pickhardt, "CT texture analysis: definitions, applications, biologic correlates, and challenges," *Radiographics*, vol. 37, no. 5, pp. 1483–1503, 2017.

[11] F. Ng, B. Ganeshan, R. Kozarski, K. A. Miles, and V. Goh, "Assessment of primary colorectal Cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival," *Radiology*, vol. 266, no. 1, pp. 177–184, 2013.

[12] C. Yip, F. Davnall, R. Kozarski et al., "Assessment of changes in tumor heterogeneity following neoadjuvant chemotherapy in primary esophageal cancer," *Diseases of the Esophagus*, vol. 28, no. 2, pp. 172–179, 2015.

[13] M. B. Andersen, S. W. Harders, B. Ganeshan, J. Thygesen, H. H. Torp Madsen, and F. Rasmussen, "CT texture analysis can help differentiate between malignant and benign lymph

nodes in the mediastinum in patients suspected for lung cancer," *Acta Radiologica*, vol. 57, no. 6, pp. 669–676, 2016.

[14] S. Liu, S. Liu, C. Ji et al., "Application of CT texture analysis in predicting histopathological characteristics of gastric cancers," *European Radiology*, vol. 27, no. 12, pp. 4951–4959, 2017.

[15] J. Yan, X. Zhao, S. Han, T. Wang, and F. Miao, "Evaluation of clinical plus imaging features and multidetector computed tomography texture analysis in preoperative risk grade prediction of small bowel gastrointestinal stromal tumors," *Journal of Computer Assisted Tomography*, vol. 42, no. 5, pp. 714–720, 2018.

[16] C. Feng, F. Lu, Y. Shen et al., "Tumor heterogeneity in gastrointestinal stromal tumors of the small bowel: volumetric CT texture analysis as a potential biomarker for risk stratification," *Cancer Imaging*, vol. 18, no. 1, 2018.

[17] L. K. Mejía, L. Jiang, and V. Gómez, "A case of ampullary follicular lymphoma," *Gastrointestinal Endoscopy*, vol. 84, no. 4, pp. 731-732, 2016.

[18] F. Pilleul, M. Penigaud, L. Milot, J. C. Saurin, J. A. Chayvialle, and P. J. Valette, "Possible small-bowel neoplasms: contrast-enhanced and water-enhanced multidetector CT enteroclysis," *Radiology*, vol. 241, no. 3, pp. 796–801, 2006.

[19] G. Yang, P. Nie, L. Zhao et al., "2D and 3D texture analysis to predict lymphovascular invasion in lung adenocarcinoma," *European Journal of Radiology*, vol. 129, article 109111, 2020.

[20] M. E. Glickman, S. R. Rao, and M. R. Schultz, "False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies," *Journal of Clinical Epidemiology*, vol. 67, no. 8, pp. 850–857, 2014.

[21] W. Wei, X. W. Hu, Q. Cheng, Y. M. Zhao, and Y. Q. Ge, "Identification of common and severe COVID-19: the value of CT texture analysis and correlation with clinical characteristics," *European Radiology*, vol. 30, no. 12, pp. 6788–6796, 2020.

[22] R. M. Gore, U. K. Mehta, J. W. Berlin, V. Rao, and G. M. Newmark, "Diagnosis and staging of small bowel tumours," *Cancer Imaging*, vol. 6, no. 1, pp. 209–212, 2006.

[23] X. H. Chen, K. Ren, P. Liang, Y. R. Chai, K. S. Chen, and J. B. Gao, "Spectral computed tomography in advanced gastric cancer: can iodine concentration non-invasively assess angiogenesis?," *World Journal of Gastroenterology*, vol. 23, no. 9, pp. 1666–1675, 2017.

[24] K. Ekert, C. Hinterleitner, and M. Horger, "Prognosis assessment in metastatic gastrointestinal stromal tumors treated with tyrosine kinase inhibitors based on CT-texture analysis," *European Journal of Radiology*, vol. 116, pp. 98–105, 2019.

[25] Z. Ma, M. Fang, Y. Huang et al., "CT-based radiomics signature for differentiating Borrmann type IV gastric cancer from primary gastric lymphoma," *European Journal of Radiology*, vol. 91, pp. 142–147, 2017.

[26] A. Ba-Ssalamah, D. Muin, R. Schernthaner et al., "Texture-based classification of different gastric tumors at contrast-enhanced CT," *European Journal of Radiology*, vol. 82, no. 10, pp. e537–e543, 2013.

[27] D. Fehr, H. Veeraraghavan, A. Wibmer et al., "Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images," *Proceedings of the National Academy of Sciences*, vol. 112, no. 46, pp. E6265–E6273, 2015.

Hindawi
BioMed Research International
Volume 2021, Article ID 8865237, 13 pages
https://doi.org/10.1155/2021/8865237

Research Article

^{18}F-FDG-PET/CT Whole-Body Imaging Lung Tumor Diagnostic Model: An Ensemble E-ResNet-NRC with Divided Sample Space

Zhou Tao,[1,2] **Huo Bing-qiang**,[1] **Lu Huiling**,[3] **Shi Hongbin**,[4] **Yang Pengfei**,[5] **and Ding Hongsheng**[1]

[1]*School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China*
[2]*Key Laboratory of Images & Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China*
[3]*School of Science, Ningxia Medical University, Yinchuan 750004, China*
[4]*Department of Urology, The General Hospital of Ningxia Medical University, Yinchuan 750004, China*
[5]*Department of Nuclear Medicine, The General Hospital of Ningxia Medical University, Yinchuan 750004, China*

Correspondence should be addressed to Huo Bing-qiang; 2916656832@qq.com and Lu Huiling; lu_huiling@163.com

Received 13 September 2020; Revised 3 November 2020; Accepted 25 February 2021; Published 1 April 2021

Academic Editor: Andrea Scribante

Under the background of ^{18}F-FDG-PET/CT multimodal whole-body imaging for lung tumor diagnosis, for the problems of network degradation and high dimension features during convolutional neural network (CNN) training, beginning with the perspective of dividing sample space, an E-ResNet-NRC (ensemble ResNet nonnegative representation classifier) model is proposed in this paper. The model includes the following steps: (1) Parameters of a pretrained ResNet model are initialized using transfer learning. (2) Samples are divided into three different sample spaces (CT, PET, and PET/CT) based on the differences in multimodal medical images PET/CT, and ROI of the lesion was extracted. (3) The ResNet neural network was used to extract ROI features and obtain feature vectors. (4) Individual classifier ResNet-NRC was constructed with nonnegative representation NRC at a fully connected layer. (5) Ensemble classifier E-ResNet-NRC was constructed using the "relative majority voting method." Finally, two network models, AlexNet and ResNet-50, and three classification algorithms, nearest neighbor classification algorithm (NNC), softmax, and nonnegative representation classification algorithm (NRC), were combined to compare with the E-ResNet-NRC model in this paper. The experimental results show that the overall classification performance of the Ensemble E-ResNet-NRC model is better than the individual ResNet-NRC, and specificity and sensitivity are more higher; the E-ResNet-NRC has better robustness and generalization ability.

1. Introduction

Lung tumors [1, 2] are one of the malignant tumors with high morbidity and mortality [3]. The data reveals that the incidence of lung tumors is increasing year by year, which is a serious threat to human health. The early clinical features of lung tumors are pulmonary nodules [4]. There are no specific clinical symptoms; hence, it is difficult to be detected and diagnosed in time. Once the disease is diagnosed, the cancer is at an advanced stage. Therefore, early diagnosis and early detection are essential for the treatment of lung cancer. Medical imaging techniques [5, 6] are widely used in the diagnosis of lung tumors, such as ultrasound, X-ray imaging, Computerized Tomography imaging(CT), Magnetic Resonance Imaging(MRI), and positron emission tomography imaging(PET). In particular, the advantages of PET and CT are combined by ^{18}F-FDG-PET/CT [7]. It can realize the same machine fusion of anatomical image CT and functional metabolism image PET and accurately locate physical characteristics of the lesion, such as the location, size, shape, and density of the lesion. Finally, the effect of "1 + 1 > 2" is achieved. Mass medical images not only provide more detailed and accurate diagnostic information but also increase the workload of clinicians. Computer-aided diagnosis system (CAD) for lung tumors is an effective solution [8, 9]. On the one hand, CAD can provide doctors with accurate quantitative analysis

services, so as to make up for the defects of human inertia and insensitivity to gray scales [10, 11]; on the other hand, it can effectively reduce the error rate of doctors' interpretation of medical images, thereby helping doctors to better diagnose diseases and improve the diagnosis rate.

Ensemble learning is a machine learning paradigm. Its essence is to use multiple classifiers to solve the same problem and finally use "majority voting" to determine the final result [12]. In recent years, deep learning has become a machine learning hot topic. It has been successfully applied in the field of medical image processing, especially in the auxiliary classification, recognition, detection, and segmentation of malignant tumors, achieving impressive results that surpass human performance. Ensemble deep learning, which couples deep learning and ensemble learning, can make full use of the advantages of the two methods and can provide a new research direction for computer-aided diagnosis. For example, Wang et al. employed transfer learning with relative majority voting to construct a convolutional neural network (CNN) model for the computer-aided diagnosis of lung tumors [13]. In another work, Xiao et al. [14] ensemble a variety of different machine learning models for the accurate diagnosis of lung cancer; five classifiers, namely, k-nearest neighbor (KNN), support vector machine (SVM), decision trees (DTs), random forest (RF), and gradient boosted decision tree (GBDTs), were ensembled to construct a multimodal ensemble model to predict the incidence of both normal and abnormal cancer. Harangi [15] uses an integrated method to integrate four types of deep neural networks, including AlexNet, GoogleNet, VGG and ResNet. Yu and Wang [16] integrated the three deep learning network models of AlexNet, GoogleNet, and VGG for computer-aided diagnosis of lung cancer; there are good generalization ability of ensemble network model. Alzubi and Bharathikannan [17] use weight optimization and maximum likelihood boosting (MLB) to achieve a better false-positive rate and accuracy. Sirazitdinov et al. [18], propose an ensemble of two convolutional neural networks, namely RetinaNet and Mask R-CNN for pneumonia detection and localization. The algorithm is validated on a recently released dataset of 26,684 images from the Kaggle Pneumonia Detection Challenge and scored among the top 3% of submitted solutions.

R-Ensembler, a parameter free greedy ensemble attribute selection method is proposed by Bania and Halder [19] adopting the concept of rough set theory by using the attribute-class, attribute-significance and attribute-attribute relevance measures to select a subset of attributes which are most relevant, significant and non-redundant from a pool of different attribute subsets in order to predict the presence or absence of different diseases in medical dataset. The main role of the proposed ensembler is to combine multiple subsets of attributes produced by different rough set filters and to produce an optimal subset of attributes for subsequent classification task. Cao et al. [20] propose an ensemble ELM (Extreme Learning Machine) combining with the SRC (En-SRC) algorithm. Rather than using the output vector from single ELM to decide the threshold for data partition, En-SRC incorporates multiple ensemble outputs to improve the reliability and classification accuracy. Jiang et al. [21] propose a contextual attention mechanism and a spatial attention mechanism for learning fine-grained representation of pulmonary nodules. an ensemble of 3D Dual Path Networks (DPNs) is used to boost the pulmonary nodule classification performance, Experimental results demonstrate the effectiveness of the proposed method.

Improving the generalizability of individual classifiers and increasing the heterogeneity of individual classifiers in ensemble learning architectures are two crucial factors that can improve the performance of ensemble learning models. Therefore, from the perspective of splitting the sample space in the framework of ensemble learning, and based on the ResNet with nonnegative representation classification (NRC), a ^{18}F-FDG-PET/CT whole-body imaging lung tumor diagnosis E-ResNet-NRC model is proposed. Firstly, three modalities of PET, CT, and PET/CT medical images of lung tumors are collected; according to the medical image modality, the medical image is divided into three sample spaces: PET, CT, and PET/CT. Secondly, constructing an individual classifier based on residual neural network in a different sample space, each individual classifier is trained by migration learning, which can ensure the rapid learning ability of the individual classifier and the difference of the individual classifier. Thirdly, using nonnegative representation classification NRC in the fully connected layer improves the sparse representation ability and classification performance of sample data. Finally, a relatively majority vote is used for ensemble learning, and the results of computer-aided diagnosis of lung tumor images are obtained.

2. Background

2.1. ^{18}F-FDG-PET/CT Whole-Body Imaging. Molecular imaging is a science that uses imaging techniques to reveal the distinct levels of tissue organization at the cellular and subcellular levels, reflecting variations in vivo at the molecular level and allowing to conduct qualitative and quantitative research on biological behaviors based on images. ^{18}F-FDG-PET/CT is an important assessment modality in molecular imaging. It can detect the initial state of the disease in the body before the disease shows clinical symptoms or changes in anatomical structure. Therefore, early intervention of the disease can be realized, and the purpose of reversing, preventing or delaying the occurrence of the disease can be achieved, and the efficiency of the disease cure can be greatly improved (Figure 1).

2.2. ResNet. The ResNet(ResiDual Neural network) is composed of convolutional layers for feature extraction and pooling layers for feature processing. After multiple convolution and pooling operations, the input image is classified and output through a fully connected layer [22]. The ResNet uses shortcut connections and fitting residual representations. The identity mapping reconstructs the learning process, redirects the network information flow, and increases the depth of the network. This improves the representation capability

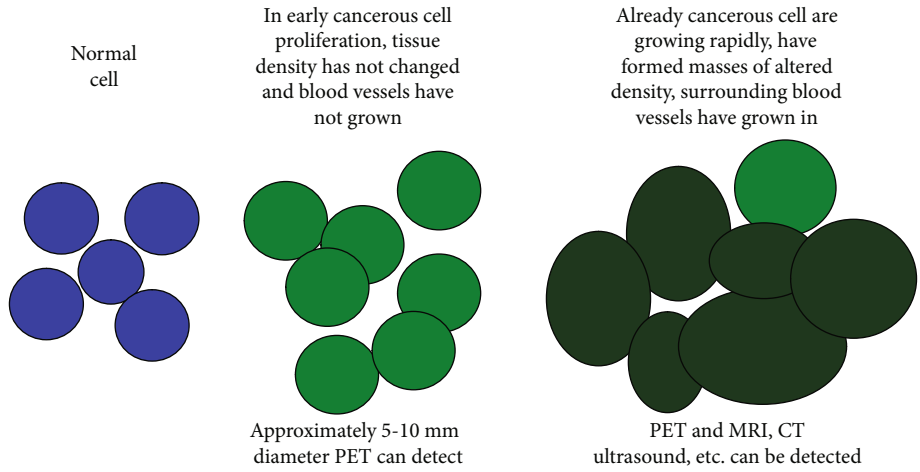


Figure 1: Molecular imaging of tumor cells.

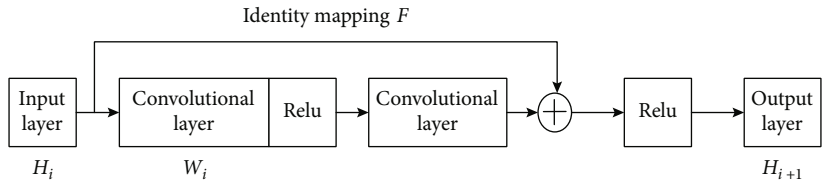


Figure 2: Residual block.

of the model, accelerates the network convergence speed, and effectively solves common issues such as network degradation and gradient vanishing. The residual neural network is composed of multiple residual block structures overlapping, while adjacent convolutional layers are connected by shortcuts to form residual blocks. The structure of the residual block is shown in Figure 2.

H_i represents the input, H_{i+1} represents the output, W_i represents the weights, and F represents residual mapping. The residual block mapping is thus represented as follows:

$$H_{i+1} = \text{Relu}(H_i + F(H_i W_i)). \tag{1}$$

When the input dimension H_i and the output dimension H_{i+1} are different, the linear projection φ is used to match the dimensions. Therefore, Equation (1) can be expressed as

$$H_{i+1} = \text{Relu}\left(\varphi(H)_i + F(H_i W_i)\right). \tag{2}$$

The residual mapping is more easily learned empirically through experimentation when compared with the original mapping. Therefore, the ResNet learns the residual mapping through the middle stacked layers. The residual mapping F is more sensitive to variations in the output, and the parameter adjustment range is comparably broader, thus speeding up learning and improving the network optimization performance. Therefore, the ResNet-50 network was chosen in this study.

2.3. NRC Algorithm. In recent years, sparse representation [23, 24] of high-dimensional feature data has become a research hot topic in the field of machine learning. Sparse representation classification (SRC) [25, 26] for high-dimensional data recognition proves advantageous in improving sparse representation and classification performance. The main concept of SRC is the association of the test sample with a linear combination of the training samples; then, the test samples are divided into their corresponding classes with the minimum distance or approximation error [27]. However, the encoding coefficient of SRC is negative, which, in practice, causes the weights corresponding to the positive and negative coefficients to offset. This affects the classification accuracy to some extent. The classification criterion of nonnegative representation classification (NRC) [28] is the classification according to the similarity of training and test samples. This approach is similar to sparse representation classification (SRC) with the difference being that the coding coefficient of NRC is limited to nonnegative [29]. The nonnegative representation can improve the representation of isomorphic samples while inhibiting the representation of heterogeneous samples, resulting in sparse encoding coefficients from the same correct class; therefore, the nonnegative representation is at the same time sparse and distinguished. Therefore, nonnegative representation tends to find homogeneous samples, which translates to higher recognition accuracy [30].

The main idea of NRC revolves around the query samples $y \in \boldsymbol{R}^{\boldsymbol{D}}$, and the training sample matrix $X = [X_1, \cdots, X_k] \in \mathrm{R}^{\mathrm{D}\times\mathrm{N}}$. Firstly, each column of Y and X is normalized to a unit L_2 standard; the encoding vector $\hat{c}$ is then calculated by querying the samples Y and X. The larger the difference between

Table 1: NRC algorithm.

Algorithm: NRC	
1	Input: training sample matrix $X = [X_1, \cdots, X_k]$ and query sample y
2	Normalize each column of matrix X and query sample y to the unit L_2 norm
3	The encoding vector of y on X is solved by the NRC model
4	Calculate the coefficient matrix: $\hat{c} = \arg \min_c \|y - Xc\|_2^2 s.t. c \geq 0$
5	Calculate residual similarity: $r_k = \|y - X_k \hat{c}_k\|_2$
6	Output label category: $Label(y) = \arg \min \{r_k\}$

reconstruction residuals, as calculated from the matric coefficients, the higher the similarity of the test sample to the training sample. The output label category is assigned based on the degree of residual similarity. The algorithm design is shown in Table 1.

2.4. Ensemble Learning. The core concept of ensemble learning is to train multiple homogeneous and different individual learning algorithms to solve the same problem [31]. Then, the final predicted result is obtained by combining the weighed outputs of all individual learners through a variety of strategies. In order to design a robust ensemble classification model, it is necessary to improve the generalization ability of individual classifiers as well as to increase the differences between the individual classifiers in the ensemble.

Ensemble learning [31] can significantly improve the generalization ability of the learning system. The most common techniques include bagging, boosting, and stacking. The conventional methods used to generate base classifiers can be roughly divided into two broad categories: the first one comprising the application of different types of learning algorithms to the same data set, with the resulting base classifier referred to as heterogeneous, and the second one consisting on the application of the same learning algorithm to different training sets, producing a homogeneous classifier [32].

The combination of strategies of ensemble learning for classifiers includes the average, voting, and learning methods. Different combinations of methods are chosen depending on the application. For example, for regression estimation, the prediction results of individual learners are usually simply averaged or weighed averaged. Meanwhile, for classification, the results of each individual classifier are usually voted to obtain the final classification result. The voting method is divided into the absolute majority voting and the relative majority voting method. The absolute majority voting method is characterized by more than half of the individual learners delivering the same answer; the output is the final classification result of the ensemble. The relative majority voting method is characterized by the majority of individual learners outputting a certain classification result, this result is the final classification result of the ensemble.

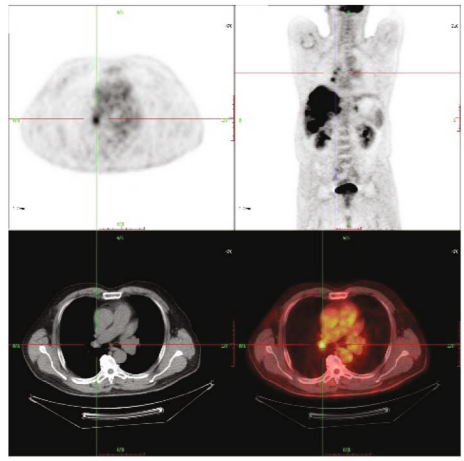


Figure 3: CT, PET, and PET/CT original images.

3. Ensemble E-ResNet-NRC Model with Partitioned Sample Space

3.1. Algorithm Rationale. In this study, an ensemble E-ResNet-NRC model with partitioned sample space is proposed. The overall design of the model is as follows:

3.1.1. Data Collection. 9000 CT, PET, and PET/CT of patient's lung images were collected from a 3A hospital in Ningxia between 2014 and 2016, including 3000 cases of each modal image. Figure 3 shows a PET image of lung tumor (upper left), CT image of a pulmonary tumor (lower left), whole-body image (upper right), and PET/CT image of lung tumor (lower right).

Sample Set Division. Lung medical image sample set is as follows: Sample_Lung, Sample Size | Sample_Lung | = 9000, according to the types of medical image (CT, PET, or PET/CT). Sample_Lung = {Sample_CT, Sample_PET, Sample_PET/CT}. The sample lung was divided into three sample subsets: Sample_CT, Sample_PET, and Sample_PET/CT. with sample sizes |Sample_CT | = 3000, |Sample_PET | = 3000, and |Sample_PET/CT | = 3000. The negative and positive samples of each sample subset are the same, i.e., Sample_CT = {Sample_CT_Negative Sample_CT_Positive}, |Sample_CT_Negative | = | Sample_CT_Positive | = 1500, Sample_PET = { Sample_PET_Negative Sample_PET_Positive}, |Sample_PET_Negative | = | Sample_PET_Positive | = 1500, Sample_PET/

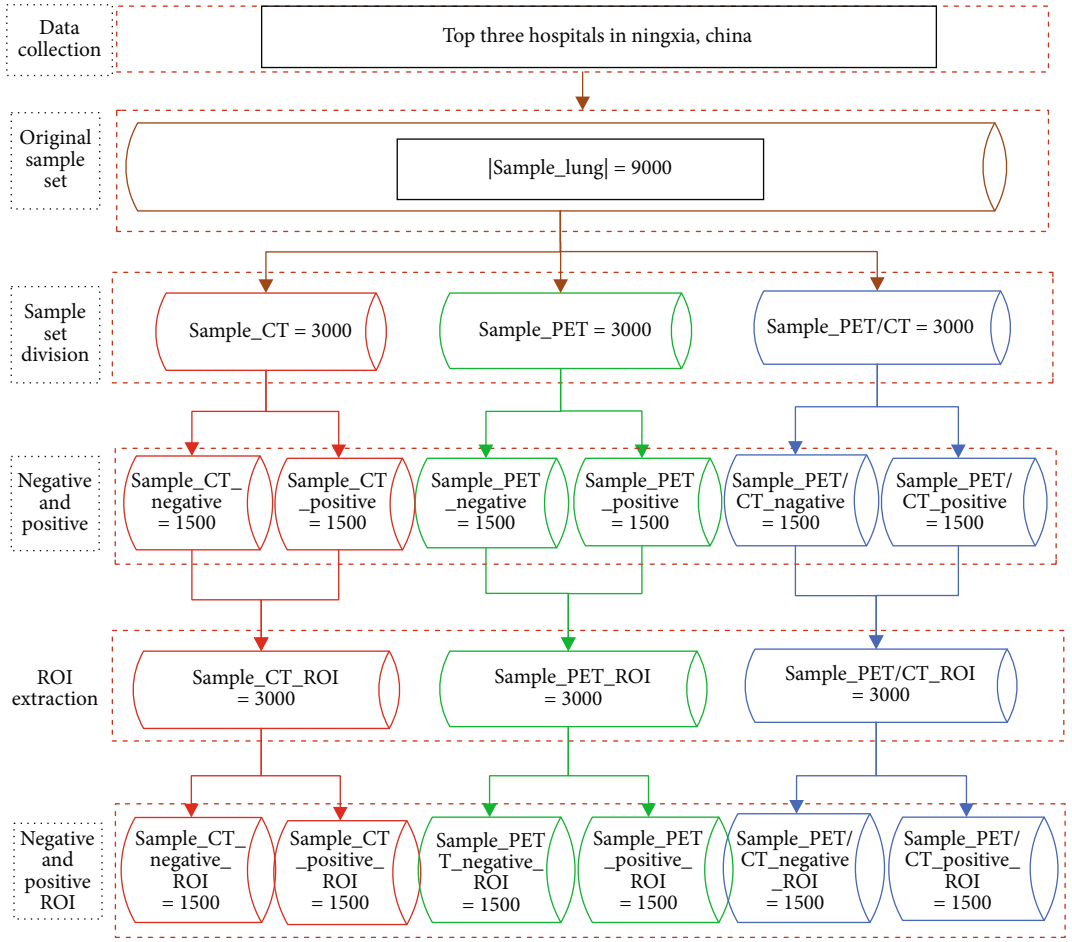


Figure 4: Sample_Lung set division.

CT = {Sample_PET/CT_Negative Sample_PET/CT_Positive}, |Sample_PET/CT_Negative| = |Sample_PET/CT_Positive| = 1500.

3.1.2. Transforming Pseudocolor into Gray Images. Sample_Lung = rgb2gra(Sample_Lung).

3.1.3. ROI. Local features (i.e., the region of interest) are extracted from the global gray images based on clinical markers corresponding to the lesion area. Then, the ROI is normalized to experimental data as 50px × 50px, sample lung = ROI lung (sample lung). The ROI extraction process for each of the three sample subsets is as follows: Sample_CT_ROI = ROI_Lung(Sample_CT), Sample_PET_ROI = ROI_Lung(Sample_PET), Sample_PET/CT_ROI = ROI_Lung(Sample_PET/CT).

3.1.4. Constructing Different Sample Spaces. The lung medical image sample set (Sample_Lung) is composed of three different medical image modalities CT, PET, and PET/CT. The local features of the lesion area are used to define the ROI and obtain the same set as the original Sample_CT_ROI: Sample_CT_ROI = {Sample_CT_ROI, Sample_PET_ROI, Sample_PET/CT_ROI}. In these three sample subsets, each of 3000 cases and each sample subset (negative and positive samples) are the same size, i.e., 1500 cases: Sample_CT_ROI = {Sample_CT_ROI − Negative Sample_CT_ROI_Positive}, |Sample_CT_ROI_Negative| = |Sample_CT_ROI_Positive| = 1500, Sample_PET_ROI = {Sample_PET_ROI_Negative Sample_PET_ROI_Positive}, |Sample_PET_ROI_Negative| = |Sample_PET_ROI_Positive| = 1500, Sample_PET/CT_ROI = {Sample_PET/CT_ROI_Negative Sample_PET/CT_ROI_Positive}, |Sample_PET/CT_ROI_Negative| = |Sample_PET/CT_ROI_Positive| = 1500.

Figure 4 shows that Sample_Lung set is divided into three sample spaces.

3.1.5. Construction of a Fivefold Cross-Experimental Dataset Based on Sample Space Division in the Three Sample Subsets, Namely Sample_CT_ROI, Sample_PET_ROI, and Sample_PET_CT_ROI. A dividing algorithm was used to separate the negative and positive sample sets of each sample subset into 5 uniform datasets, each one of 300 samples, to obtain a 5-fold cross-sample set.

3.1.6. Construction of the ResNet-NRC. Individual classifiers were designed based on sample subsets of the three image modalities.

(1) The ResNet-50 was pretrained via transfer learning. The parameters in the pretraining network were

Table 2: ResNet-50 parameters.

Layers	Output size	Parameter
Conv1	112×112	7×7, 64, stride 2
Max-Pool	112×112	3×3, 64, stride 2
Conv2_x	56×56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3_x	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4_x	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Conv5_x	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
Avg_pool	7×7	2048
FC	1×1	1000

taken as the initialization parameters: ResNet − NRC = Transfer learning (ResNet-50, NRC); Table 2 shows ResNet-50 parameters

(2) In three sample subsets Sample_CT_ROI, Sample_PET_ROI, and Sample_PET/CT_ROI, the ResNet-NRC network is retrained to get individual classifiers: ResNet − NRC − CT = Training(ResNet − NRC, Sample_CT_ROI), ResNet − NRC − PET = Training(ResNet − NRC, Sample_PET_ROI), ResNet − NRC − PET/CT = Training(ResNet − NRC, Sample_PET/CT_ROI).

3.1.7. The ResNet-NRC Classifier. The ResNet-NRC classifier was ensembled via relative majority voting to obtain three individual classifiers: ResNet − NRC = Ensemble{(ResNet − NRC − CT, ResNet − NRC − PET, ResNet − NRC − PET}.

Figure 5 shows an algorithm flow chart.

3.2. Key Technology: ResNet-NRC Model. Transfer learning refers to the initialization of a small training set of parameters by using a pretrained network with a proven learning capacity. This method can this be used to transfer existing learning abilities from one network to another. In this paper, three individual classifiers, namely, ResNet-NRC-CT in CT mode, ResNet-NRC-PET in PET mode, and ResNet-NRC-PET/CT in PET/CT mode, were constructed via transfer learning based on the ResNet-50. This model was used to identify lung tumors from CT, PET, and PET/CT medical images, respectively.

Input: The three sample subsets Sample_CT_ROI, Sample_PET_ROI, and Sample_PET/CT_ROI.

Output: Three ResNet-NRC Individual classifiers, ResNet-NRC-CT, ResNet-NRC-PET, and ResNet-NRC-PET/CT.

The process to obtain these is as follows.

(1) Transfer learning was used to train the ResNet-50: ResNet = TransferLearning(ResNet − 50)

(2) For the three modalities, the initialization parameters are taken from the pretrained ResNet-50 network. The training and extraction of the fully connected layer features carried in the ResNet: ResNet − CT = Training(ResNet, Sample_i_ROI), $i = 1, 2, 3$, where 1, 2, and 3 refer to CT, PET, and PET/CT, respectively

(3) Taking the Sample-CT-ROI as an example, for the training samples $X = [X_1, \cdots X_k], X_i \in$ Sample-CT-ROI, and testing sample $y = [y_1, \cdots y_n]$, $y_i \in$ Sample-CT-ROI. Through the ResNet-50 feature extraction, the training sample matrix of the feature space is obtained as $X' = [X'_1, \cdots X'_k]$, with a test sample matrix $y' = [y'_1, \cdots y'_n]$.

(4) Each column of the matrix X' and query sample y' are normalized to unit L_2 standard:

$$\begin{aligned} \left\|X'\right\|_2 &= \sqrt{\sum_{i=1}^{n} x'^2_i}, \\ \left\|y'\right\|_2 &= \sqrt{\sum_{i=1}^{n} y'^2_i} \end{aligned} \tag{3}$$

(5) The training sample X' in the feature space y' is nonnegative. Therefore, the nonnegative coefficient $\hat{c}$ can be obtained as

$$\hat{c} = \arg\ \min_c \left\|y' - X'c\right\| \text{s.t.} c \geq 0 \tag{4}$$

(6) Training samples are used to classify the nonnegative representations of the test samples based on their similarity as

$$r_k = \left\|y' - X'_k \hat{c}_k\right\|_2 \tag{5}$$

(7) Finally, the label category of the residual output result is defined as

$$\text{Label}\left(y'\right) = \arg\min \{r_k\}. \tag{6}$$

4. Experiments

4.1. Experimental Environment. Software environment is as follows: Windows10 operating system, MatlabR2019a;

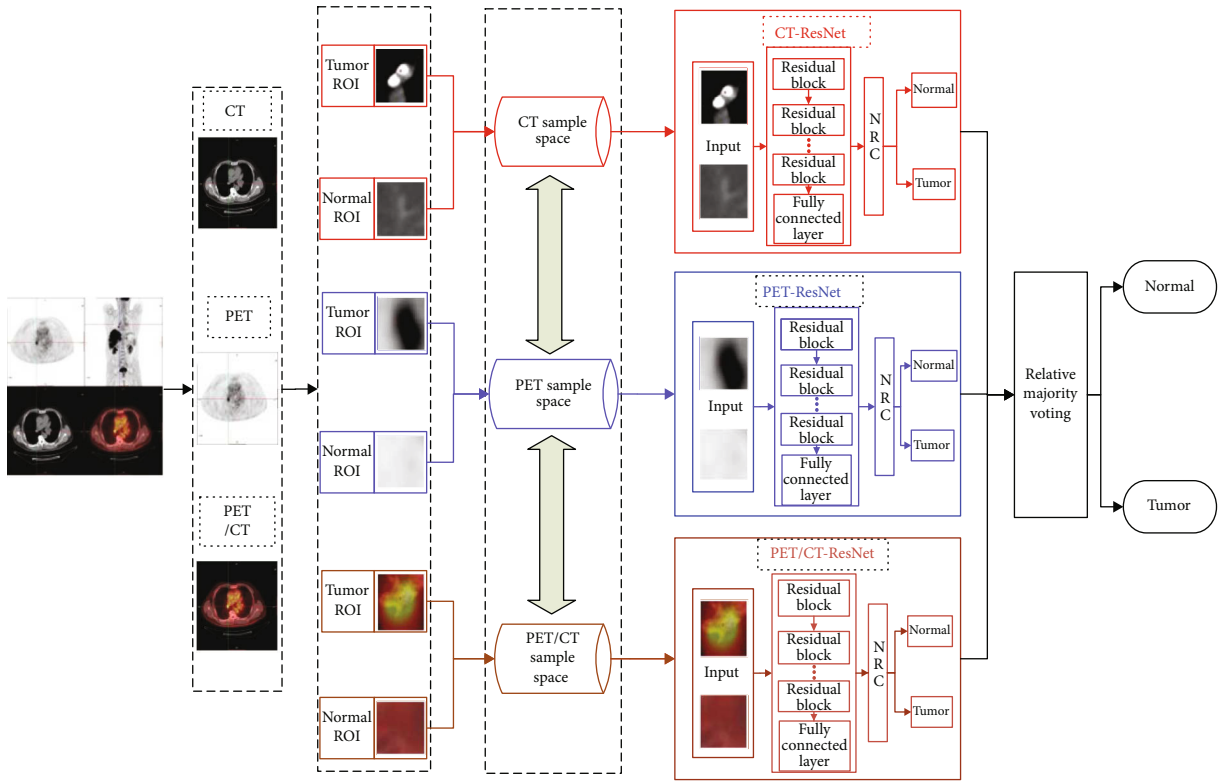


Figure 5: Algorithm flow chart of ensemble E-ResNet-NRC.

hardware environment is as follows: Intel(R)Core(TM)i5-7200U CPU @2.50GHz 2.70GHz, 4.0GB memory, 500GB hard drive.

4.2. Evaluation Metrics. In this paper, the evaluation metrics include accuracy, sensitivity, specificity, F-score value, and Matthews correlation coefficient (MCC), which are described as follows:

Accuracy, sensitivity, and specificity were calculated by true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN). TP indicates a benign tumor was predicted correctly, FP indicates a malignant tumor was predicted incorrectly, TN indicates a malignant image was predicted correctly, and FN indicates that benign tumors were predicted incorrectly. They are calculated by the following formulae:

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{Sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{Specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}}. \end{aligned} \tag{7}$$

The F-value is a summed average of the percentages of completeness and accuracy. It is used as a trade-off between accuracy and recall. The calculation formula is as follows:

$$F = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}. \tag{8}$$

MCC is a more comprehensive evaluation metric that reflects the reliability of the algorithm. When the number of categories is different, the value of the measure is considered balanced ranging from -1 to +1. The MCC takes the value of 1 when the prediction error is 0 for both FP and FN, which means that the classification is completely correct; when the prediction error is 0 for both TP and TN, the MCC takes the value of -1, which means that the classification is completely wrong. It is calculated as follows:

$$\text{MCC} = \frac{\text{TP} \times \text{TN-FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \tag{9}$$

4.3. Experimental Results and Analysis. The experiments were performed using a 5-fold cross-validation for training. The final results were averaged over five experiments. 2400 training samples and 600 test samples were used. The experiments were carried out in CT, PET, and PET/CT trimodal datasets. AlexNet and ResNet-50 were used for comparison. Classification was achieved through the nearest neighbor classification (NNC), softmax, and nonnegative representation classification (NRC) algorithms. The algorithms were pairings were as follows: AlexNet+NNC, AlexNet+Softmax, AlexNet+NRC, ResNet-50+NNC, ResNet-50+Softmax, and ResNet-50+NRC.

4.3.1. Experiment 1: Comparison of the Accuracy and Times of the Different Models. This experiment explored the effects of different network models, classification algorithms, and sample spaces on the ResNet recognition rate and training time. The following six combinations of algorithms were

Table 3: Comparison of accuracy, standard deviation, and training times in different models.

	Evaluation index	AlexNet +NNC	AlexNet +Softmax	AlexNet +NRC	ResNet-50 +NNC	ResNet-50 +Softmax	ResNet-50 +NRC
CT	Acc (%)	96.37	98.20	98.80	97.00	98.13	99.07
	SD (%)	1.29	1.14	0.93	1.51	1.34	0.50
	Training time (s)	161.48	164.33	185.91	1176.91	1182.28	1204.95
PET	Acc (%)	99.57	99.50	99.83	99.50	99.63	99.80
	SD (%)	0.75	1.31	0.25	1.31	1.30	0.48
	Training time (s)	146.63	148.77	169.46	1035.85	1036.29	1059.48
PET/CT	Acc (%)	96.60	97.63	97.97	97.53	97.90	98.33
	SD (%)	5.04	3.60	3.37	3.14	2.63	3.01
	Training time (s)	139.25	138.29	162.38	1162.93	1169.76	1187.22
Ensemble	Acc (%)	99.10	98.87	99.43	99.23	99.33	99.57
	SD (%)	2.03	2.28	0.86	1.67	1.45	1.07
	Training time (s)	422.39	420.09	491.20	3080.76	3234.90	3179.38

Table 4: Accuracies of different network models and classification algorithms.

Network model	Classification algorithm	CT (%)	PET (%)	PET/CT (%)	Ensemble (%)
AlexNet	NNC	96.37	99.57	96.60	99.10
	Softmax	98.20	99.50	97.63	98.87
	NRC	98.80	99.83	97.97	99.43
ResNet-50	NNC	97.00	99.50	97.53	99.23
	Softmax	98.13	99.63	97.90	99.33
	NRC	99.07	99.80	98.33	99.57

Table 5: Comparison of sensitivity results of different network models and classification algorithms.

Network model	Classification algorithm	CT (%)	PET (%)	PET/CT (%)	Ensemble (%)
AlexNet	NNC	99.00	100.00	99.60	99.87
	Softmax	99.20	100.00	98.27	99.27
	NRC	99.07	100.00	98.87	99.53
ResNet-50	NNC	98.80	100.00	99.00	99.93
	Softmax	98.73	100.00	98.80	99.87
	NRC	99.40	100.00	99.33	99.73

examined: AlexNet+NNC, AlexNet+Softmax, AlexNet +NRC, ResNet-50+NNC, ResNet-50+Softmax, and ResNet-50+NRC. The recognition accuracy, running time for training, and standard deviation (SD) in the sample space of CT, PET, and PET/CT are shown in Table 3.

(1) Not Using Ensemble Learning. The experiment was also carried out without using ensemble learning. In the first scenario, different network models with the same classification algorithms were used. As in Experiment 1, three groups of comparative experiments were performed, namely, AlexNet +NNC and ResNet-50+NNC, AlexNet+Softmax, and

Table 6: Comparison of specificity results of different network models and classification algorithms.

Network model	Classification algorithm	CT (%)	PET (%)	PET/CT (%)	Ensemble (%)
AlexNet	NNC	93.73	99.13	93.60	98.33
	Softmax	97.20	99.00	97.00	98.47
	NRC	98.53	99.67	97.07	99.33
ResNet-50	NNC	95.20	99.00	96.07	98.53
	Softmax	97.53	99.27	97.00	98.80
	NRC	98.73	99.60	97.33	99.40

Table 7: Comparison of F-value results of different network models and classification algorithms.

Network model	Classification algorithm	CT (%)	PET (%)	PET/CT (%)	Ensemble (%)
AlexNet	NNC	96.46	99.57	96.70	99.11
	Softmax	98.22	99.50	97.65	98.87
	NRC	98.80	99.83	97.98	99.43
ResNet-50	NNC	97.05	99.50	97.57	99.24
	Softmax	98.14	99.63	97.92	99.34
	NRC	99.07	99.80	98.35	99.57

Table 8: Comparison of MCC results of different network models and classification algorithms.

Network model	Classification algorithm	CT (%)	PET (%)	PET/CT (%)	Ensemble (%)
AlexNet	NNC	92.86	99.14	93.37	98.21
	Softmax	96.42	99.00	95.27	97.74
	NRC	97.66	99.67	95.95	98.87
ResNet-50	NNC	94.06	99.00	95.11	98.48
	Softmax	96.27	99.27	95.82	98.67
	NRC	98.14	99.60	96.69	99.13

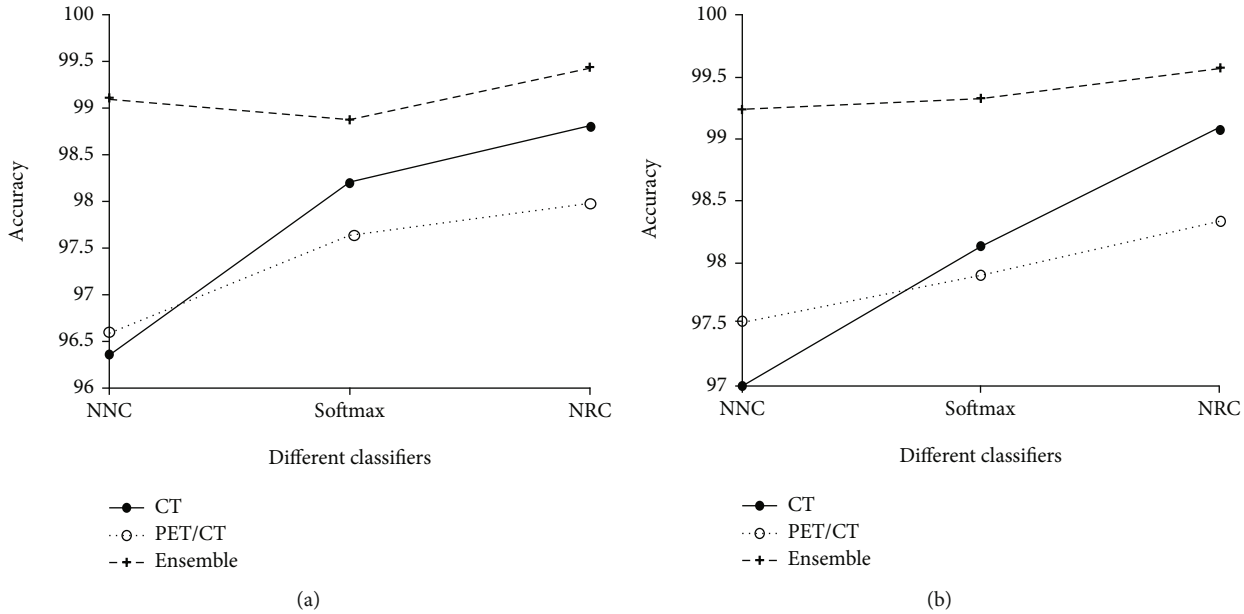


Figure 6: Accuracy of the AlexNet and ResNet-50 models. (a) Accuracy of the AlexNet model. (b) Accuracy of the ResNet-50 model.

ResNet-50+Softmax, as well as AlexNet+NRC and ResNet-50+NRC.

Taking the third group as an example, in the CT sample space, the accuracy of the proposed ResNet-50+NRC model was 0.27% higher than that of the AlexNet+NRC model with a training time of 1019.04 seconds. It is noted that the ResNet is deeper when compared with the AlexNet; for this reason, the extracted image features are richer, and the classification accuracy is higher; however, the training time is greatly increased, in this case, by 648.14%. The results of the other two groups were similar (data not shown).

In the second scenario, the same network with different classification algorithms was used. In Experiment 1, there were three groups of comparative experiments, namely, AlexNet+NNC and AlexNet+Softmax, AlexNet+NRC and ResNet-50+NNC, and ResNet-50+Softmax and ResNet-50 +NRC. Taking the second group as an example, in the CT sample space, the classification accuracy of the individual classifier ResNet-50+NRC was 2.07% and 0.94% higher than that of the ResNet-50+NNC and ResNet-50+Softmax, respectively. In terms of the training times, Net-50+NRC was 28.04 and 22.67 seconds faster than the ResNet-50 +NNC and ResNet-50+Softmax models, respectively. Compared with the first scenario, the overall training time was greatly improved; however, after the network model was determined, the increase in training time was not significant. It is noted that when using the same network architecture, the NRC model exhibits a better classification accuracy when compared with the NNC and Softmax models. This algorithm also proved suitable for handling high-dimensional data and reduced training times significantly.

(2) Using Ensemble Learning. In this experiment, the same network architecture and classification algorithms under different sample spaces were used. Six groups of comparative experiments were considered: AlexNet+NNC and E-AlexNet+NNC, AlexNet+Softmax and E-AlexNet+Softmax, AlexNet+NRC and E-AlexNet+NRC, ResNet-50+NNC and E-ResNet-50+NNC, ResNet-50+Softmax and E-ResNet-50+Softmax, and ResNet-50+NRC and E-ResNet-50+NRC.

Taking the third group in the three different sample spaces as an example, the classification accuracy of E-AlexNet+NRC model was 0.63% and 1.46% higher than that of the AlexNet+NRC in the CT and PET/CT sample spaces, respectively. When taking the sixth group in the three sample spaces as an example, the classification accuracy of the proposed E-ResNet-50+NRC model was 0.50% and 1.24% higher than that of the ResNet-50+NRC model in the sample space of CT and PET/CT, respectively. Meanwhile, the training time was improved by 1974.43 and 1992.16 seconds, respectively. It is noted that when using the same network model and classification algorithm on different sample spaces, ensemble learning can improve the classification accuracy at the expense of substantially increased training times. From the comparative experiments in Experiment 1, namely, E-AlexNet+NNC, E-AlexNet+Softmax, E-AlexNet +NRC, E-ResNet-50+NNC, E-ResNet-50+Softmax, and E-ResNet-50+NRC, the classification accuracy of the proposed E-ResNet-50+NRC model was 99.57%—the highest among the six tested models.

4.3.2. Experiment 2: Comparison of Evaluation Indexes of Different Models. In this experiment, six algorithms were examined: AlexNet+NNC, AlexNet+Softmax, AlexNet +NRC, ResNet-50+NN, ResNet-50+Softmax, and ResNet-50+NRC. Training and recognition were carried out in three sample spaces: CT, PET, and PET/CT. The algorithms were evaluated in terms of their accuracy, sensitivity, specificity, F-value, and MCC (Tables 4–8).

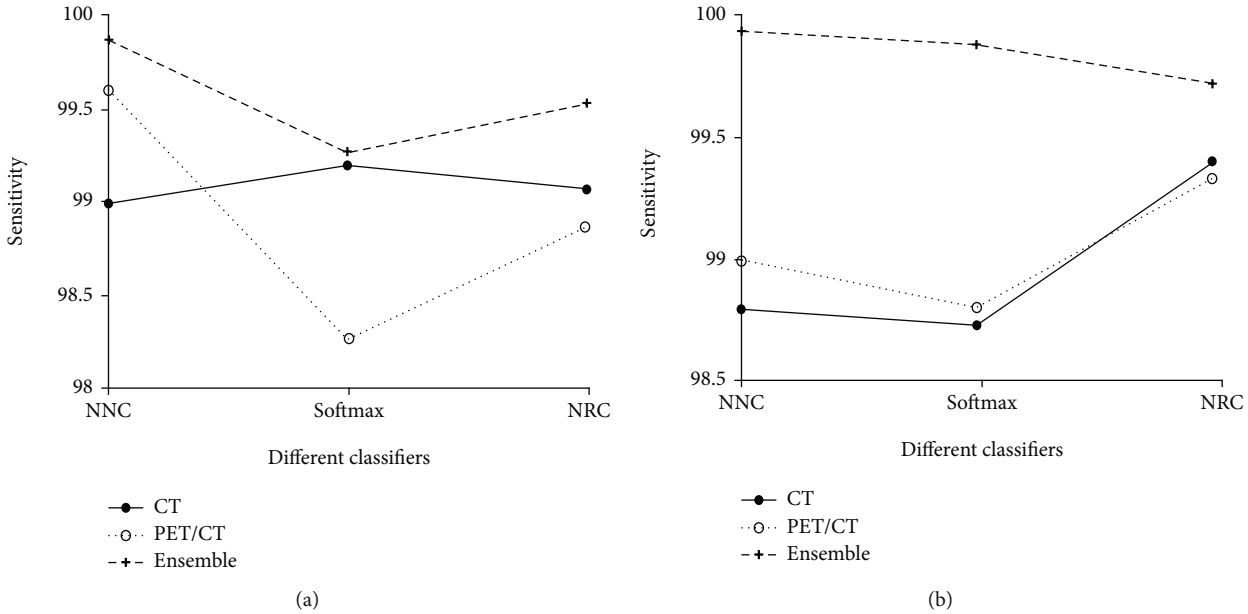


Figure 7: Sensitivity of the AlexNet and ResNet-50 models. (a) Sensitivity of the AlexNet model. (b) Sensitivity of the ResNet-50 model.

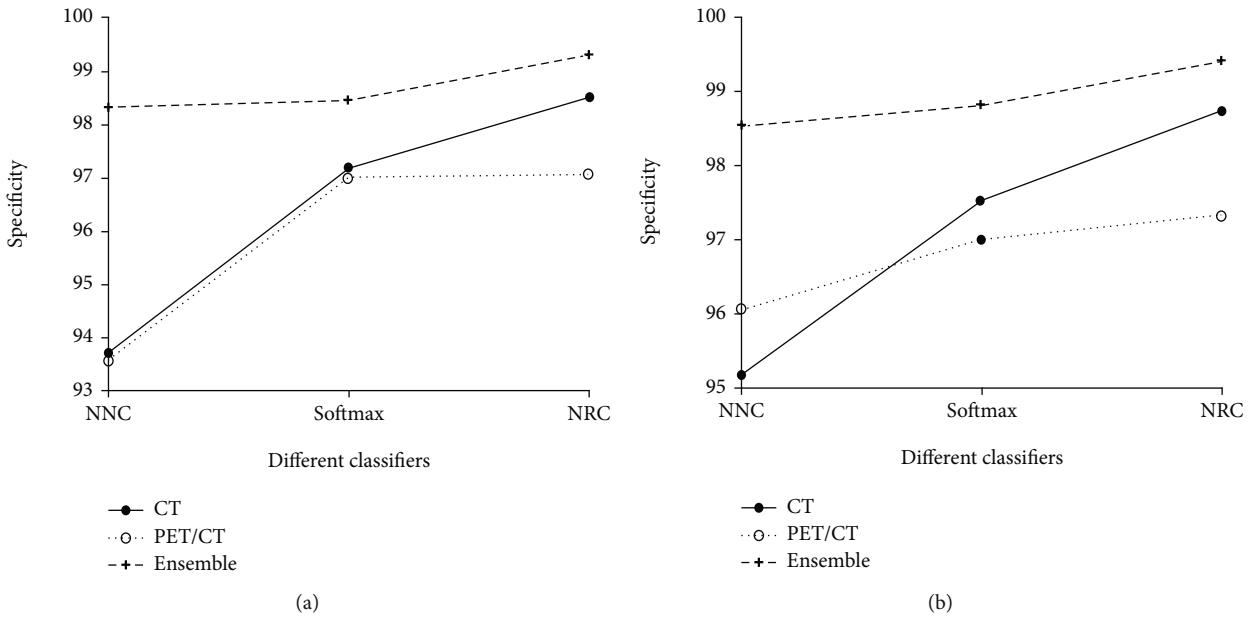


Figure 8: Specificity of the AlexNet and ResNet-50 models. (a) Specificity of the AlexNet model. (b) Specificity of the ResNet-50 model.

From Tables 4–8, it is noted that when using different network architectures with the same classification algorithm, the ResNet50-NRC showed improvements of 0.27%, 0.33%, 0.2%, 0.13%, and 0.48 seconds in accuracy, sensitivity, specificity, F-value, and MCC, respectively, when compared with the AlexNet-NRC and the Text E-ResNet50-NRC. When compared with the AlexNet-NRC, the sensitivity, specificity, F-value, and MCC were increased by 0.14%, 0.2%, 0.07%, 0.14%, and 0.36%, respectively. Plotting the average value of the indicators presented in Figures 6–10 provides with a clear visual representation of the differences between the different algorithms.

From the information derived from the above experiments and analyses, it is noted that, when using the same network architecture, the NRC algorithm exhibited a better performance when compared with the NNC and Softmax algorithms. The NCR algorithm with a ResNet proved more robust for handling high-dimensional data in the CT, PET, and PET/CT sample spaces. In terms of classification accuracy, the experimental results showed that the ResNet-50 architecture was better suited when compared with the AlexNet. The ResNet reconstructs the learning process and redirects the information flow through deep convolutional layers, which solves the issues of network degradation and

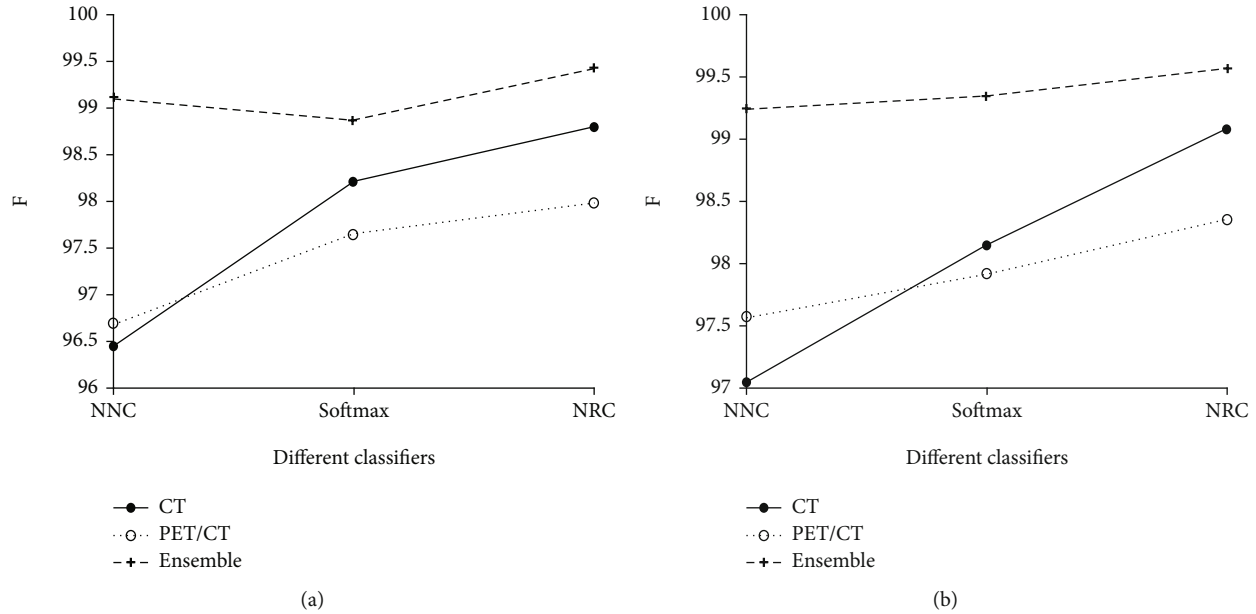


Figure 9: *F*-score of the AlexNet and ResNet-50 models. (a) *F*-score of the AlexNet model. (b) *F*-score of the ResNet-50 model.

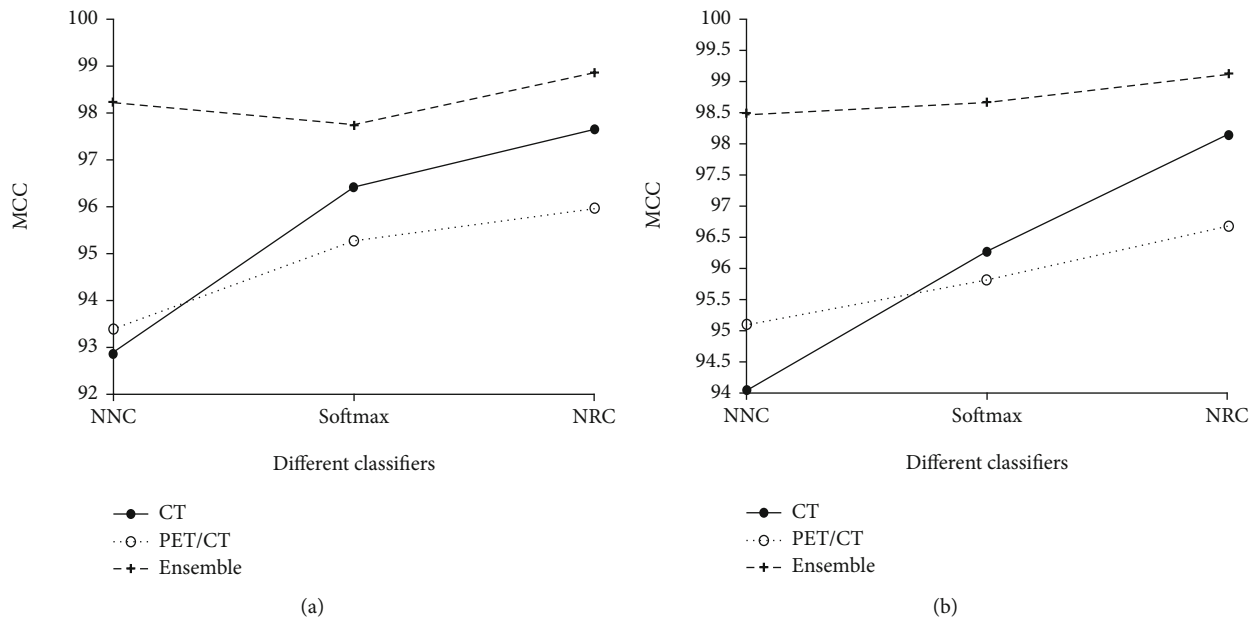


Figure 10: MCC of the AlexNet and ResNet-50 models. (a) MCC of the AlexNet and ResNet-50 models. (b) MCC of the AlexNet and ResNet-50 models.

deepens the architecture without the necessity of additional parameters and computation. The generalizability and convergence of the model are improved. When using the same network architecture and classification algorithm in the three sample spaces, the experimental results showed that the performance of the ensemble model was better suited than that of the individual classifier models. Most notably, the E-ResNet-50+NRC model proposed proved better than the other six architectures tested; this model exhibited a higher accuracy, sensitivity, specificity, *F*-value, and MCC, as well as a robust depth and generalizability. Finally, it is noted that the training times were significantly increased; this can be mitigated by the integration of more powerful hardware such as GPUs or cloud computing platforms. Additionally, in the PET sample space, the classification accuracy of all models was relatively high; this is because the PET silhouette contains less information, mainly highlighted information, which accounts for a large contrast—the explanation is not clear.

5. Conclusion

In this paper, an E-ResNet-NRC model was proposed and implemented by dividing the sample space based on

ensemble learning and using nonnegative representation classification with a ResNet for the classification of medical images of lung tumors. Firstly, the parameters were initialized using transfer learning from a pretrained ResNet. Next, the sample is divided into three different spaces (CT, PET, and PET/CT) according to the different medical imaging techniques. The ResNet extracts the ROI features and uses them to construct feature vectors. Then, an individual classifier, ResNet-NRC, was constructed by employing nonnegative NRC at the fully connected layer. Finally, the ensemble classifier E-ResNet-NRC was achieved by employing relative majority voting. The experimental results showed that the overall classification performance of the proposed E-NRC-ResNet model was better than that of the individual classifier. Its specificity and sensitivity were also higher, while possessing good robustness and generalizability.

Data Availability

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

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This work is supported by the Natural Science Foundation of China (Grant No. 62062003) and North Minzu University (No. 2020KYQD08).

References

[1] S. Xin, X. Shipeng, and L. Haibo, "Lung nodules detection based on convolutional neural network," *China Medical Imaging Technology*, vol. 34, pp. 934–939, 2018.

[2] H. L. Lu, "Computer Aided Diagnosis Research of Lung Tumor Based on Deep Convolutional Neural Network and Global Features," *BioMed Research International*, 2021.

[3] G. Savitha and P. Jidesh, "A holistic deep learning approach for identification and classification of sub-solid lung nodules in computed tomographic scans," *Computers & Electrical Engineering*, vol. 84, article 106626, 2020.

[4] S. Agarwala, M. Kale, D. Kumar et al., "Deep learning for screening of interstitial lung disease patterns in high- resolution CT images," *Clinical Radiology*, vol. 75, no. 6, pp. 481.e1–481.e8, 2020.

[5] A. Bhandary, G. Ananth Prabhu, V. Rajinikanth et al., "Deep-learning framework to detect lung abnormality - a study with chest X-Ray and lung CT scan images," *Pattern Recognition Letters*, vol. 129, pp. 271–278, 2020.

[6] J. Yang, T. Zhou, and L. Guo, "Lung tumor image recognition algorithm based on cuckoo search and deep belief network," *Journal of Computer Applications*, vol. 38, pp. 3226–3230, 2018.

[7] Z. Yuyi and Y. Zhiming, "The progress of ^{18}F-FDG PET/CT in the diagnosis of N-staging of non-small cell lung cancer," *International Journal of Radiation Medicine and Nuclear Medicine*, vol. 40, pp. 447–451, 2016.

[8] S. Diao, J. Hou, H. Yu et al., "Computer-aided pathologic diagnosis of nasopharyngeal carcinoma based on deep learning," *American Journal of Pathology*, vol. 190, no. 8, pp. 1691–1700, 2020.

[9] C. Qiangrui and X. Shipeng, "Lung tumor detection method based on deep learning," *Computer Technology and Development*, vol. 28, no. 4, pp. 201–204, 2018.

[10] W. K. Moon, Y. W. Lee, H. H. Ke, S. H. Lee, C. S. Huang, and R. F. Chang, "Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 190, article 105361, 2020.

[11] V. H. A. Ribeiro and G. Reynoso-Meza, "Ensemble learning by means of a multi-objective optimization design approach for dealing with imbalanced data sets," *Expert Systems with Applications*, vol. 147, article 113232, 2020.

[12] H. Weijian, C. Wei, F. Haozhe, Z. Tianping, Z. Zhengmao, and P. Qiaoming, "A review of deep learning methods applied to lung nodule detection in planar CT images," *Journal of Zhejiang University*, vol. 44, pp. 380–384, 2017.

[13] Y. Wang, T. Zhou, H. Lu, C. Wu, and P. Yang, "Computer aided diagnosis model for lung tumor based on ensemble convolutional neural network," *Journal of Biomedical Engineering*, vol. 34, no. 4, pp. 543–551, 2017.

[14] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 1–9, 2018.

[15] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *Journal of Biomedical Informatics*, vol. 86, pp. 25–32, 2018.

[16] C. Yu and Z. Wang, *Research on medical image diagnosis algorithm based on integrated deep learning*, Shandong University, Shandong, 2019.

[17] J. A. Alzubi and B. Bharathikannan, "Boosted neural network ensemble classification for lung cancer disease diagnosis," *Applied Soft Computing*, vol. 80, pp. 579–591, 2019.

[18] I. Sirazitdinov, M. Kholiavchenko, T. Mustafaev, Y. Yixuan, R. Kuleev, and B. Ibragimov, "Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database," *Computers & Electrical Engineering*, vol. 78, pp. 388–399, 2019.

[19] R. K. Bania and A. Halder, "R-Ensembler: A greedy rough set based ensemble attribute selection algorithm with *k*NN imputation for classification of medical data," *Computer Methods and Programs in Biomedicine*, vol. 184, article 105122, 2020.

[20] J. Cao, J. Hao, X. Lai, C. H. Vong, and M. Luo, "Ensemble extreme learning machine and sparse representation classification," *Journal of the Franklin Institute*, vol. 353, no. 17, pp. 4526–4541, 2016.

[21] H. I. Jiang, F. Gao, X. X. Xu, F. Huang, and S. G. ZhuAttentive and ensemble 3D dual path networks for pulmonary nodules classification," *Neurocomputing*, vol. 398, pp. 422–430, 2020.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9908*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., pp. 630–645, Springer, Cham, 2016.

[23] M. Xiao, Z. Fandong, and F. Jufu, "Face recognition method based on sparse representation of deep learning features," *Journal of Intelligent Systems*, vol. 11, pp. 280–286, 2016.

[24] G. Yang, J. Yang, Z. Lu, and D. Liu, "A convolutional neural network with sparse representation," *Knowledge-Based Systems*, vol. 209, article 10641917, 2020.

[25] D. Li, Q. Wang, and F. Kong, "Adaptive kernel sparse representation based on multiple feature learning for hyperspectral image classification," *Neurocomputing*, vol. 400, pp. 97–112, 2020.

[26] J. Wang, H. Cheng, and S. D. Newman, "Sparse representation of DWI images for fully automated brain tissue segmentation," *Journal of Neuroscience Methods*, vol. 343, no. 1, article 108828, 2020.

[27] X. Li, H. Tang, D. Zhang, T. Liu, L. Mao, and T. Chen, "Histopathological image classification through discriminative feature learning and mutual information-based multi-channel joint sparse representation," *Journal of Visual Communication and Image Representation*, vol. 70, article 102799, 2020.

[28] J. Xu, W. An, L. Zhang, and D. Zhang, "Sparse, collaborative, or nonnegative representation: which helps pattern classification?," *Pattern Recognition*, vol. 88, pp. 679–688, 2019.

[29] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 78–791, 1999.

[30] S. L. Chen, S. Q. Shen, and D. S. Li, "Ensemble learning method for imbalanced data based on sample weight update," *Computer Science*, vol. 45, no. 7, pp. 31–37, 2018.

[31] T. Zhou, H. L. Lu, Z. L. Yang, B. Q. Huo, H. L. Ren, and Y. L. Dong, "The ensemble deep learning model for novel COVID-19 on CT images," *Applied Soft Computing*, vol. 98, article 106885, 2021.

[32] Z. Tao, L. Huiling, W. Wenwen, and Y. Xia, "GA-SVM based feature selection and parameter optimization in hospitalization expense," *Applied Soft Computing*, vol. 75, pp. 323–332, 2019.

Hindawi
BioMed Research International
Volume 2021, Article ID 6657566, 8 pages
https://doi.org/10.1155/2021/6657566

Research Article

Association between CT-Quantified Body Composition and Recurrence, Survival in Nonmetastasis Colorectal Cancer Patients Underwent Regular Chemotherapy after Surgery

Piaopaio Ying,[1] Wenyi Jin,[2] Xiaoli Wu,[2] and Weiyang Cai[2]

[1]*Department of Pneumology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China*
[2]*Department of Gastroenterology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China*

Correspondence should be addressed to Xiaoli Wu; doctorwxl@163.com and Weiyang Cai; wish0611@126.com

Received 9 November 2020; Revised 12 January 2021; Accepted 10 March 2021; Published 25 March 2021

Academic Editor: Yong Xia

Background. Body mass index, measured at colorectal cancer (CRC) diagnosis has been associated with recurrence and survival outcomes. Computed tomography- (CT-) defined body compositions accurately reflect body mass, but there was no consistent perspective on the influence of visceral adipose tissue (VAT) and skeletal muscle mass (SM) on the prognosis of nonmetastasis CRC, especially in the patients underwent surgery and regularly standard chemotherapy. *Methods*. We investigated the associations of CT-quantified body composition (VAT and SM) with CRC patients successively underwent surgery and regular 8-12 of periods standard chemotherapy. All of the CT images were obtained at the level of the L3/4 spinal level. The prognostic value of the body compositions was analyzed using the Cox regression model, and precise clinical nomograms were established. *Results*. In XELOX-treated patients, progression-free survival (PFS) ($P = 0.025$) and overall survival (OS) ($P = 0.032$) were lower in the high-SM than in the low-SM group. The univariate analysis demonstrated that compared with low-SM patients, patients with high-SM showed a strikingly poor prognosis in both OS ($P = 0.0512$) and PFS in the T4 subgroup ($P = 0.0417$), while contrary to the T2-3 subgroup. *Conclusions*. CT-quantified body compositions have a significant influence on CRC patients successively underwent curative resection and regularly standard chemotherapy with the endpoints of 1-year, 3-year, and 5-year both OS and PFS. Patients with high-SM showed a strikingly poor prognosis in OS and PFS in the T4 subgroup; however, the prognosis role of body composition was opposite in T2-3 patients.

1. Introduction

Currently, colorectal cancer has a high incidence rate and is the secondary causation of cancer mortality worldwide, extremely posing a threat to human health [1, 2]. Therefore, it is necessary to identify the prognostic factors of progression-free survival (PFS) and overall survival (OS) in the early time and take effective and targeted interventions to improve the prognosis. According to previous literature, overweight and obesity were found associated with recurrence-free (RF) and survival among colorectal cancer patients. However, conflicting results between them are observed in different studies. Besides, traditional index, such as body mass index (BMI), is insufficient for reflecting the distinguishing between fat and muscle mass or visceral adipose tissue (VAT) and skeletal muscle (SM) [3]. Abdominopelvic computed tomography (CT) imaging is not only a routine examination as a pretreatment staging way of clinical management in cancer patients but also a more accurate method to differentiate body composition, which means that no further test exposure is required, and no additional financial burden is placed on the patient. Moreover, CT-defined body composition is widely confirmed for accurate reflecting on different types of adipose tissue as well as muscle mass [4].

The rapid growth and proliferation of tumor cells require a large amount of energy, which is a key factor resulting in high death from tumors. Based on existing researches, baseline fat and muscle distribution are closely related to postoperation recovery and complications, tolerance of chemotherapy-induced toxicity, and recurrence as well as

health-related quality of life [5–7]. Therefore, we speculate that body compositions may play a role in nonmetastasis colorectal cancer patients. Although it is acknowledged that body compartments are independent prognostic predictors, however, there was no consistent perspective on the influence on VAT and SM to the prognosis of nonmetastasis colorectal cancer, especially in the patients who successively underwent curative resection and regularly standard chemotherapy [8, 9]. Moreover, the general consensus demonstrated that the prognosis of colorectal cancer was closely associated with the staging features of TNM classification, including pathologic T stage, absence, of nodal involvement, with or without distance. Worse prognosis in these colorectal cancer patients with distant metastases, thus, we did not include these patients to avoid bias in the results of our research. In addition, previous researches about the relationship between body compositions and the prognosis of CRC, most of which were based on the CRC stage or not, there was a lack of dividing T stage into two groups to explore the connection between them [10].

Our study is aimed at exploring the relationship between CT-quantified body compositions and recurrence as well as the overall survival of colorectal cancer patients who successfully underwent curative resection and had regular 8-12 periods of standard chemotherapy. Also, the independent risk factors in these patients were analyzed. Thus, we try to construct the OS and RF nomogram to predict the 1-year, 3-year, and 5-year survival probability based on the prognostic factors derived from multivariate Cox regression analysis.

2. Methods

2.1. Study Population and Design. The records of 221 persons diagnosed with CRC at the First Affiliated Hospital of Wenzhou Medical University between January 2014 and January 2017 were reviewed. Eligible patients were defined as firstly diagnosed with primary colorectal cancer, excluded other malignant tumors, 18 years or over, stage I–III, complete pathology, laboratory, and able to provide informed consent. Simultaneously, all of the patients successively underwent surgery and regular 8-12 periods of standard chemotherapy (including XELOX and FOLFIRI/FOLFOX). Patients who did not undergo adequate abdominopelvic computed tomography scanning before starting surgery, those treated for irregular or no chemotherapy, and those who lose followed-up for <24 months were excluded. Pathologists assessed the tumor stage according to the 8th edition of the AJCC TNM staging guidelines. In general, there were 221 eligible cases selected in this study, and the pathological T stages were T2-4. All of these patients were followed up, and 91 recurrent and 66 dead patients were recorded during the follow-up. According to the American Joint Committee on Cancer (AJCC) TNM (Tumor, Nodes, Metastasis) system and the staging 8th edition of colorectal cancer, pathologic T4 stage was defined as tumor invasion of the visceral peritoneum or adherences to adjacent organ or structure. The deep tumor penetration and invasion of adjacent organs were extremely related to the risk of relapse and overall survival among CRC patients without distant metastasis. Based on many previous studies, T4 had a significant impact on affecting both the duration and effect of chemotherapy [11]. Meanwhile, combined with other existing researches on the grouping of T staging [12], we divided all patients into T2-3 and T4 groups. The cutoff time of the study was set in August 2020. The study protocols were approved by the Wenzhou Medical University Ethics Committee. All procedures adhere to the BRISQ Guidelines for reporting research on human biospecimens.

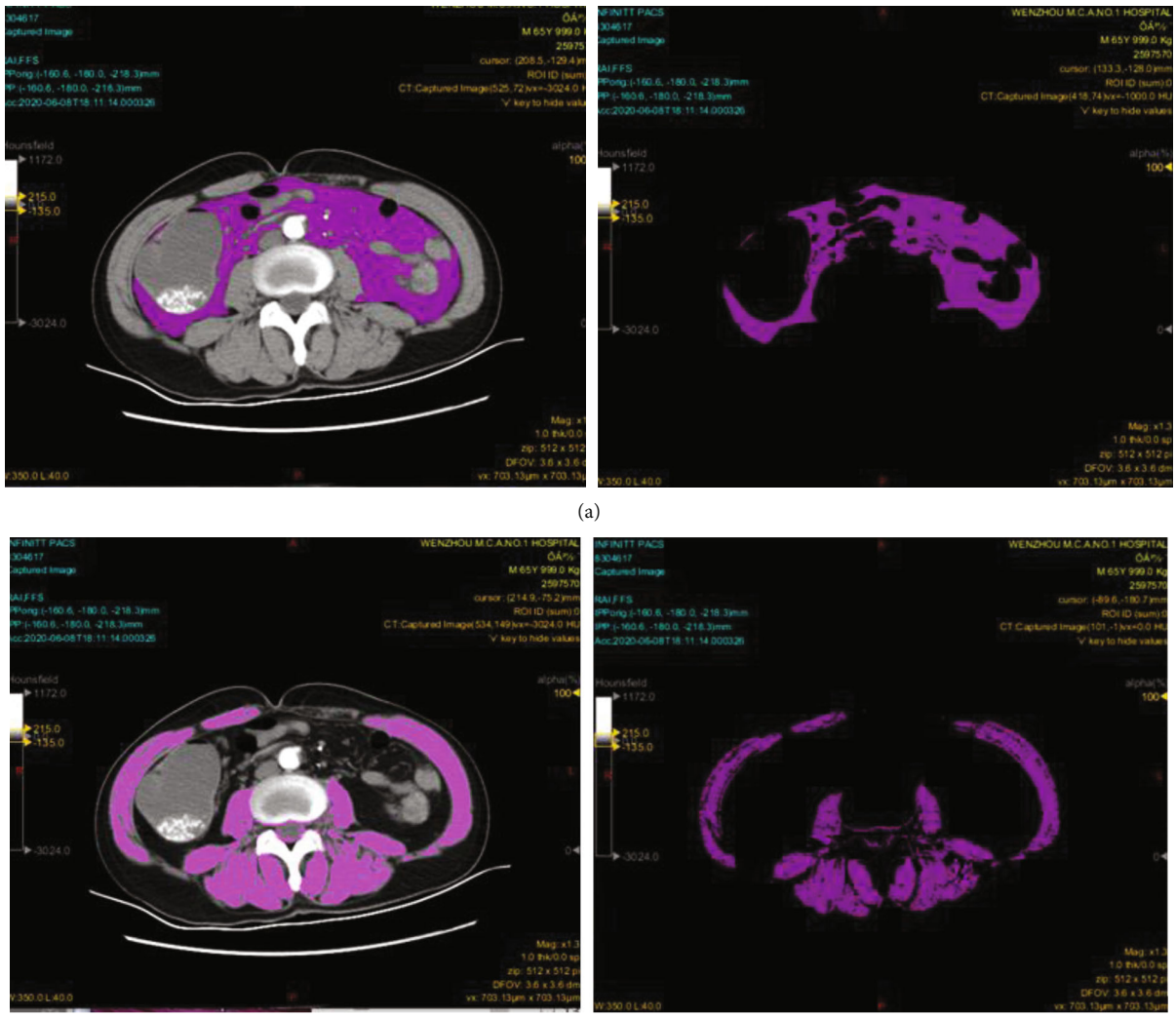
2.2. Body Composition. Muscle mass and visceral fat mass were evaluated using pretreatment CT images obtained at the level of the L3/4 spinal level in detail [13]. Patients all underwent multidetector CT scans with quantification of body composition within 15 days before surgery. Specific regions of interest (ROI) were manually determined: VAT (by defining the fascial plane of the abdominal muscle wall, using standard Hounsfield Unit (HU) ranges adipose tissue -190 to -30, Figure 1(a)) and SM (by defining the skeletal muscle using HU ranges muscle tissue 40 to 100, Figure 1(b)). All CT examinations were performed using the scanners: Brilliance-64, Philips Medical Systems, Eindhoven, The Netherlands; 128-MDCT scanner Somatom Definition, Siemens Health-care Sector, Forchheim, Germany. Two experienced radiologists drew the eligible CT planar, and then, CT analysis of the contrast-enhanced CT images was performed using LifeX software. To assess accuracy, two individuals performed scan measurements.

2.3. Body Mass Index. Patients were categorized according to their Eastern Cooperative Performance Status (ECOG-PS) into five district grades (grades 0–4) and assessed by either the treating clinician or clinical research staff. In this analysis, we gathered grade 0 as ECOG-1, grades 1-2 as ECOG-, and grades 3-4 as ECOG-3. The metabolic syndrome was internationally defined as included more than three criteria: (1) BMI was greater than 25.0 kg/m^2; (2) diagnosed with diabetes; (3) diagnosed with hypertension SBP/DBP > 140/90 mmHg; (4) blood HDL − C < 0.9 mmol/L; (5) blood TG > 1.7 mmol/L.

2.4. Statistical Analyses. R software, GraphPad Prism, and Stats were conducted for statistical analyses. The OS and PFS nomogram were constructed based on the prognostic factors derived from multivariate Cox regression analysis to predict 1-, 3-, and 5-year survival possibilities. Continuous variables were exhibited for means, medians, range, and standard deviation (SD) and compared using an independent t -test or Wilcoxon test; Spearman' correlation coefficient was used for variable correlation; Chi-square test was used to analyze categorical variables; log-rank survival analysis was employed to determine the effect of various variables on patient OS and PFS. All statistical tests were two-sided and $P < 0.05$ was considered statistically significant.

3. Results

3.1. Participators Characteristics. Of 221 eligible colorectal cancer patients, who were treated with surgery and periodic chemotherapy, were recruited from Wenzhou Medical University from 2014 January 1st to 2017 January 1st. Other



(a)

(b)

Figure 1: Example of a computed tomography (CT) scan with the area-based, densitometric quantification of adipose tissue (threshold: −190 to −30 HU) measured at spinal level L3/4: regions of interest (ROI) containing visceral fat area (VAT) (a) and an example of the densitometric quantification of muscle area (SM), dorsal and psoas muscles (threshold: 40 to 100 HU) (b).

CRC patients were excluded from the current analysis for having incomplete or no CT-based body compositions quantification or irregular chemotherapy. In our study, the proportion of patients with hypertension, diabetes, and MetS was 36.5%, 30.8%, and 12.1%, respectively. As of August 2020, 63 patients died during follow-up, none lost follow-up. Baseline clinicopathological parameters were presented in Table 1. As for the associations of adipose and muscle tissue with health-related index, we observed that SM was closely associated with ECOG and Mets score in T2-3 CRC subgroups (Table S1-S2).

Of 221 eligible patients who were diagnosed with CRC, 179 received XELOX chemotherapy and 42 received FOLFIRI/FOLFOX as first-line treatment for CRC (Table 1). The median age was 60.66 ± 12.45 months in patients treated with XELOX and 61.17 ± 16.29 months in those treated with FOLFIRI/FOLFOX. Baseline characteristics were similar between XELOX-treated and FOLFIRI/FOLFOX-treated patients except for BMI and ECOG-PS score, which was similar within the two groups.

3.2. Impact of Body Composition on Survival in CRC Patients with T Stage. The average median VAT and SM were 8.852 and 6.504, respectively. T4 stage was defined as penetrating the visceral peritoneum or directly invading or adhering to other organs or structures according to the 8th edition of the AJCC TNM staging guidelines. The degree of penetration of the tumor through the bowel wall or adhere to adjacent organs or structures played a crucial role in the prognosis of CRC. In present proof-of studies had found that the tumor stage (T4) was an independent risk factor for recurrence and OS [11, 14]. Thus, we divided the patients into $T < 4$, and T4 in the cohort. Recurrence and overall survival outcomes of <T4 and T4 patients basing on VAT and SM were shown,

Table 1: Baseline patient characteristics.

	All	Xelox ($N = 179$)	Folfox/Folfiri ($N = 42$)	P value
Characteristics	Patient (%)			
Age (range)				0.297
Median (range)	60.98 ± 12.50	60.66 ± 12.45	61.17 ± 16.29	
<50	45 (20.3%)	34 (19.0%)	11 (26.2%)	
≥50	176 (79.7%)	145 (81.0%)	31 (73.8%)	
Gender				0.161
Male	129 (53.1%)	103 (57.5%)	26 (63.4%)	
Female	92 (46.9%)	66 (42.5%)	15 (36.6%)	
T stage				0.173
T2-3	116 (19.5%)	93 (52.0%)	23 (54.8%)	
T4	105 (80.5%)	76 (48.0%)	19 (45.2%)	
Lymph node metastasis				0.135
No	138 (56.3%)	116 (64.8%)	22 (52.4%)	
Yes	86 (43.7%)	63 (35.2%)	20 (47.6%)	
TNM stage				0.197
I-II	130 (54.5%)	109 (60.9)	21 (50.0)	
III	91 (45.5%)	70 (39.1)	21 (50.0)	
BMI				**<0.001**
<18.5	12 (5.4%)	8 (4.5%)	4 (9.5%)	
18.5-25	168 (76.0%)	135 (75.4%)	33 (78.6%)	
≥25	41 (18.6%)	36 (20.1%)	5 (11.9%)	
MetS				0.893
No	196 (88.7%)	159 (88.8%)	37 (88.1%)	
Yes	25 (11.3%)	20 (11.2%)	5 (11.9%)	
ECOG				**<0.001**
1	77 (34.8%)	71 (39.7%)	6 (14.3%)	
2	119 (53.8%)	99 (55.3%)	20 (47.6%)	
>3	25 (11.3%)	9 (5.0%)	16 (38.1%)	
Status				
Alive	156 (79.1%)	133 (74.3)	23 (54.8%)	
Death	65 (20.9%)	46 (25.7)	19 (45.2%)	

respectively, in Figure S1 and Figure 2. We surprisingly found that the prognosis roles of body compositions seem to be opposite in these two subgroups. Especially, the univariate analysis demonstrated that compared with low-SM patients, patients with high-SM showed a strikingly poor prognosis in both OS ($P = 0.0512$, Figure 2(c)) and PFS in the T4 subgroup ($P = 0.0417$, Figure 2(d)). Time on the T2-3 subgroup was shorter for the low group than for the high group, although the difference was not statistically significant in OS and PFS (Figure S1).

3.3. Body Composition and Use of Chemotherapy. Totally, 179 and 42 CRC patients were treated with XELOX- and FOLFIRI/FOLFOX-treated chemotherapy, respectively, with a median of eight and twelve treatment cycles. In XELOX-treated patients, patients PFS ($P = 0.025$; Figure S2C) and OS survival ($P = 0.032$; Figure S2D) were lower in the high-SM than in the low-SM group. In FOLFIRI/FOLFOX-treated patients, overall survival ($P = 0.2108$ and 0.2701; Figure S3A and S3C), and progression-free survival ($P = 0.6163$ and 0.8542; Figure S3B and S3D) were similar between the VAT and SM high and low groups.

3.4. Construction of the CT-Based Nomogram. To establish a clinically applicable method for predicting the prognosis of CRC patients, we next established a prognostic nomogram to predict the survival probability at 1, 3, and 5 years for XELOX patients. Some independent prognostic parameters, including Mets, chemotherapy, grade, N stage, Age, VAT, and SM, were enrolled in the prediction model (Figure 3). As shown in Figure 3, the VAT and SM contributed the most risk points in CRC patients, whereas the other clinical factors contributed much less. In advanced malignant CRC patients, patients PFS and OS were higher in low-SM and VAT than in the high-SM and VAT group (Figures 3(c) and 3(d)). The trend was totally reversed in the <T3 subgroup (Figures 3(a) and 3(b)). In general, the VAT and SM were independent risk predictors for the survival of CRC patients.

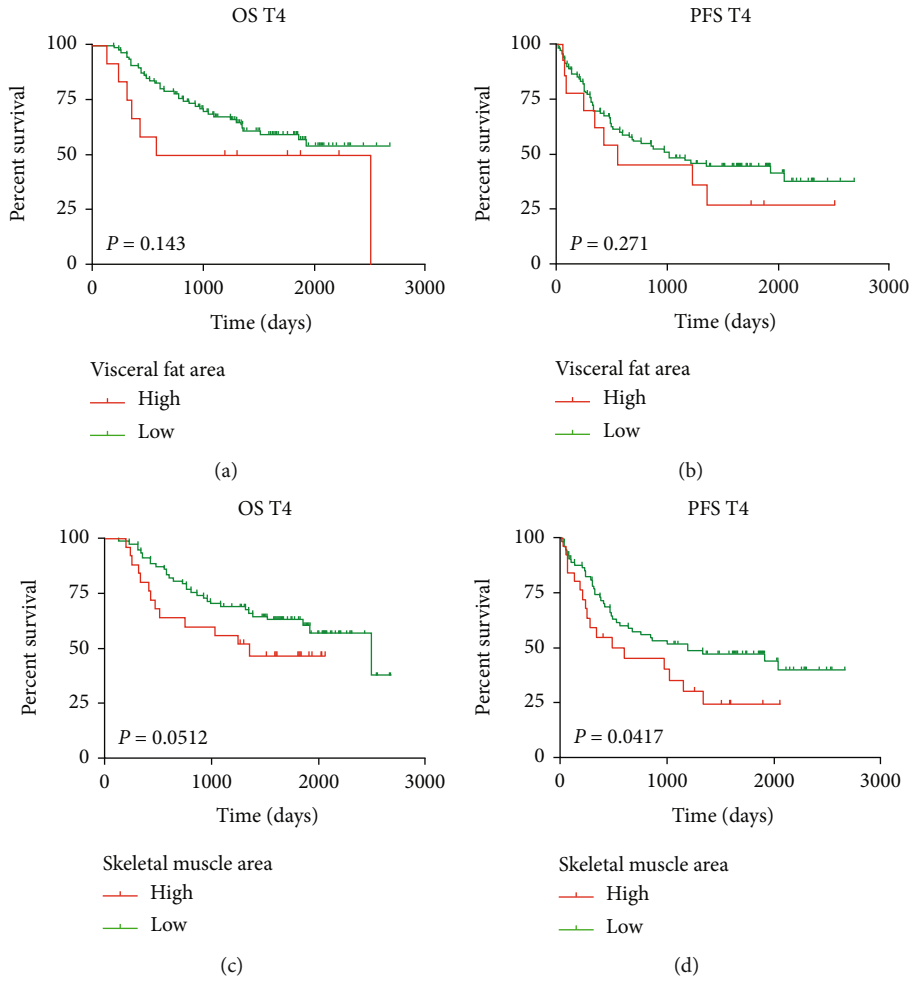


Figure 2: Outcomes T4 CRC patients based on CT body composition. Outcomes based on visceral fat in T4 GC patients from the time of diagnosis. (a) OS; (b) PFS; outcomes based on skeletal muscle in T4 GC patients from the time of diagnosis. (c) OS; (d) PFS.

4. Discussion

Body mass index has been found associated with CRC postoperative complications and survival outcomes. However, traditional indexes, such as Mets, BMI, waist/hip ratio, and ECOG-PS, do not provide detailed quantitative data for clinical reference [15]. Recently, body composition has appeared as a substitution to the traditional index. Body compositions, including skeletal muscle and visceral fat, can be estimated easily and accurately using CT images and software programs. Based on existing studies, pathological T staging was extremely associated with the duration of postoperative adjuvant chemotherapy. Furthermore, the T4 stage was the critical prognostic factor of tumor recurrence and overall survival. To our knowledge, this was the first literature to divide into subgroups according to T staging and investigate the association between body composition and disease progression, mortality, and efficacy of first-line treatment in nonmetastasis CRC patients who underwent regular chemotherapy after surgery in CT-based parameters manner. Besides, CT-quantified body composition nomograms have not systematically been estimated in CRC patients who successfully underwent curative resection and had regular standard chemotherapy.

According to the nomogram, SM and VAT played a dominant role in the prognosis of CRC patients, especially in T4 patients. We surprisingly found some conclusions worthy of consideration. High VAT and high SM were poor prognosis of RF and OS in the T4 subgroup while were protective prognosis in the T2-3 subgroup. On the other hand, body composition was associated with the risk of not only survival rate, but also chemotherapy toxicity. XELOX and FOLFOX/FOLFIRI are the most widely used first-line chemotherapy in patients with CRC. Although chemotherapy treatments have demonstrated a survival benefit and are widely approved for clinical use for CRC patients, the optimal treatment strategy remains to be determined. The therapeutic effects are unquestionably valid for patients, however, whether it can work on different physiological features of CRC patients is still controversial and has not been reported to date. Therefore, predictive markers of survival and treatment response in CRC are critically needed. In XELOX-treated patients, the high-SM group had lower progression-free survival ($P = 0.025$) and overall survival ($P = 0.032$) than

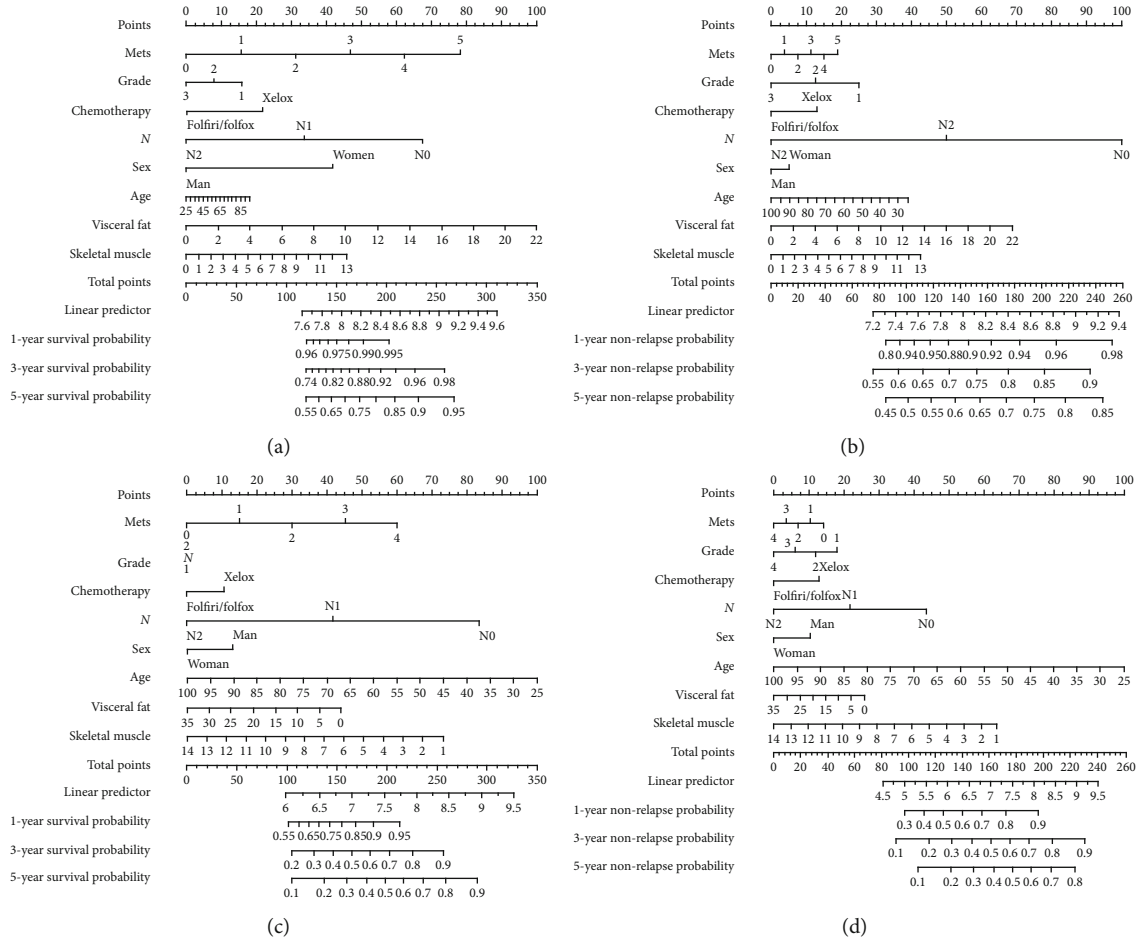


Figure 3: The nomogram to predict the 1-, 3-, and 5-year overall survival (a) and progression-free survival rates (b) of T2-T3 CRC patients. The nomogram to predict the 1-, 3-, and 5-year overall survival (c) and progression-free survival rates (d) of T4 CRC patients.

the low-SM group. Given the limited number of patients, we have not explored the associated risk of CRC progression, radiologic progression, and overall mortality in FOLFOX/-FOLFIRI-treated patients.

As was showed in Figure S1, lower VAT and SM were meaningful risk factors in CRC recurrence and overall survival among the T2-3 stages of CRC patients received regular chemotherapy after curative surgery, consistent with the view that lacking muscle mass had adverse consequences early in the malignant tumor. We speculate that low-VAT and low-SM may be related to higher chemotherapy toxicity, especially in the early stage [16]. Besides, there may be an extreme loss of fat and muscle tissue during not only the development of cancer but also the chemotherapy process [17]. T2-3 stages of CRC patients have longer OS than the T4 subgroup, higher VAT, and SM provided fat tissue and muscle tissue to be consumed and was easier to tolerate the side effects of chemotherapy, resulting in slower developing a state of cachexia and confer a survival advantage. Moreover, fewer myokines and interleukin were released as a consequence of low SM, leading to an imbalance of the immune system and having a poor impact on prognosis in the lower T stage [18]. Thus, targeted and preventive intervention strategies such as nutritional intervention and physical exercise should be given for CRC T2-3 stage patients, avoiding developing sarcopenia, thereupon then decreasing the recrudescence and overall survival rate [19, 20]. However, due to skeletal muscle did not effectively reflect muscle function, CT-based body components combined with functional measures assessed by handgrip strength and stair-climbing power would be more precise [21].

As for CRC T4 stage patients, in our research, low VAT and low SM were propitious to have a survival benefit; in addition, visceral fat had a greater impact than muscle, contradicted with T2-3 stages. The reasons why gave rise to the opposite phenomenon deserved deeper investigation. Inferring that most of the T4 stage CRC patients were at the state of cachexia, a multifactorial paraneoplastic syndrome characterized by carbohydrate, lipid, and protein metabolic disturbance, inflammatory and immunocompromised status, and higher VAT and SM would be a burden for advanced cancer patients, rather than an advantage [22]. CRC patients of the T4 stage, the higher VAT, and SM, the more enhanced inflammation, and hypermetabolism entered a vicious cycle, ending up in refractory cachexia [23]. Advanced cachexia,

nevertheless, was extremely difficult to redress for drugs, nutrition intake, and physical exercise according to existing studies. The human body reached a relative metabolic balance via abatement lipolysis and proteolysis [24]. Furthermore, high visceral fat was inclined to cause insulin resistance, which was a risk factor for cancer progression [25]. Adiposity as well as a metabolic disorder was susceptible to infection and other complications, which were contributors to shortened progression-free survival and overall survival. Moreover, immune cell expression and secretion were induced by cancer; in the pathogenesis of a malignant tumor, more fat and muscle would provide a better environment, resulting in the suppression of immune and chronic inflammation.

FOLFIRI/FOLFOX-treated was a strong combination chemotherapy regimen, and the adverse reactions caused by chemotherapy were also stronger. Patients with poor physical condition and intolerance to strong combination chemotherapy regimens were more inclined to choose the XELOX-treated regimen. Therefore, most CRC patients who choose XELOX-treated were more likely in an advanced cachexia state, which was similar to the reason for the T4 stage.

The findings outlined above suggest that high skeletal muscle mass may be associated with an increased risk of disease progression and mortality in patients with T4 nonmetastasis CRC patients. Moreover, the significance of these relationships is also significant in postoperation patients treated with XELOX. These results suggest that assessing skeletal muscle mass may be worthwhile when selecting treatments for CRC.

Nomogram is an alignment chart composed of lines of different proportions, which generates a total point to predict the likelihood of clinical events [26]. Quantify the relative contribution of each prognostic factor and convert complex regression models into visual graphics, which is more practical and convenient for evaluating the prognosis of CRC patients. According to the construction of the CT-based nomogram, some effective and targeted visceral fat and skeletal muscle mass interventions should be taken to reduce recurrence and prolong overall survival in CRC patients underwent surgery and regularly standard chemotherapy [27].

However, there are several limitations to our study. Firstly, CRC patients, without abdominal CT scans before curative surgery, were excluded. Besides, our study was comprised of a limited number of CRC patients; in particular, the diametrically inverse results in the T2-3 and T4 subgroups were found. Therefore, it will be more accurate and meaningful to establish quantified nomograms for OS and PFS if the number of samples is larger. In addition, these factors, such as correction of height, measurement of muscle density, and exclusion of fat infiltration in the muscle, had not been taken into consideration by us. Meanwhile, combined preoperative CT-quantified body component measures with muscular physical function measures will better predict the prognosis of CRC patients in the early stage. Another shortcoming of our research was the lack of regular follow-up CT scans during chemotherapy, and body composition measures were not available to verify the hypothesis.

5. Conclusions

CT-quantified body compositions have a significant influence on CRC patients successively underwent curative resection and regularly standard chemotherapy with the endpoints of 1-year, 3-year, and 5-year both OS and PFS. Patients with high SM showed a strikingly poor prognosis in OS and PFS in the T4 subgroup; however, the prognosis role of body composition was totally opposite in T2-3 patients.

Data Availability

All of the patient CT-quantified body composition data were uploaded in the supplemental file.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

Weiyang Cai and Xiaoli Wu conceived and designed the study. Piaopiao Ying and Wenyi Jin performed in data collection. Weiyang Cai, Piaopiao Ying, and Wenyi Jin analyzed the data. Piaopiao Ying, Xiaoli Wu, and Weiyang Cai wrote the manuscript.

Acknowledgments

This work was partially supported by grants from the grant LY18H030008 from Natural Science Foundation of Zhejiang Province of China, Wujieping Medical Foundation (320.6750.17396), and the First Affiliated Hospital of Wenzhou Medical University (Grant no. FHY2019002).

Supplementary Materials

Supplementary 1 Figure S1: outcomes T2-T3 CRC patients based on CT body composition. Outcomes based on visceral fat in T2-T3 CRC patients from the time of diagnosis. (a) Overall survival. (b) PFS. Outcomes based on skeletal muscle in T2-T3 CRC patients from the time of diagnosis. (c) Overall survival (d) PFS. Supplementary 2 Figure S2: outcomes CRC patients treated with XELOX based on CT body composition. Outcomes based on visceral fat in CRC patients treated with XELOX. (a) Overall survival. (b) PFS. Outcomes based on skeletal muscle in CRC patients treated with XELOX. (c) Overall survival (d) PFS. Supplementary 3 Figure S3: outcomes CRC patients treated with FOLFOX/FOLFIRI based on CT body composition. Outcomes based on visceral fat in CRC patients treated with FOLFOX/FOLFIRI. (a) Overall survival. (b) PFS. Outcomes based on skeletal muscle in FOLFOX/FOLFIRI CRC patients. (c) Overall survival (d) PFS. Supplementary 4 Table S1: correlation of VAT and body mass index. Supplementary 5 Table S2: correlation of SM and body mass index. Supplementary 6 The data registered the CRC patient CT-quantified body composition. *(Supplementary Materials)*

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] J. Ferlay, M. Colombet, I. Soerjomataram et al., "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," *International Journal of Cancer*, vol. 144, no. 8, pp. 1941–1953, 2019.

[3] P. Ying, J. Chen, Y. Ye, J. Ye, and W. Cai, *Adipose tissue is a predictor of 30-days mortality in patients with bloodstream infection caused by carbapenem-resistant Klebsiella pneumoniae*, 2021.

[4] B. H. Goodpaster, D. E. Kelley, F. L. Thaete, J. He, and R. Ross, "Skeletal muscle attenuation determined by computed tomography is associated with skeletal muscle lipid content," *Journal of Applied Physiology*, vol. 89, no. 1, pp. 104–110, 2000.

[5] J. L. A. van Vugt, R. R. J. Coebergh van den Braak, Z. S. Lalmahomed et al., "Impact of low skeletal muscle mass and density on short and long-term outcome after resection of stage I-III colorectal cancer," *European Journal of Surgical Oncology*, vol. 44, no. 9, pp. 1354–1360, 2018.

[6] H. W. Jung, J. W. Kim, J. Y. Kim et al., "Effect of muscle mass on toxicity and survival in patients with colon cancer undergoing adjuvant chemotherapy," *Support Care Cancer*, vol. 23, no. 3, pp. 687–694, 2015.

[7] R. D. Dolan, A. S. Almasaudi, L. B. Dieu, P. G. Horgan, S. T. McSorley, and D. C. McMillan, "The relationship between computed tomography-derived body composition, systemic inflammatory response, and survival in patients undergoing surgery for colorectal cancer," *Journal of Cachexia, Sarcopenia and Muscle*, vol. 10, no. 1, pp. 111–122, 2018.

[8] A. E. M. Berkel, J. M. Klaase, F. de Graaff, M. G. J. Brusse-Keizer, B. C. Bongers, and N. L. U. van Meeteren, "Patient's skeletal muscle radiation attenuation and sarcopenic obesity are associated with postoperative morbidity after neoadjuvant chemoradiation and resection for rectal cancer," *Digestive Surgery*, vol. 36, no. 5, pp. 376–383, 2019.

[9] J. J. Hopkins, R. L. Reif, D. L. Bigam, V. E. Baracos, D. T. Eurich, and M. B. Sawyer, "The impact of muscle and adipose tissue on long-term survival in patients with stage I to III colorectal cancer," *Diseases of the Colon and Rectum*, vol. 62, no. 5, pp. 549–560, 2019.

[10] G. Malietzis, O. Aziz, N. M. Bagnall, N. Johns, K. C. Fearon, and J. T. Jenkins, "The role of body composition evaluation by computerized tomography in determining colorectal cancer treatment outcomes: a systematic review," *European Journal of Surgical Oncology*, vol. 41, no. 2, pp. 186–196, 2015.

[11] R. Labianca, B. Nordlinger, G. D. Beretta et al., "Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†," *Annals of Oncology*, vol. 24, no. 6, pp. -vi64–vi72, 2013.

[12] D. M. Hari, A. M. Leung, J. H. Lee et al., "AJCC Cancer Staging Manual 7th edition criteria for colon cancer: do the complex modifications improve prognostic assessment?," *Journal of the American College of Surgeons*, vol. 217, no. 2, pp. 181–190, 2013.

[13] W. Shen, M. Punyanitya, Z. Wang et al., "Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image," *Journal of Applied Physiology*, vol. 97, no. 6, pp. 2333–2338, 2004.

[14] P. Parent, R. Cohen, E. Rassy et al., "A comprehensive overview of promising biomarkers in stage II colorectal cancer," *Cancer Treatment Reviews*, vol. 88, p. 102059, 2020.

[15] R. D. Dolan, L. E. Daly, C. P. Simmons et al., "The relationship between ECOG-PS, mGPS, BMI/WL grade and body composition and physical function in patients with advanced cancer," *Cancers*, vol. 12, no. 5, p. 1187, 2020.

[16] E. M. Cespedes Feliciano, E. Avrutin, B. J. Caan, A. Boroian, and M. Mourtzakis, "Screening for low muscularity in colorectal cancer patients: a valid, clinic-friendly approach that predicts mortality," *Journal of Cachexia, Sarcopenia and Muscle*, vol. 9, no. 5, pp. 898–908, 2018.

[17] Z. Aversa, P. Costelli, and M. Muscaritoli, "Cancer-induced muscle wasting: latest findings in prevention and treatment," *Therapeutic Advances in Medical Oncology*, vol. 9, no. 5, pp. 369–382, 2017.

[18] S. Looijaard, M. L. Te Lintel Hekkert, R. C. I. Wust, R. H. J. Otten, C. G. M. Meskers, and A. B. Maier, "Pathophysiological mechanisms explaining poor clinical outcome of older cancer patients with low skeletal muscle mass," *Acta Physiologica*, vol. 231, no. 1, article e13516, 2020.

[19] R. Oliver, "Comment on: Sufficient levels of 25-hydroxyvitamin D and protein intake required to increase muscle mass in sarcopenic older adults - the PROVIDE study," *Clinical Nutrition*, vol. 37, no. 5, p. 1770, 2018.

[20] E. Conte, E. Bresciani, L. Rizzi et al., "Cisplatin-induced skeletal muscle dysfunction: mechanisms and counteracting therapeutic strategies," *International Journal of Molecular Sciences*, vol. 21, no. 4, p. 1242, 2020.

[21] S. A. Kurk, R. K. Stellato, P. H. M. Peeters et al., "Trajectory of body mass and skeletal muscle indices and disease progression in metastatic colorectal cancer patients," *The American Journal of Clinical Nutrition*, vol. 110, no. 6, pp. 1395–1403, 2019.

[22] M. S. Razzaque and A. Atfi, "Regulatory role of the transcription factor Twist1 in cancer-associated muscle cachexia," *Frontiers in Physiology*, vol. 11, p. 662, 2020.

[23] M. J. Tisdale, "Mechanisms of cancer cachexia," *Physiological Reviews*, vol. 89, no. 2, pp. 381–410, 2009.

[24] K. Fearon, J. Arends, and V. Baracos, "Understanding the mechanisms and treatment options in cancer cachexia," *Nature Reviews Clinical Oncology*, vol. 10, no. 2, pp. 90–99, 2013.

[25] R. Mizuno, A. Miyajima, T. Hibi et al., "Impact of baseline visceral fat accumulation on prognosis in patients with metastatic renal cell carcinoma treated with systemic therapy," *Medical Oncology*, vol. 34, no. 4, p. 47, 2017.

[26] H. Shi, Y. Jiang, H. Cao, H. Zhu, B. Chen, and W. Ji, "Nomogram based on systemic immune-inflammation index to predict overall survival in gastric cancer patients," *Disease Markers*, vol. 2018, Article ID 1787424, 11 pages, 2018.

[27] B. Gigic, J. Nattenmuller, M. Schneider et al., "The role of CT-quantified body composition on longitudinal health-related quality of life in colorectal cancer patients: the colocare study," *Nutrients*, vol. 12, no. 5, p. 1247, 2020.

Hindawi
BioMed Research International
Volume 2021, Article ID 5513746, 12 pages
https://doi.org/10.1155/2021/5513746

Research Article

Computer-Aided Diagnosis Research of a Lung Tumor Based on a Deep Convolutional Neural Network and Global Features

Huiling Lu [1,2]

[1]*School of Science, Ningxia Medical University, Yinchuan 750004, China*
[2]*Key Laboratory of Images & Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China*

Correspondence should be addressed to Huiling Lu; lu_huiling@163.com

Received 8 January 2021; Revised 8 February 2021; Accepted 17 February 2021; Published 28 February 2021

Academic Editor: Lin Gu

Based on the better generalization ability and the feature learning ability of the deep convolutional neural network, it is very significant to use the DCNN on the computer-aided diagnosis of a lung tumor. Firstly, a deep convolutional neural network was constructed according to the fuzzy characteristics and the complexity of lung CT images. Secondly, the relation between model parameters (iterations, different resolution) and recognition rate is discussed. Thirdly, the effects of different model structures for the identification of a lung tumor were analyzed by changing convolution kernel size, feature dimension, and depth of the network. Fourthly, the different optimization methods on how to influence the DCNN performance were discussed from three aspects containing pooling methods (maximum pooling and mean pooling), activation function (sigmoid and ReLU), and training algorithm (batch gradient descent and gradient descent with momentum). Finally, the experimental results verified the feasibility of DCNN used on computer-aided diagnosis of lung tumors, and it can achieve a good recognition rate when selecting the appropriate model parameters and model structure and using the method of gradient descent with momentum.

1. Introduction

Lung cancer is regarded as one of the malignant tumors with high morbidity and mortality, which is a serious threat to human health and life, and it is very difficult for lung cancer patients to be discovered and diagnosed because there are no apparent symptom and typical imaging performance [1]; therefore, early detection and diagnosis are essential for lung cancer patients. In the early diagnosis of lung cancer, multilayer spiral CT (MSCT) can show the transverse, vertical, and coronal planes of the lesion area clearly by a reconstruction technique. In the interim diagnosis, MSCT jointed surface shadowing and multiple planar reconstruction can clearly show the location of the tumor, internal structure, edge features, blood supply, extent of surrounding tissue invasion, and surrounding tissue changes; it gets high accuracy [2]. Hence, the CT image provides an important reference for the diagnosis and identification of lung cancer. Aimed at the massive medical image data, it can reduce the doctors' workload, improve the recognition rate, and reduce the misdiagnosis rate and missed diagnosis rate with the help of computer-aided diagnosis (CAD).

Deep learning [3] as a new field of machine learning analyses and interprets data through simulating a human brain. In particular, a convolutional neural network with the unique deep structure can learn the complex mapping between input and output effectively. At present, the design of the DCNN model is mainly focused on the model parameters, the activation function, the size of the receptive field, the designation of the pooling layer, and so on. Based on the classical model LeNet-5 structure, Chen [4] constructed several different convolutional neural networks by adjusting the number of parameters and the interlayer connection mode, and they were used on optical digital recognition. Ma et al. [5] simplified network structure by removing the third convolutional layer of LeNet-5 and replacing the softmax classifier with an SVM classifier. Hinton constructed a CNN with five convolutional layers, and it achieved good results when used on ImageNet data set [6] in 2012. Gao [7] used whitening pretreatment and stochastic pooling based on traditional CNN;

it improved the network generalization ability for military image classification by this method. Zhang et al. [8] constructed a deep convolutional neural network (DCNN) with seven layers for the vehicle type identification; the recognition rate reached 96.8% based on comparative experiment with different model parameters. Guo [9] constructed a DCNN used on hand-printed character recognition; the experimental results show that the receptive field size has a significant influence on the number of model parameters but has little effect on the recognition rate and the running time is in a reverse trend. He and Sun [10] discussed how to balance the number of layers, the number of feature maps, and the size of the convolution kernel in the limited training time and computational complexity; they showed that the recognition performance of CNN with small convolution kernel and deep layers is better than that with large convolution kernel and shallow layers. Gunavathi et al. [11] give a review on convolutional neural network-based deep learning methods in gene expression data for disease diagnosis. Zhou et al. [12] propose a lung tumor Computer-aided diagnosis model in chest CT image based on DenseNet-NSCR (Non-negative, Sparse and collaborative representation classification of DenseNet) in this paper; the result shows that the DenseNet+NSCR model has better robustness and generalization capabilities compared with AlexNet+SVM, AlexNet+SRC, AlexNet+NSCR, GoogleNet+SVM, GoogleNet+SRC, GoogleNet+NSCR, DenseNet-201+SVM, and DenseNet-201+SRC. Zhou et al. [13] use AlexNet, GoogleNet, and ResNet to realize the ensemble deep learning model for novel COVID-19 on CT images. A novel method for stock trend prediction uses a graph convolutional feature-based convolutional neural network (GCCNN) model, in which both stock market information and individual stock information are considered in Chen et al. [14]. The deep convolutional neural network (DCNN) [15] can automatically extract the high-level features of the image and express the image effectively, and the data is mapped into a new space by the linear or nonlinear transformation of the input data; by this way, the essential feature of an image can be extracted effectively and stably. However, it is necessary to optimize the DCNN for a specific research object and application field. In this paper, the deep convolutional neural network was proposed to identify a lung tumor based on global features of the CT image; on the basis of the original DCNN, the effects of different model parameters, model structure, and optimization algorithm on the recognition performance are discussed in order to validate the feasibility of DCNN used on computer-aided diagnosis of lung tumors and provide a reference for computer-aided diagnosis of a lung tumor.

2. Method and Material

2.1. DCNN Model Structure. The deep convolutional neural network model is a simulation of simple and complex cell function in the visual cortex, and it extracted features through the alternate convolutional layer and pooling layer and combined with the corresponding classifier to realize image recognition.

2.1.1. Convolutional Layer. Each convolutional layer [16] is composed of multiple feature maps, the neurons in each feature map are connected to the neurons in the input layer or the pooling layer, and the neuron in the same feature map is connected with the neuron in the corresponding receptive field sharing weight; each output feature map can be combined with multiple feature maps:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_i^j\right), \tag{1}$$

where x_j^l is the output map of channel j in convolutional layer l, k_{ij}^l represents the convolution kernel matrix, and b_i^j is the bias; the different convolution kernel weights have different convolution operations.

2.1.2. Pooling Layer. Each pooling layer [17] is also composed of multiple feature maps, the number of feature maps are the same as the number of feature maps in the convolutional layer, and the neuron values are calculated by the maximum or average pooling. The pooling formula is as follows:

$$x_j^l = f\left(\beta_j^l * \text{down}\left(x_i^{l-1}\right) + b_i^j\right), \tag{2}$$

where down$(\cdot)$ represents the pooling function, and each output image has its own multiply bias β and additive bias b.

2.1.3. Full Connection Layer. In the full connection layer, the feature maps of all the two dimensional images are connected to one dimension features as the input of the full connection network, and the output of the full connection layer is obtained by the weighted summation of inputs and calculating the response of the activation function:

$$x^l = f\left(w^l x^{l-1} + b^l\right) \tag{3}$$

where w^l is the weight coefficient of the connected network, x^{l-1} represents feature maps, and b^l is the bias of the full connection layer.

2.2. Training Method of DCNN. The DCNN training process mainly used the backpropagation algorithm, that is, training data input, activation values of each neuron calculation, then error calculation, gradient calculation of each weight and bias, and weight and deviation adjustment.

2.2.1. Gradient Calculation of the Full Connection Layer. For the full connection layer of DCNN, the BP [18] is used to calculate the partial derivative that error function acts on weight bias. Assuming a multiclassification problem contains N training samples and C types, the formula of error function is $E^N = 1/2\sum_{n=1}^{N}\sum_{k-1}^{c}(t_k^n - y_k^n)^2$, where t_k^n represents a class label corresponding to the first k dimension between n samples and y_k^n represents the predicted output value corresponding to the first k dimension between n samples.

The "error" of the backpropagation network is regarded as the "sensitivity" of each neural unit to the deviation (residual), and the partial derivative of error with respect to network parameters is defined as follows:

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial u}\frac{\partial u}{\partial b} = \delta, \tag{4}$$

in the formula $\partial u/\partial b = 1$.

The sensitivity of output layer neurons was calculated by the following formula:

$$\delta^L = f'\left(u^L\right) \circ \left(y^n - t^n\right), \tag{5}$$

where "$\circ$" expresses the dot product; that is, the corresponding elements in the matrix were multiplied.

The sensitivity calculation formula of the full connection layer is $\delta^l = \left(W^{l+1}\right)^T \delta^{l+1} \circ f'\left(u^l\right)$.

The update rule for the weights of neurons is to multiply the input of the neuron with the triangle of the neuron. It is the inner product of the input vector and the residual vector by the vector representation:

$$\begin{aligned} \frac{\partial E}{\partial W^l} &= x^{l-1}\left(\delta^l\right)^T, \\ \Delta W^l &= -\eta\frac{\partial E}{\partial W^l}. \end{aligned} \tag{6}$$

Usually, each weight w_{ij} has a corresponding η_{ij} differently.

2.2.2. Gradient Calculation of the Convolutional Layer. Each convolutional layer l of CNN is connected with the pooling layer $l+1$; in the process of backpropagation, we need to sum up all the residuals in the layer $l+1$ corresponding to the neuron and calculate the residuals of neurons in the layer l; then, these residuals are multiplied by the corresponding weights and multiplied by the function of the current neuron. Calculate sensitivity by chain derivation:

$$\delta_j^l = \beta_j^{l+1}\left(f'\left(u_j^l\right) \circ \mathrm{up}\left(\delta_j^{l+1}\right)\right), \tag{7}$$

where $\mathrm{up}(\cdot)$ represents upsampling.

The gradient formula of bias b by δ_j^l is $\partial E/\partial b_j = \sum_{u,v}\left(\delta_j^l\right)_{uv}$.

The gradient formula of bias k by using MATLAB is

$$\frac{\partial E}{\partial k_{ij}^l} = \mathrm{rot180}\left(\mathrm{conv2}\left(x_i^{l-1}, \mathrm{rot180}\left(\delta_j^l\right), 'valid'\right)\right). \tag{8}$$

2.2.3. Gradient Calculation of the Pooling Layer. In the backpropagation process of the pooling layer, the residual graph is first calculated, and then, the two learning parameters with β and b are updated.

δ can be calculated by MATLAB:

$$\delta_j^l = f'\left(u_j^l\right) \circ \mathrm{conv2}\left(\delta_j^{l+1}, \mathrm{rot180}\left(k_j^{l+1}\right), 'full'\right). \tag{9}$$

The gradient of the additive bias b is the sum of the elements in the residual graph:

$$\frac{\partial E}{\partial b_j} = \sum_{u,v}\left(\delta_j^l\right)_{uv}. \tag{10}$$

The gradient of the multiplicative bias β is $\partial E/\partial \beta_j = \sum_{u,v}\left(\delta_j^l \circ d_j^l\right)_{uv}$.

2.3. Evaluation Indicator. In this paper, six evaluation indexes are selected to measure the experimental results: accuracy, sensitivity, specificity, Matthews correlation coefficient (MCC), F_1 score [13], and training time, and they are calculated by true positive (TP), false positive (FP), true negative (TN), and false negative (NN). Besides, TP indicates that the normal image is predicted to be normal, FP indicates that the abnormal image is predicted to be normal, TN indicates that lung tumor images are predicted as lung tumor images, and FN indicates that the normal lung image is predicted to be abnormal.

(1) Training time: it is the time that the algorithm spends from start to finish; in the process of convolution operation, the total time of the whole training process and testing process is expressed when the specified iteration times are reached

(2) Accuracy: it is the description of the correct classification of the lung CT image; the value is between 0 and 1; the greater the value, the better the classifier; and this value reflects the performance of the correct identification

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}. \tag{11}$$

(3) Sensitivity and specificity: sensitivity indicates that the proportion of the normal lung image is accurately recognized, and specificity indicates that the proportion of lung tumor images is accurately identified:

$$\begin{aligned} \mathrm{sensitivity} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \\ \mathrm{specificity} &= \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}. \end{aligned} \tag{12}$$

(4) MCC: it is a more balanced evaluation standard, which takes into account the true false positives and

false negatives; especially, in the case of different numbers, it is generally considered to be a balanced measure. MCC is essentially a correlation coefficient between the observed and predicted binary classifiers and returns the value between -1 and 1, where 1 represents a perfect prediction, 0 represents a random prediction, and -1 indicates that the classification result is completely wrong. The MCC formula is as follows:

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}}. \quad (13)$$

(5) F_1 score: it is an index measuring the recognition performance of two classification models, taking into account the classification accuracy and recall rate, which can be regarded as a kind of weighted average with precision and recall. Its maximum value is 1, the minimum value is 0, and the value that is closer to 1 indicates that the accuracy is higher. The formula is as follows:

$$\begin{aligned}
\mathrm{Precision} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \\
\mathrm{Recall} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \\
\mathrm{NegPrecision} &= \frac{\mathrm{TN}}{\mathrm{FN} + \mathrm{TN}}, \\
\mathrm{NegRecall} &= \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \\
F_1\ \mathrm{score} &= \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} + \frac{\mathrm{NegPrecision} * \mathrm{NegRecall}}{\mathrm{NegPrecision} + \mathrm{NegRecall}}.
\end{aligned} \quad (14)$$

2.4. Computer-Aided Diagnosis Model of a Lung Tumor Based on DCNN and Global Fuzzy Feature

2.4.1. Algorithm Idea. In order to verify the feasibility of the convolutional neural network used on medical diagnosis of a lung tumor, a deep convolutional neural network is designed for lung tumor recognition based on the global features of CT images. Although the deep learning model has good generalization ability and robustness [19], the performance of different models is different for different image recognition. According to the fuzzy characteristics of lung CT images and the complexity of medical images, three aspects model parameters, model structure, and optimization algorithm are discussed. Firstly, the effects of different resolution and iteration times on the identification results are discussed. Secondly, different DCNN models are constructed and analyzed from the different convolution kernel sizes, feature dimensions, and network layers. Finally, the different optimization methods on how to influence the DCNN performance from three aspects including sampling methods (maximum pooling and mean pooling), activation function (sigmoid and ReLU), and training algorithm (batch gradient descent and gradient descent with momentum) were discussed. In a word, the research objective is to explore the optimal DCNN model for lung tumor computer-aided diagnosis.

The specific steps for computer-aided diagnosis research of a lung tumor based on DCNN and global features are as follows:

(1) Data collection: 5000 CT images with DICOM format were collected from the General Hospital of Ningxia Medical University. The 2500 images were selected as the lung tumor images according to the doctor's mark and the doctor's advice, and the 2500 normal CT images were taken as contrast images

(2) Image preprocessing: the collected images were converted into gray image and normalized to the same size of experimental data, and then, lung tumor CT data set was constructed for DCNN training and testing

(3) Construction of DCNN: construct a deep convolutional neural network with eight layers aimed at global features of lung cancer, including an input layer, 3 convolutional layers, 3 pooling layers, 2 full connection layers, and an output layer with a softmax classifier

(4) Discussion based on different model parameters with the same model structure: the influence of different resolution and the number of iterations on the DCNN recognition rate and training time is discussed aimed at the CT global feature set

(5) Discussion of different model structures: based on the initial construction of DCNN, the recognition performance for the lung tumor by changing the convolution kernel size and the number of feature maps and network layers is discussed

(6) Comparison analysis of different optimization algorithms: after choosing the suitable model, comparative experiments with different methods were done, including pooling method (mean pooling and maximum pooling), activation function (sigmoid function and ReLU function), and training algorithm (batch gradient descent method and gradient descent method with momentum)

(7) Evaluation: Construct an optimal DCNN model through the analysis of experiment and comparison of different model parameters and structures, and it is used on computer-aided diagnosis of a lung tumor based on global features, in order to improve the recognition rate, reduce the training time, and enhance the robustness and generalization ability

2.4.2. Model Construction. The convolutional neural network can directly input the original image and has obvious advantage to the complex image recognition; besides, CT is widely used in the diagnosis of a lung tumor, but the lesion area

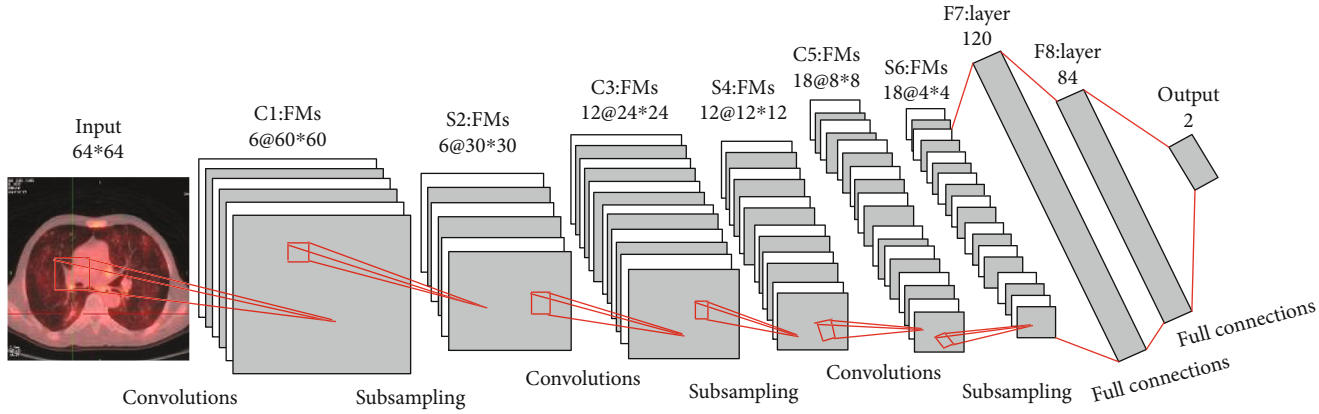


Figure 1: DCNN model construction.

accounted for a smaller region in the whole CT images, and the characteristic is not obvious and it is difficult to distinguish, so a deep convolution neural network is constructed to extract the hidden features of the lung tumor for computer-aided diagnosis. The DCNN model structure is shown in Figure 1.

(1) Input layer: the input image is the whole CT image of 64 ∗ 64, and it is also the lung global feature input for DCNN classification

(2) C1 layer: it is the first convolutional layer, each neuron is convoluted with the local receptive field of the input image 5 ∗ 5, feature maps' size is 60 ∗ 60, and it contains 6 different feature maps

(3) S2 layer: it is the first pooling layer, the pooling method was used on 2 ∗ 2 neighborhood, feature maps' size is 30 ∗ 30, and it contains 12 different feature maps

(4) C3 layer: it is the second convolutional layer, 12 convolution kernels of 7 ∗ 7 are used for convolution operation, feature maps' size is 24 ∗ 24 after convolution, and it contains 12 different feature maps

(5) S4 layer: it is the second pooling layer, and it got 12 feature maps of 12 ∗ 12 after subsampling without repetition

(6) C5 layer: it is the third convolutional layer, and it contains 18 feature maps of 8 ∗ 8

(7) S6 layer: it is the third pooling layer, and it contains 18 feature maps of 4 ∗ 4

(8) F7 layer: it is the first full connection layer, and it contains 120 neurons and connects with the S6 layer

(9) F8 layer: it is the second full connection layer, and it contains 84 neurons and connected with the upper layer and the output layer

(10) Output layer: it connects with the softmax classifier that is used to calculate the probability of different images that belong to which types. The formula is as follows:

$$d_j^{(i)} = \frac{\exp\left(W_j^T x^{(i)} + a_j\right)}{\sum_{j=1}^{2} \exp\left(W_j^T x^{(i)} + a_j\right)}, \quad (15)$$

where $W = [W_1, W_2] \in R^{d\times 2}$ and $a = [a_1, a_2] \in R^{d\times 2}$ are classifier parameters and $d_j^{(i)}$ is a possibility prediction for which $x^{(i)}$ belongs to class j; finally, the output is of two types: normal and abnormal lung images.

3. Discussion and Conclusion

3.1. Experiment Platform. The software and hardware environment is as follows:

Software environment: Windows 7 operating system, MATLAB R2014b

Hardware environment: Intel Xeon CPU E5-2407 v2 @ 2.40 GHz, 32.0 GB memory, 3.5 TB HD

3.2. Experimental Data

3.2.1. Data Sources. 5000 CT images were collected from a hospital in Ningxia of China: the 2500 images were selected as the lung tumor images according to the doctor's mark and the doctor's advice, and the 2500 normal CT images were taken as contrast images; the original image is shown in Figure 2.

3.2.2. Data Preprocessing. Firstly, the CT image of a lung tumor was selected according to the marker of three PET/CT modality images; secondly, the normal lung CT image was selected according to the DICOM documents and the doctor's advice; then, the experimental data is converted into a gray image; finally, the experimental data are normalized to the same size: 4000 cases are selected as training data randomly and 1000 cases are regarded as test data. The pretreated images are shown partly in Figure 3: abnormal

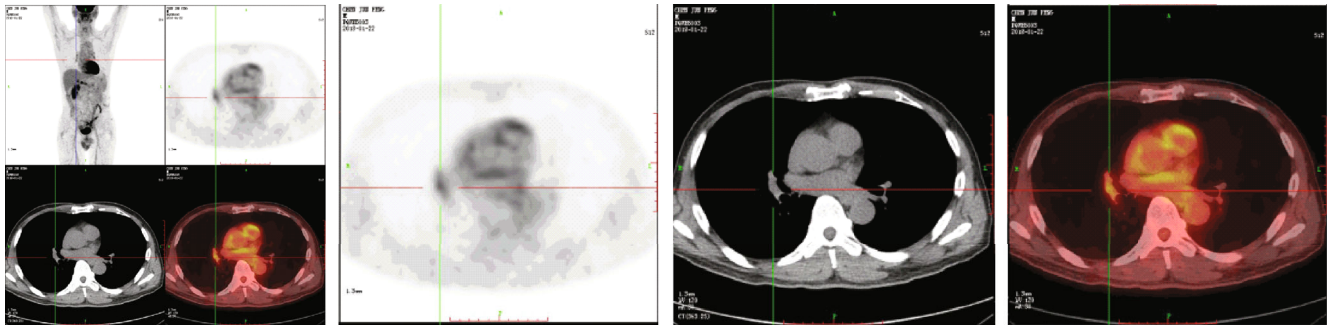


Figure 2: Original experimental data.

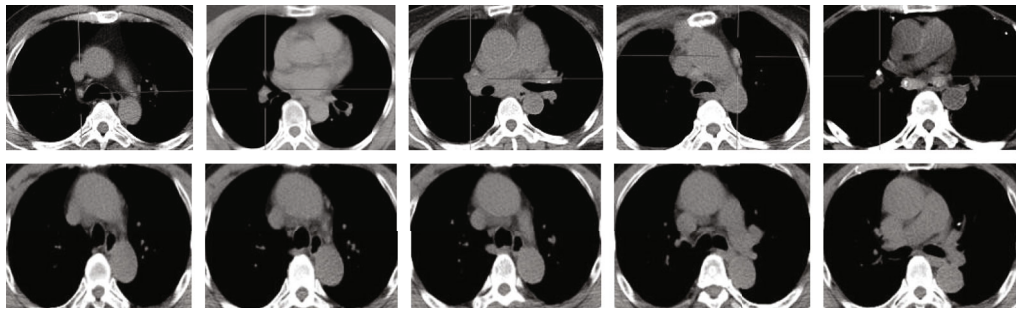


Figure 3: Experimental data after pretreatment.

images are shown in the first line and normal images are shown in the second line.

3.3. Analysis of Experimental Results

3.3.1. Experiment 1: Research on Different Model Parameters Based on the Same Model Structure

(1) Output of Intermediate Feature Maps. Image recognition using the deep convolution neural network is based on the abstract features of the hidden layer, and three convolutional layers and three pooling layers are used to extract and output features from different angles after input original images. The output of the intermediate feature maps is shown in Figure 4. From left to right, the feature maps are shown of which the original input image is C1, S2, C3, S4, C5, and S6. It clearly showed that the edge information and contour information of the input image are extracted by the first two layers; that is, the characteristics of low level, such as image edges, lines, and angles, are extracted by the bottom of the convolutional layers. And the abstraction of higher semantic information and essential information is performed by back layers. It cannot be identified with the naked eye, and it has also shown the superior learning ability of deep learning. In a word, the bottom layer of DCNN can learn the physical features such as edge and shape. With the increase in the number of hidden layers in the network, more complex and abstract visual features can be learned.

(2) The Influence of Different Resolution on Recognition Results. Due to the different resolution of the images used in the training samples, different convolution and downsampling operations will affect the recognition rate of the model, so based on the same model structure of the convolutional neural network, it was selected among different lung CT images with different resolution for the experiments, including $28 * 28$, $32 * 32$, $64 * 64$, and $100 * 100$ different resolution images. The experimental results are shown in Table 1.

According to the table, we can see that (1) the higher the resolution of the image, the longer the training time is; that is, the more complex the image is, the longer the training time is. (2) The higher the resolution, the higher the recognition rate, because of the low resolution of the image which means that the input information of the image is lost in different degrees. (3) The sensitivity is generally higher than the specificity regardless of the resolution. The results show that the lung tumor image is easy to be recognized as a normal image, and it also accords with the current situation of pulmonary nodules missed. (4) As for the MCC and F_1 score, the higher the resolution, the higher the value.

In short, high-resolution images will not only lead to more processing time but also reduce the quality of spatial resolution, but the deep convolutional neural network got high accuracy for high-resolution image recognition, so the CT images of $64 * 64$ resolution are chosen for subsequent experiments that are about different model structures and optimization methods based on the consideration of time complexity and accuracy.

(3) The Influence of Iterations on Recognition Results. The iterative method is used to calculate the weight of the convolutional neural network model, the weight and the error will be adjusted in each iteration, and the experimental results

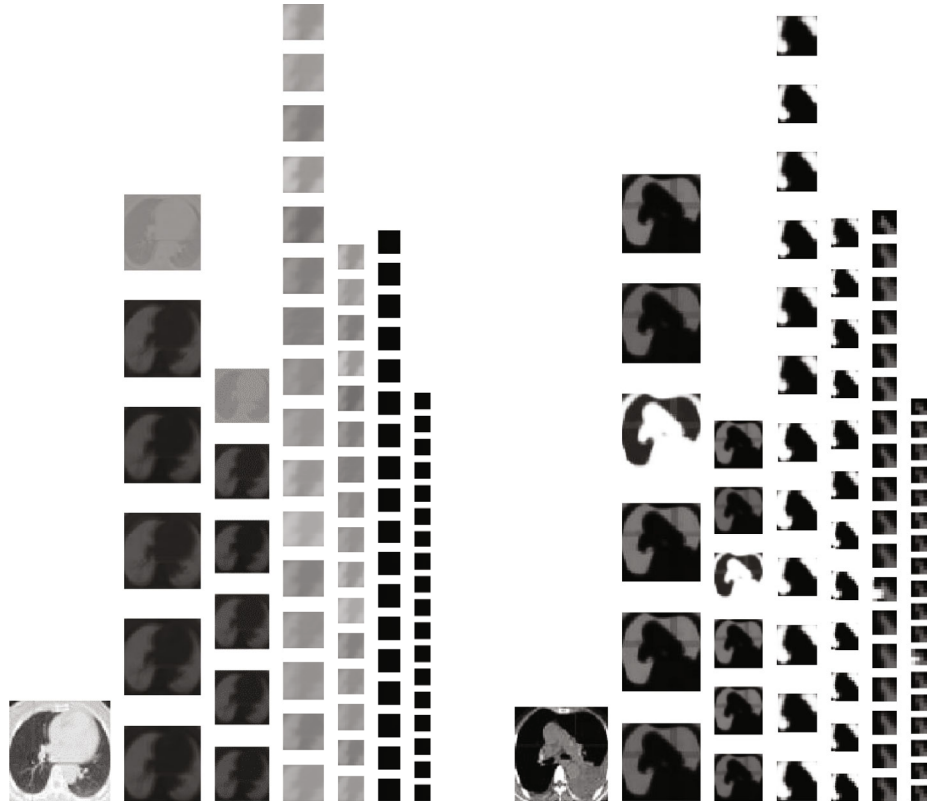


Figure 4: Output of the intermediate feature maps.

Table 1: Effects of different resolution samples on experimental results.

Different resolution	Time (s)	Accuracy (%)	Error (%)	Sensitivity (%)	Specificity (%)	MCC	F_1 Score
28 ∗ 28	68.66	76.20	23.80	99.40	53.00	0.59	0.75
32 ∗ 32	92.21	68.50	31.50	100.00	37.00	0.47	0.65
64 ∗ 64	605.97	87.30	12.70	95.40	79.20	0.76	0.87
100 ∗ 100	1328.17	89.70	10.30	99.00	80.40	0.81	0.90

Table 2: Effects of different iterations on the experimental results.

Iterations	1	10	30	50	100	150	200	250	300
Accuracy (%)	50.00	59.98	77.25	75.55	83.43	86.83	84.13	83.93	85.33
Time (s)	25.08	256.78	700.35	1250.68	2429.52	3750.23	5016.18	6274.23	7803.60

will be different with different iterations; in this experiment, the ideal weight parameters are obtained by many iterations, and then, the influence of the number of iterations on the recognition results is discussed. The experimental results are shown in Table 2.

It can be seen from Table 2 that, with the increase in the number of iterations, the accuracy increases first and then decreases, while the training time increases with the iterations increasing. The main reason lies in the idea that the learning of the convolutional neural network is not sufficient when the number of iterations is lower than the normal times. With the increase in the number of iterations, the network has achieved a high recognition rate in the training and learning processes. However, when the number of iterations increases to a certain degree, the recognition rate will decrease. It shows that the network model was trained under the appropriate iterations, the parameters have been optimized to the optimal state, the network also entered the convergence phase, and the network model got the best performance. The increase in the number of iterations will affect the change of the training time, and this change has a positive correlation, and the test time and the number of iterations are not directly linked.

Table 3: Effect of different convolution kernel sizes on experimental results.

Convolution kernel size			Time (s)	Accuracy (%)	Error (%)	Sensitivity (%)	Specificity (%)	MCC	F_1 score
k1	k2	k3							
5	5	5	868.26	69.70	30.30	99.80	39.60	0.49	0.67
5	7	5	779.64	85.30	14.70	80.80	89.80	0.71	0.85
5	9	9	950.25	80.90	19.10	100.00	61.80	0.67	0.80
5	11	11	1150.56	86.30	13.70	99.60	73.00	0.75	0.86
7	5	7	706.79	83.40	16.60	97.60	69.20	0.70	0.83
11	11	9	1765.49	82.30	17.50	99.60	65.40	0.69	0.82

3.3.2. Experiment 2: Recognition Based on Different Model Structures. The deep convolutional neural network is constructed for the first time, and it contains 1 input layer, 3 convolutional layers, 3 pooling layers, 2 full connection layers, and 1 output layer. Input images are lung CT images of 64 $*$ 64; the number of feature maps of three convolutional layers are, respectively, 6, 12, and 18; the convolution kernel size is 5 $*$ 5, 7 $*$ 7, and 5 $*$ 5; sigmoid function is regarded as activation function; the output layer connected with the softmax classifier, and the outputs are of two types: normal and abnormal lung images. On the basis of the construction of the deep convolution neural network structure, the convolution kernel size, the number of feature maps, and the depth of the network are changed, and they are used to do further experiments.

(1) Different Convolution Kernel Sizes. In order to discuss the influence of different convolution kernel sizes on DCNN recognition results, the deep convolutional neural network structure was fixed and the DCNN was trained by using different convolution kernels; the results are shown in Table 3.

At first, the three convolution kernels of 5 $*$ 5, 7 $*$ 7, and 5 $*$ 5 are adopted, and the recognition rate was 85.3%. Then, the convolution kernel size was reduced to 5 $*$ 5, and the recognition rate reduced to 69.7%. Next, the recognition rate reached 80.9% when the convolution kernel sizes were 5 $*$ 5 , 9 $*$ 9, and 9 $*$ 9. When the recognition rate reached 86.3%, the convolution kernel sizes were 5 $*$ 5, 11 $*$ 11, and 11 $*$ 11 ; however, when the convolution kernel sizes increase to 11 $*$ 11, 11 $*$ 11, and 9 $*$ 9, the recognition rate decreased. In a word, with the increase in convolution kernel size, the running time increases; the smaller the convolution kernel size, the less the training time, because small convolution kernels have less training parameters and the space complexity and time complexity are reduced. However, when the convolution kernel is too large or too small, the recognition rate will be reduced. When the convolution kernel is too small, it cannot extract the valid local features; when the convolution kernel is too large, the complexity of the extracted feature may be difficult to express by the convolution kernel. In general, small convolution kernel can handle images finely, but it needs more layers to achieve good results; the large convolution 5kernel can extract abstract features of images, but it needs more training parameters.

When the convolution kernels are 5 $*$ 5, 11 $*$ 11, and 11 $*$ 11, the sensitivity is 99.6% and the sensitivity is higher than the specificity. The MCC and F_1 score are consistent with the recognition rate, and the F_1 score reached 0.86 with the choice of the optimal convolution kernel. Therefore, the convolution kernel size should be set reasonably combined with the input image size, and it is very important to improve the performance of CNN but also to protect the CNN parameter tuning. After the discussion of different convolution kernel sizes, it chooses the convolution kernel of 5 $*$ 5, 11 $*$ 11, and 11 $*$ 11 and carries on the following experimental analysis under the high recognition rate.

(2) Different Feature Maps. The number of feature maps is the number of features extracted from each layer, with the same number of convolution kernels per layer. On the premise that the convolution kernel size is invariable, in this paper, we change the number of feature maps on the basis of 6-12-18 and discuss the influence of the extracted feature dimensions on the recognition results; the experimental results are shown in Table 4.

As we can see from Table 4, the number of feature maps is reduced and the running time is reduced, but the recognition rate is not significantly increased. With the increase in the number of feature maps of the third layer, the running time is obviously rising and the recognition rate is also increasing. Although the number of feature maps continues to increase, the number of CT image features is more and the training time is also increasing, and the recognition rate and other evaluation indexes are decreased; especially, the number of features achieves 16-32-200; it took more than two hours, but the recognition rate was only 71.7%. On the whole, with the increase in the number of extracted features, the recognition rate, sensitivity, specificity, MCC, and F_1 score were all increased; when the number of features is 6-12-24, the recognition rate reaches 89.3%. The experimental results show that the number of feature maps of the first layer is less, the number of features in the back layer was increased by 2 times, and it can achieve the highest recognition rate.

Because of the small number of features, the feature description is not sufficient and the large number of feature maps will be overfitting; therefore, we should refer to the size of the data and complexity of the actual sample to select the number of feature maps or convolution kernels and adjust the feature dimension. Generally, using more convolution

Table 4: Effect of feature maps on the experimental results.

Feature maps			Time (s)	Accuracy (%)	Error (%)	Sensitivity (%)	Specificity (%)	MCC	F_1 score
FM1	FM2	FM3							
2	4	8	214.06	86.50	13.50	80.40	92.60	0.74	0.86
3	6	12	443.72	85.40	14.60	81.20	89.60	0.71	0.85
6	12	18	779.64	85.30	14.70	80.80	89.80	0.71	0.85
6	12	24	1161.70	89.30	10.70	92.00	86.60	0.79	0.89
6	12	36	1189.80	86.80	13.20	98.20	75.40	0.76	0.87
6	12	120	1392.19	87.00	13.00	98.60	75.40	0.76	0.87
6	12	150	1463.30	85.10	14.90	78.80	91.40	0.71	0.85
6	12	200	1582.14	86.80	13.20	83.80	89.80	0.74	0.87
6	16	32	2054.51	86.70	13.30	81.60	91.80	0.74	0.87
6	16	120	1812.94	87.90	12.10	93.80	82.00	0.76	0.88
6	16	200	2070.36	85.30	14.70	77.80	92.80	0.71	0.85
16	32	200	8616.49	71.70	28.30	100.00	43.40	0.53	0.69

Table 5: Relationship between the number of layers and recognition performance.

Layer	Convolution and pooling layers	Time (s)	Accuracy (%)	Error (%)	Sensitivity (%)	Specificity (%)	MCC	F_1 score
2	C1-S1	117.76	50.00	—	—	—	—	—
4	C1-S1-C2-S2	485.38	82.30	17.50	99.60	65.40	0.69	0.82
5	C1-S1-C2-S2-C3	1161.70	83.80	16.20	72.40	95.20	0.69	0.84
6	C1-S1-C2-S2-C3-S3	1765.49	89.30	10.70	92.00	86.60	0.79	0.89
7	C1-S1-C2-S2-C3-S3-C4	1779.64	85.40	14.60	97.00	73.80	0.73	0.85
8	C1-S1-C2-S2-C3-S3-C4-S4	1881.06	76.00	24.00	65.60	86.40	0.53	0.76

kernel will get better performance, and the appropriate features maps will certainly be helpful to achieve the desired classification results. Through the discussion of the number of different feature maps, this paper selects the number of feature maps with 6-12-24 and then carries on the following experiment on the premise of ensuring the high recognition rate and the appropriate running time.

(3) Different Network Layers. Although the most essential difference between deep and shallow learning is the number of hidden layers, generally, the more the hidden layers are, the easier it is to learn the underlying features of the image. Aimed at lung CT global features of $64 * 64$ size, the deep convolutional neural network is constructed, and the influence of the DCNN model on the recognition results is discussed by changing the network layer in this experiment. The results of the specific layer assignment and experiment are shown in Table 5; C1 is the first convolutional layer, S1 is the first pooling layer, C2 is the second convolutional layer, and so on.

As we can see from Table 5, with the increase in the number of network layers, from 2 to 8 layers, the recognition rate increased first and then decreased. The recognition rate has only 50% probability with only one convolutional layer and a pooling layer, and it showed that too small number of layers exerts a tremendous influence on recognition result. As the number of layers increases to 6, the recognition rate is up to 89.3%, and then, by adding a convolutional layer, the recognition rate began to decline to 85.4%; the recognition rate is reduced to 76% when constructing the hidden layer of the 8 layers. In general, the deep network structure can promote the reuse of features and get more abstract features with high-level expression; with the increase in network layers, the recognition rate is also increasing; however, there are too many layers in the network, which requires the much convolution and downsampling operations, and the parameters are increasing. In short, the increasing network layers appropriately will ensure the less running time and the higher recognition rate, but too many layers will lead to excess parameters, and the phenomenon of overfitting occurs and the recognition rate reduced. The two indicators MCC and F_1 score have the same change trend with the recognition rate. When the number of hidden layers is 6, the maximum value of the two indicators shows that the recognition efficiency and the fitting effect of the network structure are the best.

3.3.3. Experiment 3: Recognition Based on Different Optimization Methods. Based on the study of the DCNN model structure and model parameters, the optimization of the DCNN model structure for different optimization methods was discussed. Firstly, two kinds of pooling method were analyzed; then, the two activation functions ReLU and sigmoid were analyzed, and finally, two optimization training methods that are the batch gradient descent method and gradient descent method with momentum are compared.

Table 6: Experimental results of different sampling methods.

Feature maps	Convolution kernels	Activation function	Pooling	Iterations/accuracy (%)								Time (s)
				5	10	12	15	20	30	50	100	
6-16-120	5-10-10	Sigmoid	Max	60.08	65.37	79.94	73.75	71.76	70.26	69.86	63.67	2738.62
			Avg	50.00	68.66	76.65	68.06	65.37	61.58	57.49	58.48	2718.68

Table 7: Experimental results of different sampling methods.

Feature maps	Convolution kernels	Activation function	Pooling	Iterations/accuracy (%)								Time (s)
				50	100	140	150	160	200	250	300	
6-16-24	5-10-10	Sigmoid	Max	75.55	83.34	86.23	86.83	84.63	84.13	83.93	85.33	7524.27
			Avg	69.46	81.74	81.24	83.43	80.54	82.14	83.03	83.83	7620.68

Table 8: Experimental results of different activation functions.

Feature maps	Convolution kernels	Pooling	Activation function	Iterations	Accuracy (%)	Time (s)
6-16-120	5-10-10	Max pooling	Sigmoid	150	73.82	543.73
			ReLU	3	72.07	11.37
20-50-500	5-10-10	Max pooling	Sigmoid	8	80.00	4971.42
			ReLU	1	72.82	1043.57

(1) Mean Pooling and Maximum Pooling. The deep convolutional neural network model is composed of two kinds of special hidden layers: the convolutional layer and the pooling layer; however, the pooling layer can reduce the feature dimension, reduce the amount of computation, prevent overfitting, and provide a certain degree of translation and rotation invariance. At present, the commonly used sampling methods are mean pooling and max pooling [20]: mean pooling is the average of the feature points in the neighborhood and max pooling is the maximum value of the feature points in the neighborhood. Two groups of experiments were conducted to discuss the effects of different sampling methods on the final results, the learning rate is 0.0005, and the batch size is 200; the experimental analysis was carried out using two DCNN, and the results are shown in Tables 6 and 7.

The results show that the recognition rate of using the max pooling is higher than that of the mean pooling with the same DCNN model, but the two methods have little effect on the training time. When the number of feature maps is 6-16-120 and the number of iterations is 12, the recognition rate of using the max pooling method is up to 79.94%, but the recognition rate of using the mean pooling method is 76.65%. With the increasing iterations, the recognition rate increased first and declined next based on two methods. When the number of feature maps is reduced to 6-16-24, the convergence rate is slow, and the recognition rate of using max pooling is 86.83% and the recognition rate of using mean pooling is about 83.43%. In conclusion, using max pooling is superior to the mean pooling method for the identification of the CT global features on lung tumors.

According to the theory of pattern recognition, the error of feature extraction mainly comes from two aspects: the first is the increase in the variance of the estimation result caused by the limited size of the neighborhood and the other is the deviation of the estimated mean value caused by the parameter error of the convolutional layer [21]. In general, the mean pooling can reduce the first error and more retain the background information of the image, while the max pooling can reduce the second kinds of error and retain the texture information. In the average sense, it is similar to the mean pooling, and in the local sense, it obeys the principle of max pooling; therefore, we should pay more attention to the ROI of the lung CT image; that is to say, the lesion area is more reserved for the local area; hence, the max pooling is better than the mean pooling for the lung tumor recognition based on global features.

(2) Sigmoid Function and ReLU Function. The activation function can be joined by the nonlinear factor, because the expression ability of the linear model is not enough, and the function is mapped to the specified range by the activation function. There are two kinds of activation functions: sigmoid function and ReLU function [22]. Sigmoid function is one of the most commonly used activation functions; the formula is as follows:

$$S(z) = \frac{1}{1 + \exp(-z)}. \tag{16}$$

The calculation of the ReLU activation function can be greatly reduced, and it is helpful to the characteristic effect; the formula is as follows:

$$R(z) = \max(0, x). \tag{17}$$

Two different DCNN models were selected aimed at different activation functions, and the effects of the two common activation functions on the global feature recognition

Table 9: Effect of batch size on experimental results.

Batch size	Time (s)	Accuracy (%)	Error (%)	Sensitivity (%)	Specificity (%)	MCC	F_1 score
20	1816.22	90.50	9.50	98.80	82.20	0.82	0.90
50	1708.21	91.70	8.30	99.40	84.00	0.84	0.92
100	1619.94	89.90	10.10	97.20	82.60	0.81	0.90
200	1533.37	86.30	13.70	99.60	73.00	0.75	0.86
300	1526.66	85.60	14.40	99.40	71.80	0.74	0.85
500	1508.10	68.40	31.60	100.00	36.8	0.47	0.65

Table 10: Results of the batch gradient descent and gradient descent method with elastic momentum.

Training method	Time (s)	Accuracy (%)	Error (%)	Sensitivity (%)	Specificity (%)	MCC	F_1 score
Batch gradient descent	5809.73	91.70	8.30	99.40	84.00	0.84	0.92
Gradient descent method with elastic momentum	5016.18	96.40	3.60	97.60	95.20	0.93	0.96

of lung tumors were discussed on the basis of the max pooling. The results are shown in Table 8.

It can be seen from the table that when the structure of the DCNN model is unchanged, the sigmoid function is used to achieve the recognition rate of 73.82% when the iteration is 150 times, and the recognition rate reached 72.07% by using the ReLU function in the iteration of 3; compared with the saturation activation function, ReLU has a faster convergence rate and lower training error. Although the recognition rate of using ReLU activation function is not significantly higher than that of the sigmoid function, it converges quickly and the training time is significantly reduced, so we can use the ReLU function to speed up the convergence rate, reduce the training time, and improve the recognition performance.

(3) Batch Gradient Descent Method and Gradient Descent Method with Elastic Momentum. The batch gradient descent method [23] is used to do iterative training and parameter tuning by selecting different batch sizes; in the experiment, the effect of batch size on classification results was discussed, and the effects of two optimization methods batch gradient descent method and gradient descent method with elastic momentum were compared. The experimental results are shown in Table 9.

As we can see from the table, the batch size is closely related to the identification results; the smaller the batch size, the longer the running time, but the recognition rate will continue to increase; when the batch size is too small, the recognition rate will be maintained at a certain level, because when the batch size is too small or too large, training is not enough and the adjustment of the parameters is not enough, so it makes the recognition rate decreased. Therefore, it is necessary to combine the size of the training set and select the appropriate batch size, in order to ensure each parameter adjustment based on adequate training and backpropagation.

It can be seen from Table 10 that the gradient descent method with the elastic momentum is higher than the batch gradient descent method, its recognition rate achieved 96.4%, the sensitivity and specificity were above 95%, and MCC and F_1 score are close to 1. It is indicated that the gradient descent method with elastic momentum is more suitable for lung CT recognition based on DCNN. The gradient descent method with elastic momentum is used to train the network, which reduces the oscillation of the learning process of the neural network, and the network can converge quickly. It can reduce the sensitivity of the local detailed feature for error surface and suppress the network into local minima effectively.

3.4. Conclusions. In the paper, the DCNN is directly used for lung tumor recognition based on CT global features, because of the better feature representation ability of the deep convolutional neural network, and image processing and feature extraction are nonessential. There are three comparison experiments with different model parameters, network structure, and training algorithm; the results verified the feasibility of DCNN for the global characteristics of CT lung tumors, and the experiment results show that the appropriate convolution kernel size, the number of feature maps, and the number of layers of the network can be used to ensure a good recognition performance; being too large or too small will make the feature learning not sufficient for parameter fitting. For lung tumor image recognition, the maximum pooling results are better than the average pooling results; the choice of ReLU activation function can speed up the convergence and reduce the running time; the gradient descent method with momentum not only improves the recognition rate but also makes the recognition rate of DCNN for lung cancer reach 94.6%. Thus, the good feature learning ability and good generalization ability and robustness of the deep convolutional neural network are proven.

In a word, with the deep convolutional neural network, the more the layers, the more the feature maps, the larger the feature space can be represented by the network, and the stronger the feature learning ability of the network. However, the computational complexity is larger, and the phenomenon of overfitting is easy to appear. Therefore, it is necessary to select appropriate layers, the number of feature maps, the convolution kernel size, and other parameters

in the practical application of the specific field; in this way, it can train a better model and ensure relatively little training time.

Data Availability

Some or all data and models during the study are available from the corresponding author on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 62062003).

References

[1] S. Q. Yu and J. H. Peng, "Imaging diagnosis of early lung cancer," *Chinese Journal of Modern Drug Application*, vol. 8, no. 23, pp. 205-206, 2014.

[2] Z. F. Zhang and C. Zhang, "The value of multi-slice spiral CT in the diagnosis of lung cancer," *Practical Clinical Medicine*, vol. 17, no. 1, pp. 53-54, 2016.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[4] X. C. Chen, *Research on Algorithm and Application of Deep Learning based on Convolutional Neural Network*, Zhejiang Gongshang University, 2014.

[5] M. Ma, F. Chen, M. Guo, and Y. L. Chen, "Identification method of street view house number based on improved LeNet-5," *Journal of Yunnan University*, vol. 38, no. 2, pp. 197–203, 2016.

[6] K. Alex, S. Ilya, and H. Geoff, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 1106–1114, 2012.

[7] H. L. Gao, "Military image classification based on convolutional neural network," 2016, http://www.arocmag.com/article/02-2017-10-012.html.

[8] J. Zhang, T. Zhang, Z. L. Yang, X. S. Zhu, and B. X. Yang, "Vehicle model recognition method based on deep convolutional neural network," *Transducer and Microsystem Technologies*, vol. 35, no. 11, pp. 19–22, 2016.

[9] P. Guo, *Deep Convolutional Neural Network and Its Application in Handwritten Chinese Character Recognition*, Sichuan Normal University, 2016.

[10] K. M. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 5353–5360, Boston, Massachusetts, 2015.

[11] K. Gunavathi, P. Sivasubramanian, C. Keerthika, and C. Paramasivam, "A review on convolutional neural network based deep learning methods in gene expression data for disease diagnosis," *Materials Today: Proceedings*, 2020.

[12] Z. Tao, H. Bingqiang, L. Huiling, Y. Zaoli, and S. Hongbin, "NSCR-based DenseNet for lung tumor recognition using chest CT image," *BioMed Research International*, vol. 2020, Article ID 6636321, 9 pages, 2020.

[13] T. Zhou, H. L. Lu, Z. L. Yang, B. Q. Huo, H. L. Ren, and Y. L. Dong, "The ensemble deep learning model for novel COVID-19 on CT images," *Applied Soft Computing*, vol. 98, article 106885, 2021.

[14] W. Chen, M. R. Jiang, W. G. Zhang, and Z. S. Chen, "A novel graph convolutional feature based convolutional neural network for stock trend prediction," *Information Sciences*, vol. 556, pp. 67–94, 2021.

[15] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[16] J. Li, *Automatic Identification Research of Tobacco Diseases based on Convolutional Neural Network*, Shandong Agricultural University, 2016.

[17] D. X. Gong and C. R. Cao, "Plant leaf classification based on CNN," *Computer and Modernization*, vol. 4, pp. 12–19, 2014.

[18] Y. LeCun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[19] T. Zhang, Y. J. Li, and H. H. Hu, "A gender classification model based on cross-connected convolutional neural networks," *Acta Automatica Sinica*, vol. 42, no. 6, pp. 858–865, 2016.

[20] B. Y-Lan, P. Jean, and L. C. Yann, "A theoretical analysis of feature pooling in visual recognition," *International Conference on Machine Learning*, vol. 32, no. 4, pp. 111–118, 2010.

[21] W. J. Liu, X. J. Liang, and H. C. Qu, "Learning performance of convolutional neural networks with different pooling models," *Journal of Image and Graphics*, vol. 21, no. 9, pp. 1178–1190, 2016.

[22] N. Vinod, E. H. Geoffrey, and C. Farabet, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–881, Haifa,Israel, 2010.

[23] D. Silver, A. Huang, C. J. Maddison et al., "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

Hindawi
BioMed Research International
Volume 2021, Article ID 6618918, 9 pages
https://doi.org/10.1155/2021/6618918

Research Article

A Myocardial Segmentation Method Based on Adversarial Learning

Tao Wang[1], **Juanli Wang**[1], **Jia Zhao**[2] and **Yanmin Zhang**[1,3]

[1]*Department of Pediatric Cardiovascular Medicine, Xi'an Children's Hospital, Xi'an 710003, China*
[2]*School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China*
[3]*Shaanxi Provincial Institute for Pediatric Diseases, Xi'an Children's Hospital, Xi'an 710003, China*

Correspondence should be addressed to Yanmin Zhang; ymzh628@126.com

Received 16 October 2020; Revised 9 December 2020; Accepted 2 February 2021; Published 26 February 2021

Academic Editor: Dominique J Monlezun

Congenital heart defects (CHD) are structural imperfections of the heart or large blood vessels that are detected around birth and their symptoms vary wildly, with mild case patients having no obvious symptoms and serious cases being potentially life-threatening. Using cardiovascular magnetic resonance imaging (CMRI) technology to create a patient-specific 3D heart model is an important prerequisite for surgical planning in children with CHD. Manually segmenting 3D images using existing tools is time-consuming and laborious, which greatly hinders the routine clinical application of 3D heart models. Therefore, automatic myocardial segmentation algorithms and related computer-aided diagnosis systems have emerged. Currently, the conventional methods for automatic myocardium segmentation are based on deep learning, rather than on the traditional machine learning method. Better results have been achieved, however, difficulties still exist such as CMRI often has, inconsistent signal strength, low contrast, and indistinguishable thin-walled structures near the atrium, valves, and large blood vessels, leading to challenges in automatic myocardium segmentation. Additionally, the labeling of 3D CMR images is time-consuming and laborious, causing problems in obtaining enough accurately labeled data. To solve the above problems, we proposed to apply the idea of adversarial learning to the problem of myocardial segmentation. Through a discriminant model, some additional supervision information is provided as a guide to further improve the performance of the segmentation model. Experiment results on real-world datasets show that our proposed adversarial learning-based method had improved performance compared with the baseline segmentation model and achieved better results on the automatic myocardium segmentation problem.

1. Introduction

Congenital heart defect (CHD), also known as congenital heart anomaly or congenital heart disease, is a structural defect of the heart or large blood vessels that occurs at birth. Symptoms vary widely, depending on the specific type of defects [1], ranging from mild to life-threatening. Symptoms typically include shortness of breath, bluish to purple skin color, abnormal weight gain, and fatigue. CHD is usually associated with complications of heart failure without causing chest pain, while most CHD are unrelated to other diseases. CHD is the most common birth defect [2]. In 2015, about 48.9 million people globally suffered from CHD [3]. In different countries and regions, CHD affects 4 to 75 cases per 1,000 live births, and moderate or severe problems can occur in 6 to 19 people per 1,000 [1, 4]. CHD is the main cause of death associated with birth defects. Among many types of CHD, the most common involves the inner walls of the heart, valves, or large blood vessels that pump blood into and out of the heart. Some minor defects do not require treatment, but moderate and severe cases can be effectively treated with catheter-based or cardiac surgery. However, many operations are often required, potentially even including heart transplants. Nevertheless, the death rate from CHD can be greatly reduced, given appropriate treatment is provided.

Cardiovascular magnetic resonance imaging (CMRI) is a noninvasive medical imaging technique used to evaluate the function and structure of the cardiovascular system. By using electrocardiographic (ECG) gated control and high time

resolution, regular MRI is adapted to cardiovascular imaging, and its importance is paramount in the evidence-based diagnosis and treatment of cardiovascular diseases [5]. Accurate diagnosis is essential for the development of appropriate treatment regimens for CHD. CMRI can safely provide comprehensive information about CHD, without the use of X-rays or intrusions. This technique is often used in conjunction with other diagnostic techniques, such as echocardiography and diagnostic cardiac catheterization. The use of CMRI for blood pools and myocardium segmentation is a prerequisite for surgical planning and patient-specific heart models for children with complex CHD. The use of existing tools for manual segmentation of 3D images is time-consuming and laborious, which greatly impedes the routine clinical use of 3D heart models. Therefore, automatic myocardial segmentation algorithms and related computer-aided diagnosis systems were developed.

Traditional automatic myocardial segmentation algorithms are generally based on semiautomatic segmentation algorithms. Using prior knowledge, steps such as manual selection of initial contour and initial seed points are automated to realize the automation of the entire cardiac segmentation task. Common myocardial segmentation algorithms include horizontal set segmentation algorithm [6], regional growth segmentation algorithm [7], and threshold segmentation algorithm [8]. However, the segmentation results using this kind of automatic cardiac segmentation algorithms are not ideal, and algorithm robustness is not adequate. With the continuous improvement of hardware equipment and the development of technology, deep learning has been increasingly applied in image processing, resulting in the deep learning-based image segmentation algorithm, surpassing the traditional image segmentation algorithm in many specific tasks [9].

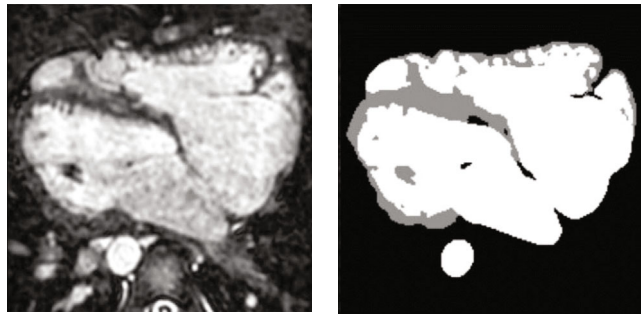
In recent years, with the increase of available data volume and the improvement of computing power, deep learning has made breakthrough progress in various applications in the field of computer vision [10, 11]. Based on these successful experiences, deep learning is now also widely applied in medical image processing [12], including myocardial segmentation. However, common problems in the field of medical image analysis still exist, namely, the low volume of labeled data and networks prone to overfitting. In the field of cardiac segmentation in particular, due to the complex structure of the heart, cardiac labeling is often time-consuming and laborious, which results in the lack of labeled cardiac data. Simultaneously, due to the complex shape of myocardium, myocardium and other surrounding organs and tissues are poorly differentiated in CMR images, and due to the influence of factors such as the tortuous segmentation boundary, room for improvement remains in the final myocardial image segmentation model.

This paper applies the idea of antagonistic learning to the segmentation of myocardium, as an attempt to address these issues. Through a discriminant model, additional supervisory information is given to the segmentation model as a guide to further improve its performance. The myocardial segmentation algorithm based on antagonistic learning is mainly composed of two modules: (1) a segmentation network and (2) discrimination networks. Similarly, to the generation of the maximum and minimum game against the network, the segmentation network accepted the input image and generated the segmentation probability graph. The discriminant network received images and corresponding segmentation results simultaneously and determined whether the input segmentation results came from the segmentation network or from manual annotation. We evaluated the method on the HVSMR2016 dataset and the experimental results showed that our method can achieve good results. An example of the raw image and its segmentation regions is showed in Figure 1.

2. Related Work

2.1. Myocardial Segmentation. Algorithms based on probability models are commonly used to solve the problem of myocardial segmentation among the traditional methods, especially the Gaussian mixture model (GMM) [13, 14]. According to the maximum likelihood (ML) estimation criterion, the expectation maximization (EM) algorithm is usually employed to calculate the parameters in the GMM [15]. On this basis, the Naive Bayes classifier is used to classify each pixel or voxel. Ngo et al. [16] proposed a fully automatic myocardial segmentation method based on depth learning and the level-set algorithm; Mukhopadhyay [17] proposed a fully automatic myocardial segmentation algorithm based on a variational random forest; Tziritas [18] proposed a fully automatic myocardial segmentation algorithm based on the 3D Markov random field; Shahzad et al. [19] proposed a fully automatic myocardial segmentation algorithm that combines the multiple atlas and level-set algorithms. To address the issue of performance and robustness, myocardial segmentation algorithms based on deep learning have been the subject of research. Yu et al. proposed a fully automatic myocardial segmentation algorithm based on 3D fractal convolutional neural networks and dense connection convolutional neural networks [20, 21]. Wolterink et al. [22] proposed a fully automatic myocardial segmentation algorithm based on dilated convolutional neural networks. Avendi et al. [23] applied two deep structures, using convolutional neural networks to automatically detect the left ventricle and a stack automatic encoder to infer the shape of the left myocardium. The inferred shape was then combined into the variability model, to improve the segmentation accuracy. Tran [24] applied fully convolutional network to myocardial MRI segmentation for the first time, extracting ROI regions, and then using the network structure proposed by ROI region pairs to train left and right ventricular segmentation using the stochastic gradient descent (SGD) optimization algorithm. Tao et al. [25] propose a novel shape-transfer GAN for LGE images, which can (1) learn to generate realistic LGE images from bSSFP with the anatomical shape preserved and (2) learn to segment the myocardium of LGE images from these generated images. It is worth to note that no segmentation label of the LGE images is used during this procedure.

2.2. Generative Adversarial Networks (GAN). Generative adversarial networks (GAN), proposed by Goodfellow et al.



(a) Raw MRI image (b) Segmentation regions

Figure 1: An example of the raw image and its segmentation regions.

[26], learn by pitting two neural networks in a zero-sum game with each other. In recent years, GAN have become the most popular learning method of complex probability distribution. They consist of a generator and a discriminator. The goal of the generator is generating samples that are as close to the real data distribution as possible in an attempt to deceive the discriminator, and the goal of the discriminator is to correctly distinguish whether the data belongs to the real distribution or to the generator. The generator and discriminator of conditional generative adversarial networks (CGAN) [27] also use additional condition information, to make the generated data satisfy certain constraints. On the basis of CGAN, Luc et al. [28] used GAN for semantic image segmentation. Xue et al. [29] proposed a novel end-to-end adversarial network architecture called SegAN for MRI image semantic segmentation tasks. Inspired by the original GAN [26], the training process of SegAN is similar to the minimax game, training the segmented network and discriminant network alternately, minimizing and maximizing the objective function, respectively, and combining multi-scale loss in SegAN.

3. Dataset

The dataset used in this experiment was the HVSMR 2016 dataset. This dataset included 20 MR images with various congenital heart defects, where in 10 cases, the image data and their corresponding manual segmentation labeling have been made public (training set). The remaining 10 cases constitute the test set, which did not include manual segmentation tagging, and the segmentation results needed to be submitted to an online test platform that returned the test results.

The images of this dataset were acquired during clinical practice at Boston Children's Hospital, Boston, MA, USA. Some subjects included in the dataset have undergone interventions. Imaging was done in an axial view on a 1.5 T scanner (Phillips Achieva), without contrast agent, using a steady-state free precession (SSFP) pulse sequence. The subjects breathed freely during the scan, and ECG and respiratory gating were used to eliminate the effects of cardiac and respiratory movements for the duration of the imaging. Manual segmentation of the ventricular myocardium was performed by a trained rater and validated by two clinical experts.

There were three classes of labeling: blood pool, myocardial layer, and background. The blood pool class included the left and right atria, left and right ventricles, aorta, pulmonary veins, pulmonary arteries, and the superior and inferior vena cava. The myocardium class included the thick muscle surrounding the two ventricles and the septum between them. Image dimensions and image spacing vary across subjects, with an average of $390 \times 390 \times 165$ pixels and $0.9 \times 0.9 \times 0.85$ mm, respectively, in the full-volume training dataset.

4. Method

4.1. Data Preprocessing. The data processing part of the experiment consisted of two steps: data standardization and random block taking.

4.1.1. Data Standardization. The preprocessing process of the experiment was standardized using the Z-score:

$$x^* = \frac{x - \bar{x}}{s}, \tag{1}$$

where x is the input image, $\bar{x}$ is the average of the gray value of each voxel in the input image, that is, $\bar{x} = 1/n\sum_{i=1}^{n} x_i$, and s is the sample standard deviation of the input image, that is, $s = \sqrt{1/n - 1\sum_{i=1}^{n}(x_i - \bar{x})^2}$. After standardization, the mean value of x^* was 0 and the standard deviation was 1. Data standardization is the most commonly used standardization method and was performed to eliminate the systematic deviation between data as much as possible and be robust to abnormal data values.

4.1.2. Sliding Window Block Taking. Limited by the very small dataset size, data augmentation was a necessary data preprocessing process. In addition, the original 3D image was large in size, therefore, too expensive to input directly into the network for training. Consequently, the method of sliding window block taking was adopted in this experiment. An image block with a size of $64 \times 64 \times 64$ from a standardized input image was extracted along three spatial dimensions with independent uniform distribution, and the corresponding tensor was cut out from the segmentation label according to the corresponding spatial position as the segmentation label of the extracted image block. In order to further expand the size of the training set and to consider

the possible direction invariance caused by the acquisition process of MRI, random 90°, 180° and 270° rotations in the axial plane and symmetric flip about the axial plane were also introduced.

4.2. Overall Framework. The cardiac muscle segmentation algorithm based on adversarial learning consisted mainly of two modules, namely, the segmentation network and the discrimination network. The segmentation network received the input image and generated the segmentation probability map. The discrimination network then received the image and the corresponding segmentation results simultaneously and determined whether the input segmentation result came from the segmentation network or from manual annotation. The outline of the algorithm is shown in Figure 2. The discrimination network can be regarded as a special loss function, different from the commonly used cross-entropy loss-function and dice loss function, which directly depends on the value of each pixel and defines a complete loss function. The discrimination network analyzed the image and segmentation results jointly and had a deep network structure and a large number of learnable parameters, therefore, it was able to provide advanced guidance information for the segmentation network.

For the input 3D image block and its corresponding segmentation result x, y, the segmentation probability map $S(x)$ given by the network was obtained through the segmentation network S, using forward reasoning calculation, and the segmentation loss function J_{seg} and the adversarial loss function J_{adv} were calculated, during training. Similar to the training process of GAN, the segmentation network S and the discrimination network D were trained in turn, and the parameters of the corresponding network model were updated by the back-propagation algorithm.

In the prediction process, the input 3D image was preprocessed and then several image blocks were extracted in a certain step along the three spatial dimensions and input into the segmentation network S, respectively, to obtain the segmentation probability map of the corresponding image blocks. Finally, the segmentation probability maps corresponding to the image blocks at different positions were synthesized, and the segmentation results corresponding to the input 3D images were obtained after postprocessing.

4.3. Segmentation Network. A 3D full convolutional neural network was used to segment the cardiac muscle and blood pool. In theory, the full-convolutional neural network can process input images of any size. However, the input image size is directly related to the size of the characteristic tensor of each layer of the network, demanding a lot of runtime memory for oversized input images. Additionally, because the input image and convolution kernel are 3D tensors, the computational complexity will increase significantly with the increase of input image size. Therefore, the sliding window block strategy of size $64 \times 64 \times 64$ was used for the input image, during both training and testing. The training process blocked the input image at random positions. This step can be seen as a form of data augmentation, which expands the size of the training data set and also creates reasonable constraints on the size of the input image block, so that the network model can complete the training process with limited memory and within reasonable calculation time. The input images were taken along three spatial dimensions with overlapping blocks at equal intervals during the test process, and the extracted image blocks were input into the segmentation model to obtain a segmentation probability map; then, the segmentation probability maps at different positions were divided according to the input image block. The spatial position was arranged, and the overlapping part adopted the voting strategy to average the segmentation probability map and finally obtain the segmentation result of the original image.

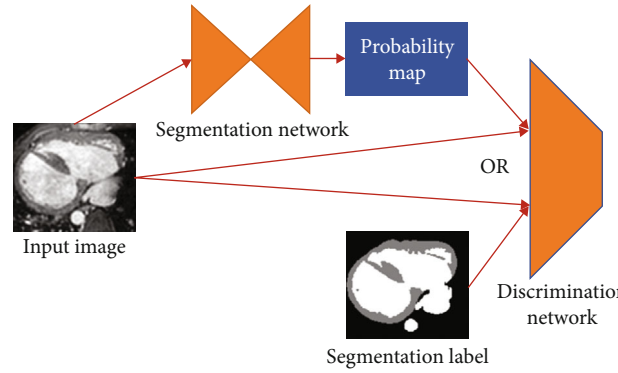


Figure 2: Outline of the cardiac muscle segmentation algorithm based on adversarial learning.

A full-convolutional neural network model with a structure similar to 3D U-Net was designed in this study. As a segmentation network part, the network structure is shown in Figure 3. The network model used a symmetric encoder-decoder structure to extract the characteristics of the input image and obtain the segmentation probability map through forward reasoning calculation. The network used jump connections, connecting the shallow and deep layers of the network, and was able to simultaneously use high-dimensional semantic features and low-dimensional grayscale, texture, and other image detail features to jointly participate in the final segmentation probability map calculation.

Each scale part of the encoder part was composed of two identical stacked modules, with each module including a convolutional layer with a kernel of $3 \times 3 \times 3$, a step size of 1, a batch normalization (BN) layer, and linear rectification function (Rectified Linear Unit, ReLU). Each time the maximum pooling was performed, the spatial scale of the feature map was halved, but the number of feature channels was doubled to retain a certain amount of information. The decoder part was generally symmetrical with the encoder part and had a similar structure. The kernel size and stride of deconvolution are $2 \times 2 \times 2$ and 2, respectively. The input tensor of each scale consisted of the output of the previous layer after deconvolution, while the output features of the encoder of the corresponding size were spliced together. After deconvolution, the spatial scale of features was doubled, and the number of feature channels was halved. Finally, the network used a convolutional layer with an output channel of 3 and the SoftMax activation function to obtain a

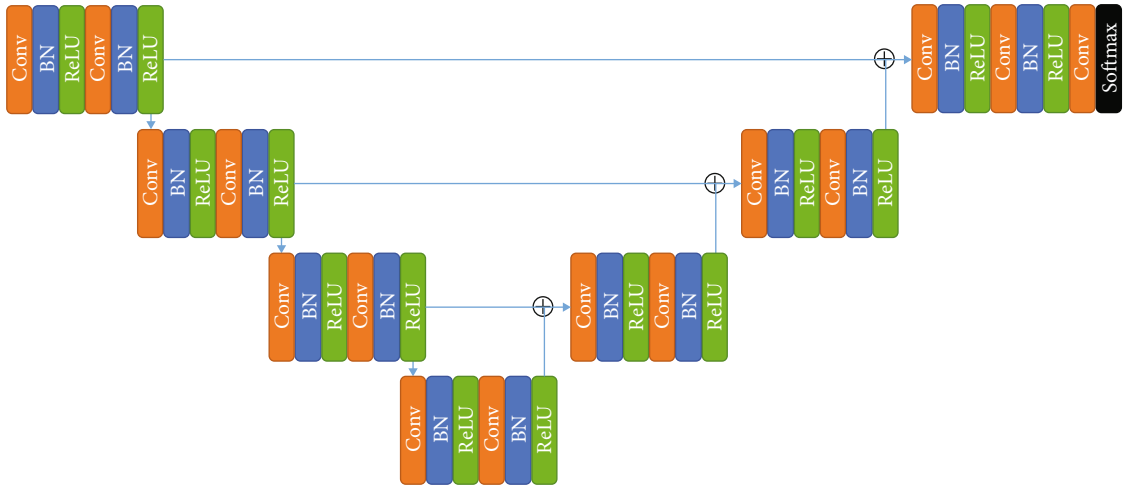


Figure 3: Outline of the segmentation network model.

segmentation probability map. The three channels represent the three category labels of cardiac muscle, blood pool, and background, respectively.

4.4. Discriminant Network. A full-convolutional neural network model is presented in this paper. The number of layers is shallow, and the structure is similar to the VGG network. The network structure is shown in Figure 4. The basic module consisted of a convolutional layer with a step size of two, followed by batch normalization and a linear rectification activation function. The convolutional layer with a step size of two can extract features and reduce the scale of the feature map. The input of the discriminant network was the input image and segmentation results. After the processing of four basic network modules, the discriminate results were obtained by global average pooling and sigmoid activation function. The output value, ranging from 0 to 1, represented the probability that the segmentation result was derived from the manual annotation.

Compared with the segmented network, the discriminant network was shallower, and the number of parameters used is lower. The reason for this design was that the segmentation network was tasked with the relatively complex task of generating segmentation results, which was the main part of the model. Conversely, the sole output of the discriminant network was one probability value, and too many parameters are easy to over fit, which is not conducive to the convergence of the model.

4.5. Loss Function. The most commonly used loss function in image segmentation tasks is the pixel by pixel loss entropy error function. The value of the loss function on each pixel (voxel) was calculated independently. The pixel classification prediction was compared with the standard vector encoded by one-hot to measure the difference between them. The calculation formula of the cross-entropy loss function is shown in equation (2):

$$J_{\mathrm{CE}} = -\frac{1}{N}\sum_{\text{classes}}\sum_{i=1}^{N} g_i \ln p_i, \tag{2}$$

where N is the total number of pixels (voxels), classes represent each category, g_i denotes whether the i-th pixel is marked as the true label of the current category, and p_i is the prediction probability that the i-th pixel is predicted as the current category. It is clear from the formula that the cross-entropy loss function was evaluated separately for each pixel and then the contribution of all pixels was averaged to obtain the final loss value. The segmentation network model combined with cross entropy loss function was the basic method used to address the image segmentation problem by deep learning, and it was also the baseline method of the experiment in this paper.

Adversarial loss function, J_{adv} is a minimum-maximization function, defined as equation (3):

$$\min_S \max_D J_{adv}(S, D) = E_{x,y\sim P_{data}(x,y)}\{\log D(x, y) - \log D[x, S(x)]\}, \tag{3}$$

where S and D are the segmentation network and the discrimination network, respectively, X and Y are the input image block and the corresponding segmentation result annotation, respectively, P_{data} is the data distribution composed of the training data set, and $D(x, y)$ is the prediction probability of the segmentation result corresponding to X, determined by the segmentation network. When the parameters of the segmentation network were fixed, the discrimination network minimized the binary cross entropy loss function. When the parameters of the discriminant network were fixed, the discrimination network was minimized as follows:

$$-E_{X,Y\sim P_{\mathrm{data}}(x,y)} \log D[x, S(x)]. \tag{4}$$

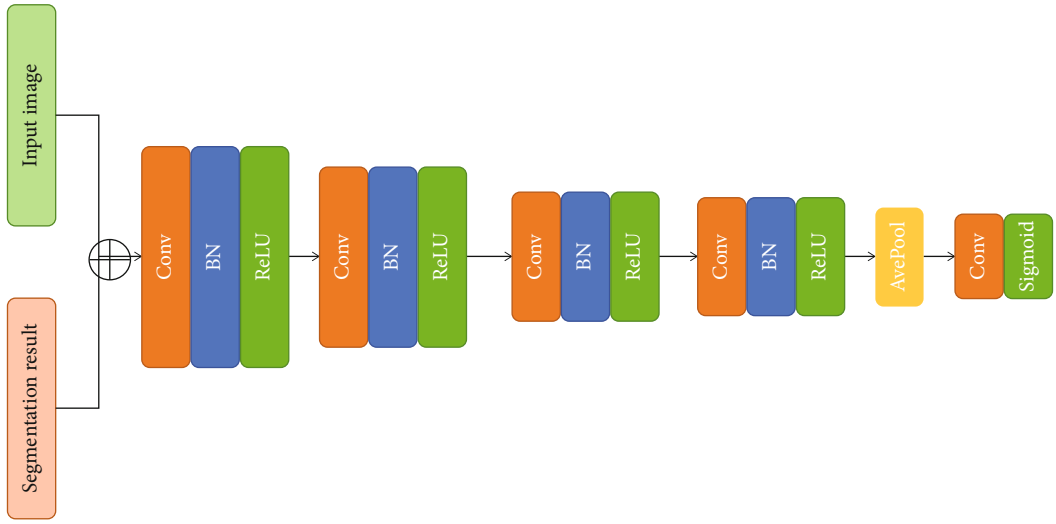


Figure 4: Outline of the discrimination network model.

In other words, the segmentation network was induced to produce a more realistic segmentation result.

The cross-entropy loss could effectively measure the difference between the classification prediction value of each pixel and the gold standard, while the counter loss function could comprehensively measure the difference between the predicted image segmentation results and the gold standard from a global perspective, complementing each other. The two loss functions were used at the same time in this study, in order to utilize both their advantages, and the total loss function was defined as shown in equation (4):

$$J = J_{CE} + \alpha J_{adv}, \tag{5}$$

where α is a super parameter used to adjust the relative weight of the above two loss functions. Larger values of α lead to larger relative weight of J_{adv}, which also cause the influence of the adversarial network on segmentation results to be more explicit. On the other hand, smaller α values lead to larger relative weight of J_{CE}, causing the influence of adversarial network on segmentation results to be less explicit. Experimental results showed that when α was set as 0.15, the overall performance of the model is optimal. The related parameters are discussed in detail in the discussion part.

4.6. Evaluation Index and Implementation Details. The commonly used image segmentation task evaluation index Dice coefficient (DSC) was used to evaluate the performance of myocardial and blood pool segmentation. The definition of DSC is as follows:

$$\mathrm{DSC} = \frac{2|X \bigcap Y|}{|X| + |Y|}, \tag{6}$$

where X and Y are the predicted segmentation result and the manually annotated segmentation result, respectively. DSC is a dimensionless number between 0 and 1 that measures the similarity of two sets. High DSC values are associated with close match between the predicted segmentation result of the model and that of manual annotation, meaning better model performance.

Table 1: Ablation study experimental results of the discriminative network.

Model	Myocardial dice	Blood pool dice
3D UNet (baseline)	0.712	0.926
3D UNet + discrimination network	0.753	0.929

The experiment was based on the deep learning framework Keras. The Adam adaptive optimization algorithm was used to complete the training and testing of the network model, using an NVIDIA 1080Ti GPU hardware platform. The network was trained on the HVSMR 2016 dataset. The Leave-One-Out scheme was used in the study, since there were only 10 samples in the training set. One sample was selected as the validation set, and the remaining nine samples were selected as the training set that the 10-fold cross-validation experiment was performed in turn to verify the effectiveness of the proposed method and the discussion experiment of super parameter. The final model was then tested online with the complete training set and the optimized hyperparameter training. The learning rate is 0.001, and the batch size is 16.

5. Experiments and Results

We demonstrated the myocardial segmentation algorithm based on adversarial learning and analyzed its effectiveness by conducting ablation experiments. The training datasets of the HVSMR 2016 datasets were used to conduct cross-validation using the leave-one-out method and use the average value of the Dice score to evaluate the performance of the model. The experimental results are shown in Table 1.

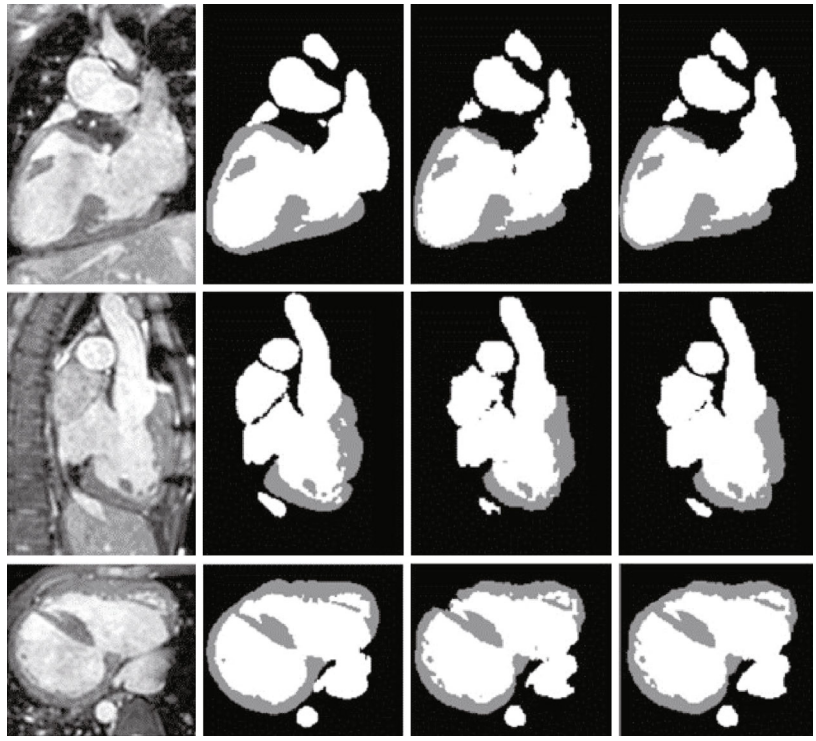


Figure 5: Comparison of segmentation results on the validation set. Each row corresponds to a different case sample. The first column is the CMR image slice of the case sample, the second column is the manually marked segmentation results, and the third column is the baseline model. The fourth column is the segmentation result of the complete model.

The segmentation network shown in Figure 2 was used as the baseline model on the HVSMR 2016 datasets. The discrimination network, as shown in Figure 3, was then added. We conducted relevant experiments again to verify the effectiveness of the discriminative network. Comparing the experimental results in Table 1, it is reasonable to conclude that after using the discriminative network and introducing the adversarial learning mechanism, the performance of the network model in myocardial segmentation was considerably improved and the Dice score increased from 0.712 to 0.753. The improvement of the blood pool segmentation was very small, because the blood pool has a simple shape and no internal texture and structure. The blood pool was also relatively easy to segment compared to the complex-shaped myocardium with a thin layer structure. The baseline model achieved good results, and improved space was relatively small. Figure 5 shows the segmentation results of the learning-based segmentation of the partial validation set. It can be seen that the network model that introduced the adversarial learning mechanism gained a better segmentation result than the baseline model, thereby achieving more accurate myocardial segmentation.

The network model trained on the HVSMR 2016 datasets was tested online on the test datasets and compared with other methods published in recent years. These methods are mainly divided into two categories based on traditional machine learning and deep learning, as shown in Table 2.

Table 2: Quantitative comparison of the method presented in this paper to segmentation performance of different methods from the literature.

Method	Myocardial dice	Blood pool dice
Mukhopadhyay [17]	0.495	0.794
Tziritas [18]	0.612	0.867
Shahzad [19]	0.747	0.885
3D UNet [30]	0.707	0.926
Ours	0.762	0.928
Wolterink et al. [22]	0.802	0.926
Yu et al. [20]	0.786	0.931

The traditional machine learning algorithms included a variation random forest algorithm proposed by Mukhopadhyay [17], a 3-D MRF model random field proposed by Tziritas [18], and methods combining multiatlases and level sets proposed by Shahzad et al. [19]. Limited to the characteristics of manual design, the overall performance was slightly worse than the methods based on deep learning. However, in terms of deep learning models, the 3D FractalNet proposed by Yu et al. introduced the idea of recursion, with the network model being a complex fractal structure [20]. Wolterink et al. proposed a convolutional neural network model with two-dimensional holes [22] that combines the

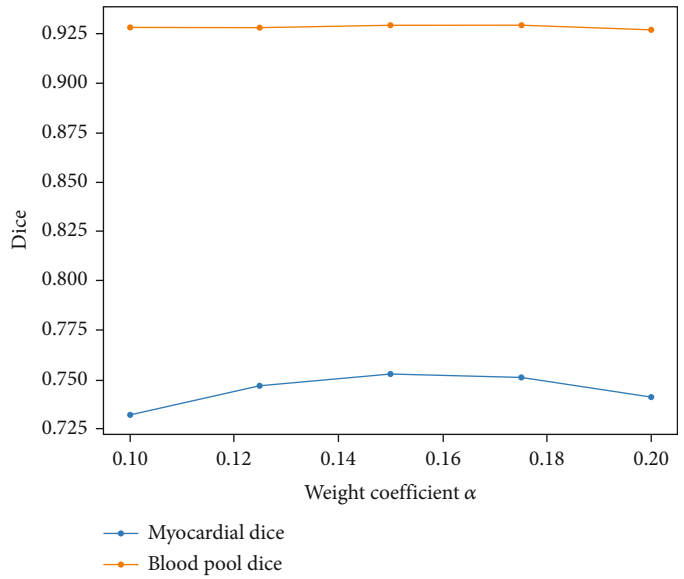


Figure 6: The influence of weight coefficient α on model performance for the myocardial layer and blood pool.

ideas of multiple perspectives. 3D UNet has a relatively simple structure and a wide range of applications. Even though a difference in performance was still present between the method of 3D UNet + discrimination network proposed in this paper and the optimal method, compared with the baseline model, the performance was notably improved.

6. Discussion

The influence of the discriminator on the final segmentation result was affected by using different loss function weight coefficients α, further affecting the final average Dice coefficient. In the 10-fold cross-validation experiment conducted on the training datasets, the α value was adjusted depending on the average dice coefficient. The performance indexes of different α values on the validation datasets are shown in Figure 6. Experiments showed that the model achieves best performance when $\alpha = 0.15$.

7. Conclusion

A myocardial segmentation algorithm based on adversarial learning was proposed in this paper, and experiments were designed to comparatively analyze the effectiveness of the adversarial learning mechanism on myocardial tissue segmentation tasks. The introduction of adversarial learning mechanism for model focus on the overall spatial structure and context consistency was successful, and a more accurate segmentation result was obtained. Our method improved the quantitative segmentation performance index considerably, compared with the baseline model.

Data Availability

The dataset used in this experiment was the HVSMR 2016 dataset, which is available at: http://segchd.csail.mit.edu/data.html

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 81974014 and 81470452.

References

[1] S. Mendis, P. Puska, and B. Norrving, *Global Atlas on Cardiovascular Disease Prevention and Control*, World Health Organization, Geneva, 2011.

[2] T. Vos, R. M. Barber, B. Bell et al., "Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013," *The Lancet*, vol. 386, no. 9995, pp. 743–800, 2015.

[3] T. Vos, C. Allen, M. Arora et al., "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," *The Lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016.

[4] A. Milunsky and J. M. Milunsky, *Genetic Disorders and the Fetus: Diagnosis, Prevention, and Treatment*, John Wiley & Sons, 2015.

[5] V. K. B. Florian, G. Pilz, and J. Schulz-Menger, "Representation of cardiovascular magnetic resonance in the AHA/ACC guidelines," *Journal of Cardiovascular Magnetic Resonance*, vol. 19, no. 1, 2017.

[6] S. O. R. Fedkiw and S. Osher, "Level set methods and dynamic implicit surfaces," *Applied Mathematical Sciences*, vol. 153, no. 3, 2003.

[7] Song Chun Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image

segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, 1996.

[8] S. S. Al-Amri and N. V. Kalyankar, "Image segmentation by using threshold techniques," 2010, https://arxiv.org/abs/1005.4020.

[9] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 9, pp. 60–88, 2017.

[10] M. Yang, Y. Liu, and Z. You, "The Euclidean embedding learning based on convolutional neural network for stereo matching," *Neurocomputing*, vol. 267, pp. 195–200, 2017.

[11] J. Yu, X. Yang, F. Gao, and D. Tao, "Deep multimodal distance metric learning using click constraints for image ranking," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4014–4024, 2017.

[12] 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.

[13] D. Feng, L. Tierney, and V. Magnotta, "MRI tissue classification using high-resolution Bayesian hidden Markov normal mixture models," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 102–119, 2012.

[14] Z. Ji, Y. Xia, Q. Sun, Q. Chen, D. Xia, and D. D. Feng, "Fuzzy local Gaussian mixture model for brain MR image segmentation," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 339–347, 2012.

[15] D. Karlis and E. Xekalaki, "Choosing initial values for the EM algorithm for finite mixtures," *Computational Statistics & Data Analysis*, vol. 41, no. 3-4, pp. 577–590, 2003.

[16] T. A. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Medical Image Analysis*, vol. 35, pp. 159–171, 2017.

[17] A. Mukhopadhyay, "Total variation random forest: fully automatic MRI segmentation in congenital heart diseases," *Lecture Notes in Computer Science*, vol. 10129, pp. 165–171, 2017.

[18] G. Tziritas, "Fully-automatic segmentation of cardiac images using 3-D MRF model optimization and substructures tracking," *Lecture Notes in Computer Science*, vol. 10129, pp. 129–136, 2017.

[19] R. Shahzad, S. Gao, Q. Tao, O. Dzyubachyk, and R. van der Geest, "Automated cardiovascular segmentation in patients with congenital heart disease from 3D CMR scans: combining multi-atlases and level-sets," *Lecture Notes in Computer Science*, vol. 10129, pp. 147–155, 2017.

[20] L. Yu, X. Yang, J. Qin, and P. A. Heng, "3D FractalNet: dense volumetric segmentation for cardiovascular MRI volumes," *Lecture Notes in Computer Science*, vol. 10129, pp. 103–110, 2017.

[21] L. Yu, J.-Z. Cheng, Q. Dou et al., "Automatic 3D cardiovascular MR segmentation with densely connected volumetric convnets," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017: 20th International Conference*, pp. 287–295, Quebec City, QC, Canada, 2017.

[22] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. Reconstruction," in *Reconstruction, Segmentation, and Analysis of Medical Images*, vol. 10129, pp. 95–102, Springer, Cham, 2017.

[23] M. R. Avendi, A. Kheradvar, and H. A. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Medical Image Analysis*, vol. 30, pp. 108–119, 2016.

[24] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis MRI," 2016, https://arxiv.org/abs/1604.00494.

[25] X. Tao, H. Wei, W. Xue, and D. Ni, "Segmentation of multimodal myocardial images using shape-transfer GAN," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 271–279, Springer, Cham, 2020.

[26] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," 2014, https://arxiv.org/abs/1406.2661.

[27] D. Michelsanti and Z. H. Tan, "Conditional generative adversarial networks for speech enhancement and noise-robust speaker verification," 2017, https://arxiv.org/abs/1709.01703.

[28] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," 2016, https://arxiv.org/abs/1611.08408.

[29] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: adversarial network with multi-scale L_1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, no. 3-4, pp. 383–392, 2018.

[30] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, vol. 9901, pp. 424–432, Springer, Cham, 2016.

Hindawi
BioMed Research International
Volume 2021, Article ID 8824395, 18 pages
https://doi.org/10.1155/2021/8824395

Research Article

A New Robust Adaptive Fusion Method for Double-Modality Medical Image PET/CT

Tao Zhou[1,2] **Huiling Lu**[3] **Fuyuan Hu**[4] **Hongbin Shi**[5] **Shi Qiu**[6] **and Huiqun Wang**[7]

[1]*School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China*
[2]*Key Laboratory of Images & Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China*
[3]*School of Science, Ningxia Medical University, Yinchuan 750004, China*
[4]*School of Electronic & Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China*
[5]*Department of Urology, The General Hospital of Ningxia Medical University, Yinchuan 750004, China*
[6]*Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China*
[7]*School of Public Health and Management, Ningxia Medical University, Yinchuan 750004, China*

Correspondence should be addressed to Tao Zhou; zhoutaonxmu@126.com and Huiling Lu; lu_huiling@163.com

Received 13 September 2020; Revised 24 October 2020; Accepted 12 January 2021; Published 4 February 2021

Academic Editor: Andrea Scribante

A new robust adaptive fusion method for double-modality medical image PET/CT is proposed according to the Piella framework. The algorithm consists of the following three steps. Firstly, the registered PET and CT images are decomposed using the nonsubsampled contourlet transform (NSCT). Secondly, in order to highlight the lesions of the low-frequency image, low-frequency components are fused by pulse-coupled neural network (PCNN) that has a higher sensitivity to featured area with low intensities. With regard to high-frequency subbands, the Gauss random matrix is used for compression measurements, histogram distance between the every two corresponding subblocks of high coefficient is employed as match measure, and regional energy is used as activity measure. The fusion factor d is then calculated by using the match measure and the activity measure. The high-frequency measurement value is fused according to the fusion factor, and high-frequency fusion image is reconstructed by using the orthogonal matching pursuit algorithm of the high-frequency measurement after fusion. Thirdly, the final image is acquired through the NSCT inverse transformation of the low-frequency fusion image and the reconstructed high-frequency fusion image. To validate the proposed algorithm, four comparative experiments were performed: comparative experiment with other image fusion algorithms, comparison of different activity measures, different match measures, and PET/CT fusion results of lung cancer (20 groups). The experimental results showed that the proposed algorithm could better retain and show the lesion information, and is superior to other fusion algorithms based on both the subjective and objective evaluations.

1. Introduction

The main purpose of medical image fusion is to generate a composite image by integrating the complementary information from multiple medical source images of the same scene [1]. Molecular images and anatomical images are integrated by PET/CT fusion; the fused image contains information on the pathophysiology of different modality images and improves the identifiability of the lesion areas. It not only provides images for the differential diagnosis of benign or malignant lesions and the detection rate of local space-occupying lesions but also carries out whole body imaging in tumor exploration. Medical image fusion plays an important role in clinical applications such as image-guided radiotherapy, image-guided surgical procedure, and noninvasive diagnosis, thereby helping the diagnosis and differential diagnosis, treatment planning, therapeutic monitoring, and prognostic evaluation of many serious diseases [2]. In Reference

[3], by analyzing the Piella framework and multiscale analysis theory, four construction methods of pixel level fusion rules are presented on the basis of the Piella framework; a self-adaption fusion algorithm of lung cancer PET/CT based on Piella frame and DT-CWT is proposed in first fusion path.

The general framework of multiresolution image fusion method was firstly proposed by Zhang and Blum [4]. On this basis, Piella [5] has been developing and extending the framework by categorizing the key technology technologies multiresolution image fusion method into two parts, which includes multiresolution transform and fusion rule, making the multiresolution image fusion method more systematic and standardized. At present, the research on PET/CT image fusion method can be divided into two aspects. One aspect is multiresolution transformation; the fusion schemes based on multiresolution transform are widely used in practical applications. The commonly used methods based on wavelet transform are to overcome the limitations of spectral distortion, but they can only obtain limited directional information. Novel multiresolution transform-based approaches are proposed, such as ridgelet transform [6], curvelet transform [7], bandlet transform [8], contourlet transform [9], nonsubsampled contourlet transform (NSCT) [10], and shearlet transform [11]. The medical image fusion algorithm based on weighted contourlet transformation coefficient weighting is studied in [12]. In addition, the medical image fusions of NSCT transform and contourlet transform are compared [13] and the experimental results show that the NSCT transform can improve the contrast of the fused image in the fusion process. The other is the design of the fusion rule based on the Piella framework; the purposes of which are to explore how to construct the match measure and the activity measure by improving and optimizing the traditional fusion rule [14, 15].

In recent years, researchers have proposed many new methods of image fusion [16–19], such as CT image feature-level fusion with rough sets using in pulmonary nodule detection [20], GA-SVM [21], COVID-19 on CT images [22], high-dimensional feature reduction based on variable precision rough set, and genetic algorithm in medical image [23]. Fusion method based on compressed sensing domain is also emerging in recent years to solve the high time complexity in the process of medical image fusion. The compressed sensing theory is applied to image fusion firstly by Wan and Canagarajah [24]. Luo et al. [25] proposed the classification-based image fusion method, with the data similarity as the adjustment term of the weights in fusion process, the standard of measuring the energy of the original image by mean observed value of observation of the fusion rule. In 2011, the superiority and effectiveness of compressed sensing theory in image fusion applications are demonstrated [26]. Medical image fusion based on compressed sensing can improve the time and space efficiency in the network transmission process and provide technical support for mobile medical services and medical treatment.

The characteristics of medical image fusion based on compression perception domain are as follows. Firstly, it is not a simple fusion based on pixel, but a small amount of sampling data fusion processing. Secondly, the characteristics of the compression measurement are different from the traditional transformation coefficient and the fusion rules are designed to be adapted. Thirdly, the time and space efficiency of image fusion is improved. Fourthly, the redundant information between different source images is reduced.

In general, specific integration framework can be summarized as follows. Firstly, the source image is transformed with the appropriate sparse representation methods. Secondly, the sparse coefficients are sampled by using the compression measurement matrix, according to the characteristics of observed value and designed fusion rules. Thirdly, the fused image is reconstructed by performing the corresponding inverse transform over the merged coefficients. The fusion process is shown in Figure 1.

In this paper, we propose a self-adaptive fusion algorithm of PET/CT based on compressed sensing and histogram distance as described in the Piella framework. The fusion rule of low-frequency coefficients in the NSCT transform domain is calculated by using the pulse-coupled neural network (PCNN) method [27]. For high-frequency images, the Gauss random matrix is used for compression measurement after NSCT transform of subbands, the regional energy is measured as the activity measure, and the histogram distance between subbands in eight directions is calculated as the matching measure. The fusion factor d is calculated by using the match measure and the activity measure. According to the fusion factor, the high-frequency measurement value is fused, and the high-frequency fusion image is reconstructed by using the orthogonal matching pursuit algorithm of the high-frequency measurement after fusion. By combining NSCT and compressed sensing, high-frequency subbands with sparsity are obtained after NSCT transform and the fused image can be reconstructed by a few of the observed data extracted from a large number of redundant data generated from the multiresolution decomposition. Experimental results show that the algorithm reduces the workload of high-frequency signal sampling and improves the image contrast. In addition, the fused image has good visual effect and can be improved by the objective index.

2. Method and Material

2.1. The Piella Framework. The multiresolution image fusion framework of Piella is shown in Figure 2. The two original images A and B are decomposed using the multiresolution, which has been transitioned gradually from the conventional pyramid decomposition to the wavelet transform and curvelet transform. Decomposition coefficient C_I (I = A, B) is obtained by multiresolution transform, and the fusion rules of the decomposition coefficient are summarized as four modules outlined in the dashed box in Figure 2: match measure, activity measure, decision module, and combination module. The match measure M_I (I = A, B) is used to measure the matching and similarity between two original images (I = A, B); the activity measure a_I (I = A, B) is used to extract the feature information and highlight different parts. The activity and match measures are used in the decision module, and the degree of similarity and feature information of the decomposition coefficient are obtained by the decision factor d. The

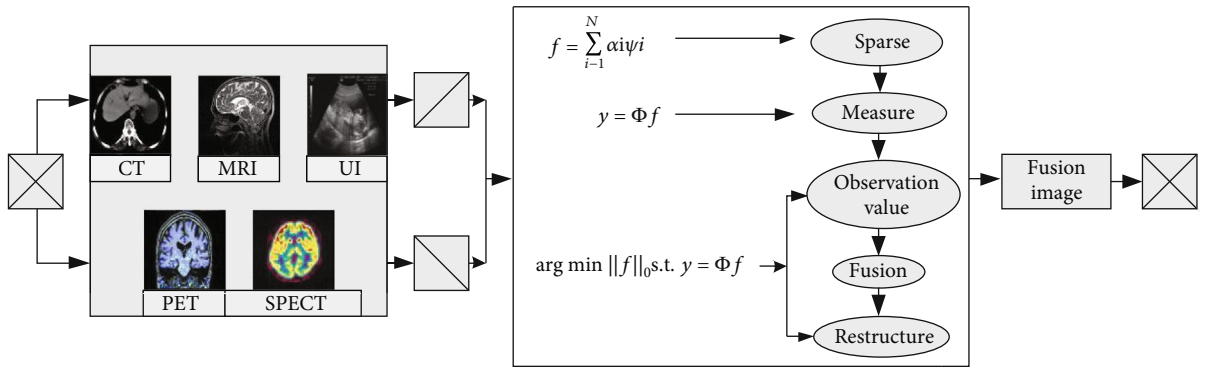


Figure 1: Medical image fusion based on compressed sensing.

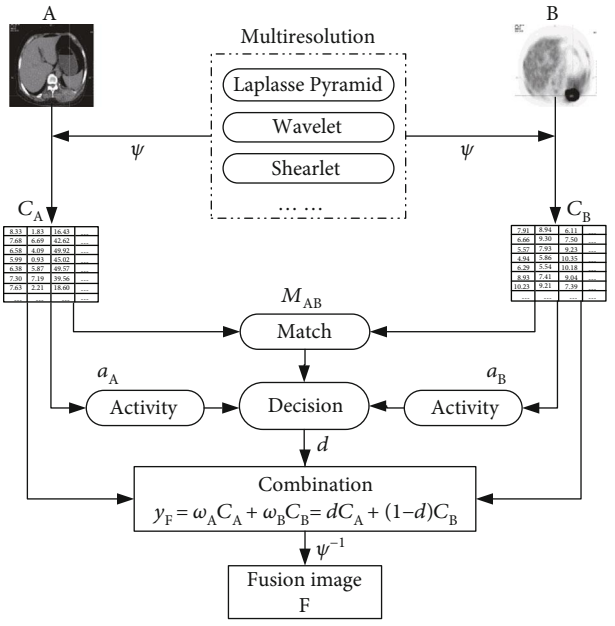


Figure 2: Multiresolution image fusion framework of Piella.

decomposition coefficient y_F of the fusion image is obtained by synthesizing the image decomposition coefficients according to the decision factor d, and the fusion image F is obtained by multiresolution inverse transform of y_F.

2.2. Image Fusion Based on NSCT and Compressed Sensing. NSCT was proposed by Cunha in 2006. It can be constructed through double filter banks approach, nonsubsampled pyramid structure (NSP), and nonsubsampled directional filter banks (NSDFB) [28]. The source image is filtered firstly by passing through NSP filter to get $K + 1$ subband images with equal size to the source images assuming that the NSP decomposition is with K level. After NSP decomposition, the obtained high-frequency subbands are subjected to transformation in direction at I level by passing through filter composed of iterated two-channel NSDFB to generate 2^I high-frequency images. Therefore, a low-frequency subband $\sum_{K=1}^{I} 2^{I_K}$ and high-frequency subbands can be obtained after the NSCT decomposition.

In the compressed sensing theory, the original signal can be recovered by a small amount of observation data, which are applied to the image to extend the one-dimensional signal to the operation of the two-dimensional matrix. Assuming that the observation matrix $\Phi \in R^{M \times N}$ is used to measure the image $f \in R^{N \times N}$, the observation vector $y \in R^M$ is obtained.

$$y = \Phi f. \tag{1}$$

In this process, the image data from the N dimension are reduced to the observation data of M dimension, and the compressed sampling is realized [29]. However, the prerequisite for data compression is to satisfy the prior condition of sparsity, i.e., the sparse representation can be obtained by orthogonal basis transformation or tight frame transformation.

$$f = \psi a, \tag{2}$$

where α is the representation of image in the Ψ domain in Equation (2).

If the nonzero K of α is much smaller than N, that image is sparse and α is the image sparse coefficient in the Ψ domain. In this paper, we take the NSCT as the sparse basis of original image. Equation (3) can be obtained by transformation of Equations (1) and (2).

$$y = \Phi f = \Phi \psi a = \Theta a. \tag{3}$$

Among them, $\Theta(M \times N)$ is the sensing matrix; the number of equations is far less than the number of unknowns ($M << N$); there is no determined solution for the equation. Since the signal is K sparse ($K << M$), the uniqueness of the solution can be warranted if the sensing matrix satisfies the restricted isometry property (RIP). For any given signal f with K sparse and constant δ_K, the following expression should be met.

$$(1 - \delta_K)\|f\|_2 \le \|\Theta f\|_2 \le (1 + \delta_K)\|f\|_2. \tag{4}$$

In Equation (4), constant $\delta_K \in (0, 1)$ is known as the RIP constant [30]. The image is subjected to sparse transformation and measurement matrix, and the fusion rule is designed

to fuse the compressed measurement value. The high-dimensional image is restored from the fused measurement value using reconstruction algorithm. The above process is transformed into a minimum L_1 norm problem and expressed mathematically as follows:

$$f^* = \arg\min \|f\|_1 \text{s.t.} y = \Phi f. \tag{5}$$

This is a convex optimization problem in mathematics [31]. A convex relaxation algorithm can be used to solve the l_1 norm optimization problem and total variation (TV) optimization. In addition, the signal can be reconstructed by other methods, such as relaxation of l_0 norm to l_P norm followed by optimization problem solving, or reconstruction of image using Bayesian method on the basis of introducing sparsity by prior distribution.

Taken together, we can get one low-frequency subband and some high-frequency subbands after NSCT transform and different image fusion methods should be applied to high- or low-frequency subbands. Since high-frequency subbands of NSCT transform usually contain multidirectional image information, a large amount of computations are required in the process of fusion, thereby making the process time-consuming and ineffective. In contrast, the combination of NSCT and compressed sensing can significantly reduce the amount of computation and space of data storage in image fusion.

2.3. Self-Adaptive Fusion Algorithm of PET/CT Based on Compressed Sensing and Histogram Distance

2.3.1. Algorithm Idea. By analyzing the features of PET and CT images, an adaptive fusion algorithm of PET/CT fusion based on compressed sensing and histogram distance is proposed. The main steps of the algorithm are as follows:

First step: monolayer NSCT transform of PET and CT source images is registered (PET for image A, CT for image B) and a low-frequency subband L_I (I = A and B) and eight high-frequency subbands H_I^ξ(I = A, B, ζ is the direction number: $\zeta = 1, 2, \cdots, 8$) in different directions are obtained.

Second step: L_I mainly contains the low-frequency signal with poor sparsity. In this paper, the fusion rule of the PCNN is used to fuse the low-frequency L_I and get fusion image L_F since PCNN has high sensitivity to the featured region of an image.

Third step: H_I^ξmainly contains the detailed information of the original image with higher sparsity, and a higher reconstruction accuracy can be obtained by compressive sampling. Therefore, Gauss random matrix Φ is used for compression measurement of H_I^ξ to get the measured value Y_I.

Fourth step: the fusion rule of H_I^ξ is determined according to that used for the Piella framework:

(1) Match measure: H_I^ξ is divided into blocks and the histogram distance between the blocks is calculated, getting the match measure M_{AB}

(2) Activity measure: the regional energy of H_I^ξ is calculated and used as the activity measure a_I

(3) Decision module: the fusion factor d of the self-adaptive decision model is calculated using M_{AB} and a_I

(4) Combination module: the measured value Y_I is fused based on fusion factor d and the fusion measurement value Y_F is obtained. The high-frequency image H_F is then reconstructed using the orthogonal matching pursuit algorithm

Fifth step: the final fusion image F is obtained by NSCT inverse transform of L_F and H_F. The framework of fusion algorithm is shown in Figure 3.

2.3.2. Lowpass Subband Fusion Rule. The gray levels of the PET and CT images are different and usually of mutually exclusive property since the imaging mechanisms are different between PET and CT. Therefore, a PET scan shows the metabolically active malignant lesion as a dark spot, while CT scan provides detailed images of bones and organs inside the body. The low-frequency image of the source image obtained by the NSCT transform mainly contains the approximate components of the source image with very low sparsity. If the random measurement matrix is used for compressive sampling, the reconstruction accuracy of the fusion of low-frequency subbands will be affected. Since the low-frequency subband image mainly shows the background information, the fusion rule of PCNN based on the fact that human visual system is more sensitive to the featured objects or regions is employed to highlight the lesions in the whole image.

The PCNN of a single neuron is composed of a receiving section, a modulation section, and a pulse generator as shown in the following [29]:

$$\begin{cases} F_{pq}(n) = e^{-a_F} F_{pq}(n-1) + V_F \sum_{kl} M_{pqkl} Y_{kl}(n-1) + S_{pq}, \\ I_{pq}(n) = e^{-a_I} I_{pq}(n-1) + V_I \sum_{kl} W_{pqkl} Y_{kl}(n-1), \\ D_{pq}(n) = e^{-a_D} D_{pq}(n-1) + V_D O_{pq}(n), \\ U_{pq}(n) = F_{pq}(n)\left(1 + \beta I_{pq}(n)\right), \\ O_{pq}(n) = \begin{cases} 1, & U_{pq}(n) > D_{pq}(n-1), \\ 0, & \text{other}, \end{cases} \end{cases} \tag{6}$$

where F_{pq} is the feedback input, I_{pq} is the link input, D_{pq} is the dynamic threshold, U_{pq} is the internal activity term, O_{pq} is the pulse output, pq is the neuron label, and n is the number of iterations. S_{pq} is the external stimulus, W_{pqkl} and M_{pqkl} are the synaptic connection weights between neurons, i.e., the strength of the connection between the feedback domain in neuron p and the input domain in neuron q. α_F, α_L, and α_D are the time decay constants; V_F, V_L, and V_D

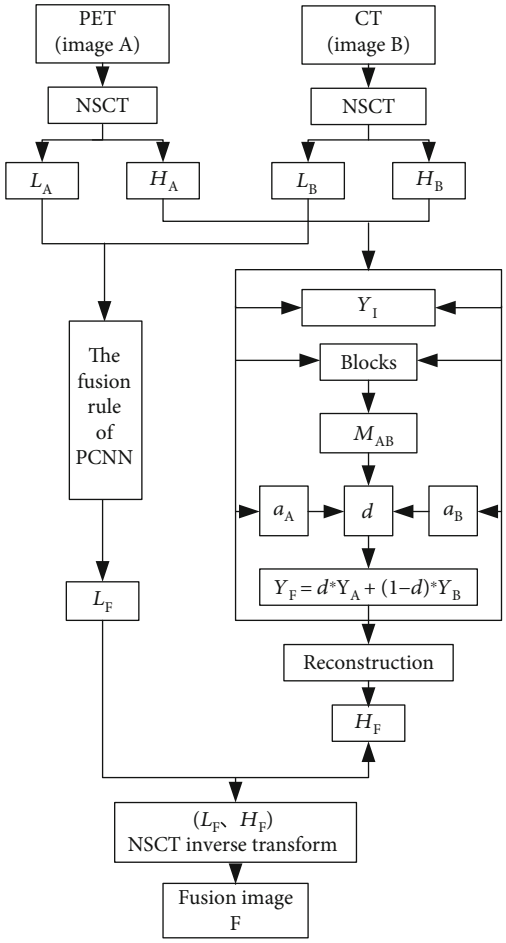


Figure 3: The framework of the self-adaptive fusion algorithm of PET/CT based on compressed sensing and histogram distance.

are the coefficients of method, and β is the coefficient of internal active connection.

The specific steps for applying PCNN to the low-frequency image fusion are as follows:

(1) The low-frequency coefficient L_I of the source images obtained by the NSCT transform is used as the input of the neuron

(2) A neuron pulse is generated according to Equation (6) and the number of ignition is calculated and used as the basis for the selection of low-frequency fusion coefficient. The formula is as follows:

$$T_{pq}(n) = T_{pq}(n-1) + O_{pq}(n). \tag{7}$$

(3) The low-frequency subband coefficients with more ignition times are selected as low-frequency fusion coefficients, and the low-frequency fusion image L_F is obtained as follows:

$$L_F(i,j) = \begin{cases} L_A(i,j), & T_{A,pq} \geq T_{B,pq}, \\ L_B(i,j), & T_{A,pq} < T_{B,pq}. \end{cases} \tag{8}$$

2.3.3. Highpass Subband Fusion Rule

Definition 1. Histogram distance refers to the accumulated value of the difference between two histograms, i.e., the sum of the difference in frequency between gray scales corresponding to two images.

In this paper, the histogram distance between the corresponding blocks of high-frequency coefficients is calculated. Assuming the numerical interval of a coefficient block is $[u, v]$, then the histogram function of this coefficient block is $p(r_k) = n_k/n$, where n is the total number of coefficients in this coefficient block, n_k is the k^{th} sum of numerical value in the coefficient block, and r_k is the k^{th} numerical value, $k = u,\ u+1, \cdots, v$.

Histogram distance is used to evaluate the similarity between frames in video image processing. Prompted by this, in this paper, distance histogram is introduced to the similarity measure of high-frequency subbands. Since high-frequency subbands mainly contain the detailed characteristics and edge information of the image, therefore, the high-frequency subbands obtained by multiresolution conversion of images of lung cancer are of multidirectional characteristics and structural similarity. In this paper, image is blocked with high-frequency coefficients and the histogram distance between the corresponding coefficient blocks is used to determine the similarity of the high-frequency decomposition coefficients. The calculation procedure of histogram distance is shown in Figure 4.

The specific steps are as follows:

First step—compressed measurement: linear measurement of high-frequency subband coefficient H_I^ξ is performed using Gaussian random measurement matrix Φ and the measurement value Y_I of the corresponding subband coefficient is obtained.

Second step—computation of match measure: the high-frequency subband H_I^ξ is divided into blocks with the size of 8×8. The coefficient blocks are extracted in accordance with the principle of top to bottom, left to right. The histogram distance T_1 between the high-frequency subbands of the two source images A and B is calculated according to Equation (9) and the obtained T_1 is used as the match measure.

$$T_1 = \sum_{i=1}^{u}\sum_{j=1}^{v}\left|p\left(H_A^\xi(r_k)\right) - p\left(H_B^\xi(r_k)\right)\right|. \tag{9}$$

In Equation (9), $p(H_A^\xi(r_k))$ is the histogram of the coefficient blocks of high-frequency subbands of the source image; the numerical interval of the coefficient block is $[u, v]$, and r_k is the k^{th} numerical value. The smaller the T_1 value, the higher the similarity between the two coefficients and the smaller the difference in histogram frequency of coefficient

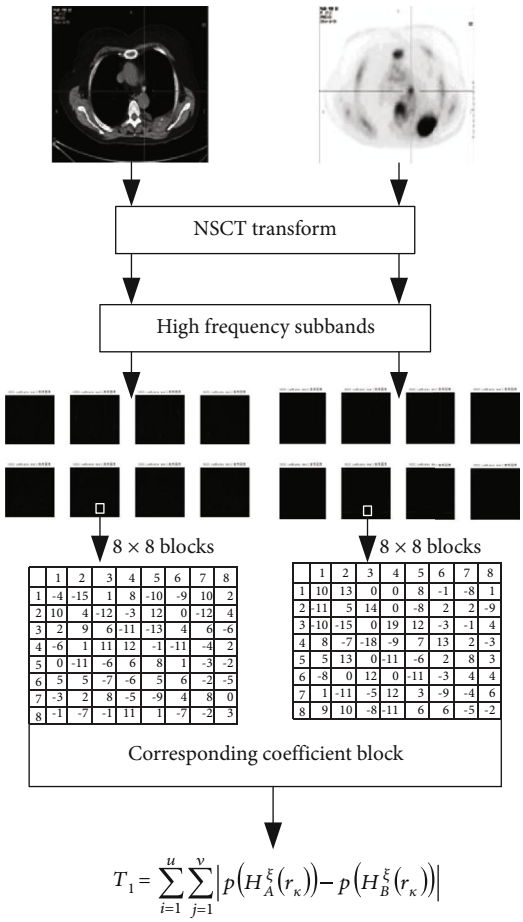


Figure 4: The calculation procedure of histogram distance.

blocks. In contrast, the bigger the T_1 value, the lower the similarity between source images A and B.

Third step—calculation of active measures: regional energy can better maintain correlation between two images and retain original useful information, thereby generating fused images with better visual effect. Thus, the regional energy is used as the active measure of the high-frequency fusion rule. Neighborhood division of high-frequency subbands H_I^{ξ} by $\omega(M \times N)$ is conducted by defining a 3×3 neighborhood window to calculate the regional energy with the following equation:

$$E_I^{\xi}(i,j) = \sum_{(i,j)\in\Omega(M,N)} \omega(i,j) H_I^{\xi}(i+M, j+N)^2 (I = \mathrm{A}, \mathrm{B}), \quad (10)$$

where $E_I^{\xi}(i,j)$ is the energy of the coefficient points (i,j) in the neighborhood range of $\omega(M \times N)$ and $w(i,j) = 1/8$ $[1\ 2\ 1; 2\ \ 1\ \ 2; 1\ \ 2\ \ 1]$ is the 3×3 neighborhood window used to calculate regional energy of high-frequency subbands.

Fourth step—self-adaptive decision module: the match measure M_{AB} and activity measure a_I are used to construct the self-adaptive decision model and calculate the fusion factor d for the high-frequency subbands. The threshold T is set as follows: it is assumed that if $T_1 < T$, the similarity of coefficient blocks is higher, then the fusion factor of low-frequency subbands is calculated by using the self-adaptive weighted calculation with regional energy. If $T_1 \geq T$, the difference in regional energy between the two high-frequency subbands is significant, then the high-frequency subbands with higher energy are used as the high-frequency subband coefficients of fusion image. Because the histogram distance of the high-frequency coefficient blocks of the two source images is quite different, the mean value of the histogram distance between the coefficient blocks is used as the threshold value in order to make it more flexible. Self-adaptive threshold setting can improve the accuracy of decision-making module. Decision factor d is calculated as follows:

$$d = \begin{cases} 1, & T_1 \geq T, E_{\mathrm{A}}(i,j) > E_{\mathrm{B}}(i,j), \\ 0, & T_1 \geq T, E_{\mathrm{A}}(i,j) < E_{\mathrm{B}}(i,j), \\ \dfrac{E_{\mathrm{A}}(i,j)}{E_{\mathrm{A}}(i,j) + E_{\mathrm{B}}(i,j)}, & T_1 < T. \end{cases} \quad (11)$$

Fifth step—combination module: after the determination of the match measure and activity measure, as well as the calculation of the decision factor d, the high-frequency measurement value Y_I is then determined, thereby getting the combination module. The specific expression of the combination module is as follows:

$$Y_{\mathrm{F}}(i,j) = d \times Y_{\mathrm{A}}(i,j) + (1-d) \times Y_{\mathrm{B}}(i,j), \quad (12)$$

where $Y_{\mathrm{A}}(i,j)$ and $Y_{\mathrm{B}}(i,j)$ are the measured values of high-frequency subband coefficients of source images A and B after multiresolution decomposition. Y_{F} is the high-frequency fused measurement value of the fusion image F.

Sixth step—reconstruction and recovery: the orthogonal matching pursuit algorithm is used to reconstruct the fused measurement value Y_{F} to get the high-frequency fusion image H_{F}. The final fusion image F is obtained by NSCT inverse transform of L_{F} and H_{F} simultaneously.

3. Discussion and Conclusion

3.1. Experimental Environment

3.1.1. Hardware Environment. The hardware platform used for the simulation experiment is Dual-Core (R) CPU E6700 Pentium, 3.2 GHz, 2.0 GB memory, with the operating system of Windows 7.

3.1.2. Software Environment. The software environment used is R2012b MATLAB version.

3.1.3. Experimental Data. As shown in Figure 5, the CT and PET images with the size of 256×256 were from two groups of registered patients with lung cancer.

3.1.4. Parameter Settings of NSCT Transform. The filter level is set to 1 and the direction of the series is set to 3, in which the NSP structure uses the double orthogonal wavelet for decomposition and NSDFB uses the trapezoidal filter.

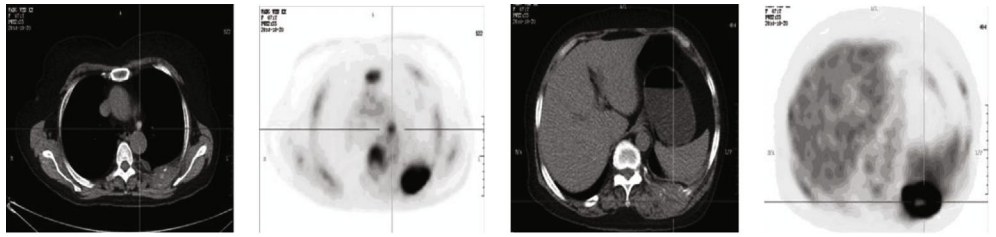


Figure 5: Source images.

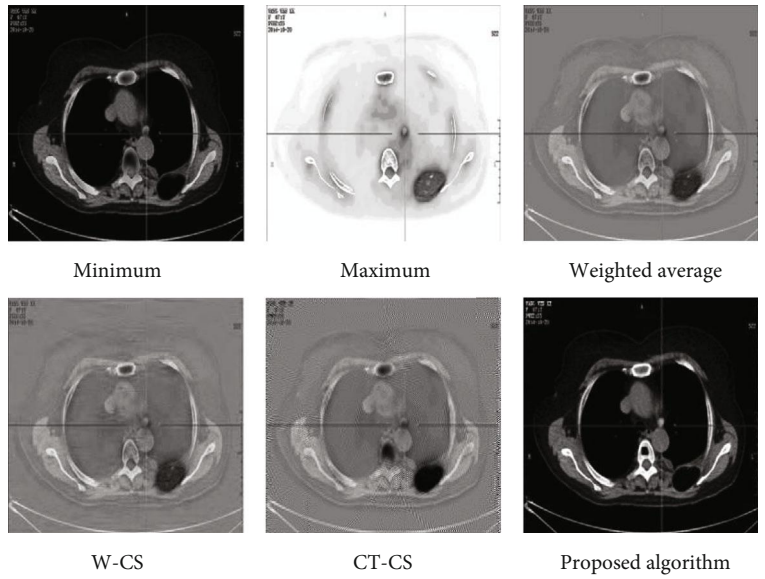


Figure 6: Fusion results of six methods.

3.2. Experimental Results and Analysis. In order to verify the superiority of the proposed algorithm, the proposed algorithm was compared with other fusion methods, including the traditional pixel image fusion methods—maximum method, minimum method, and weighted average method; image fusion methods based on compressed sensing—compressed sensing image fusion based on wavelet transform (W-CS) and compressed sensing image fusion based on contourlet transform (CT-CS). On this basis, a further experiment was conducted to study the effect of the activity measure and the match measure on the Piella framework, and to analyze the effect of different active measures and match measures on the performance of PET/CT image fusion.

The evaluation of the fusion image includes subjective evaluation and objective evaluation. Subjective evaluation is the most reliable for image quality inspection, especially in medical image fusion, which plays an important role in helping doctors make diagnosis. However, it is not easy to conduct a subjective evaluation since it not only requires equipment and strict working conditions but also requires a close cooperation of related persons. Therefore, those parameters for objective evaluation of the quality and performance of the fusion image were employed in this paper, including standard deviation (SD), average gradient (AG), spatial frequency (SF), peak signal-to-noise ratio (PSNR), information entropy (IE), mutual information (MI), and edge preserving quantity ($Q^{AB/F}$).

3.2.1. Experiment One: Comparative Experiment of Fusion Methods. Our proposed algorithm in this paper was compared with several other algorithms, including maximum method, minimum method, weighted average method, compressed sensing image fusion based on wavelet transform (W-CS), and compressed sensing image fusion based on contourlet transform (CT-CS). Among these algorithms, the sparse transformation matrix of W-CS is weighted by the weighted average method as the fusion rule and the low-frequency subbands in CT-CS are weighted by Gauss membership function as the fusion rule. For the high-frequency subbands, a method based on average gradient and regional energy is used to fuse the high-frequency measurement. In the fusion method based on compressive sensing, the measurement matrix is Gauss random matrix and the reconstruction algorithm is orthogonal matching pursuit algorithm, with the sampling rate of 50%. Figure 6 shows the fusion results of six methods.

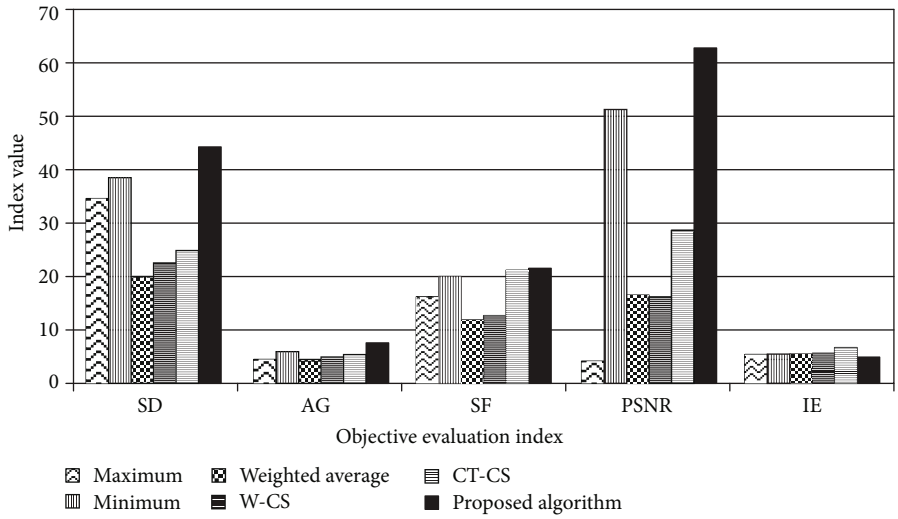


Figure 7: The histogram of six fusion methods for objective evaluation index.

As shown in Figure 6, grayscale fluctuations of the fusion images obtained with the simple pixel-level image fusion methods were generally smaller. For example, the pixel value of the fusion image obtained with the minimum method was lower and the brightness of the bone was dark. The pixel value of the fusion image with the maximum method was higher and the contrast of lesions was high, with a severely damaged spatial resolution. The fusion image obtained with the average weighted method, which calculates the median of maximum and minimum, was weakened and the lesions could not be accurately identified. While the W-CS and CT-CS methods had a twofold compression (50% reduction in file size) of the amount of data relative to the source images, the fusion quality is not high. The fusion image obtained with the W-CS method, for example, exhibited a water wave-like pattern horizontally, with a fuzzy texture and a blurred contour. The phenomenon of slight spectral overlap was observed in the fusion image obtained with the CT-CS method, which was resulted from the contourlet transform. The fusion method proposed in this paper could not only show clearly and completely the metabolic function of the lesions and the surrounding tissues but also increase the contrast between bone and soft tissues and organs. In contrast to the traditional fusion methods, the proposed method in this paper could reduce the fusion operation of the high-frequency image with 256×256 to 128×256 dimension and the reduction of the dimension of the data by compressed measurement greatly reduces the amount of data in the image fusion. The combination of compressed sensing and multiresolution transform can achieve the purpose of reducing the storage space in the process of image fusion.

The histogram of objective evaluation results for image fusion is shown in Figure 7. It could be seen that the proposed algorithm is superior to the other five methods in the evaluation of the objective index. With 50% of the sampling rate, the standard deviation (SD), average gradient (AG), spatial frequency (SF), and the peak signal-to-noise ratio (PSNR) of the proposed algorithm fusion image were higher than those of other five algorithms. Regarding the information entropy (IE), W-CS and CT+CS fusion algorithms were better than the proposed algorithm, which was caused by the fact that the fusion image of W-CS showed significant blur in some area and the contourlet transform could produce the phenomenon of spectral overlap, with a big change in gray level and a large edge fluctuation in the fusion image, thereby resulting in a larger quantity of image information. Except for the information entropy, the index values of the proposed fusion method were better than those of CT+CS and W-CS algorithm. Therefore, the multiresolution transform could make the image more sparse than the wavelet transform, producing a fusion image with a more detailed and complete information.

3.2.2. Experiment 2: Comparative Experiment of Activity Measure. The similarity between CT and PET was measured by the histogram distance which was used as the match measure, and the commonly used methods in image fusion, including energy, gradient, variance, and signal intensity, were used as the activity measure to investigate the effect of different activity measures on the fusion performance. Experimental results are shown in Table 1.

As shown in Table 1, overall, the changes in activity measure had no significant effect on the final result of the image fusion and there was no significant difference among the seven evaluation indexes. However, the standard deviation (SD), the average gradient (AG), spatial frequency (SF), and mutual information (MI) of the proposed algorithm were the highest of the fusion results in the four different active measures. When signal intensity was used as the activity measure, the peak signal-to-noise ratio (PSNR) and information entropy (IE) were the highest. The edge preserving quantity ($Q^{AB/F}$) was the highest when regional variance was used as the activity measure. Therefore, with the same match measure, it still can be concluded that the activity measure based on the regional energy has a better stability and a wide applicability and that the activity measure based on regional gradient has the least impact on the performance of image fusion.

Table 1: Objective evaluation index values of different activity measures for image fusion.

Methods (sampling rate = 50%)	SD	AG	SF	PSNR	MI	IE	$Q^{AB/F}$
Regional energy	**52.3534**	**6.7850**	**17.0333**	59.5880	**4.7553**	6.0011	0.5044
Regional gradient	52.0898	6.2285	15.3709	58.1131	4.2094	6.0039	0.4050
Regional variance	52.3518	6.7766	17.0194	59.5400	4.7473	6.0030	**0.5058**
Signal strength	52.3503	6.7729	16.9984	**59.7836**	4.7489	**6.0109**	0.5046

Table 2: Objective evaluation index values of different match measures for image fusion.

Methods (sampling rate = 50%)	SD	AG	SF	PSNR	MI	IE	$Q^{AB/F}$
Histogram distance	**52.3534**	**6.7850**	**17.0333**	**59.5880**	**4.7553**	*6.0011*	0.5044
Gradient ratio	52.3496	6.7836	*17.0158*	59.4785	4.7404	6.0041	*0.5041*
Energy ratio	52.3507	6.7781	17.0177	*59.4676*	4.7356	6.0040	0.5042
Signal strength ratio	52.3500	6.7829	17.0092	59.5400	*4.7355*	**6.0139**	**0.5047**
Structural similarity	*52.3435*	*6.7721*	17.0290	59.5362	4.7543	6.0083	0.5043

3.2.3. Experiment 3: Comparative Experiment of Match Measure. On the basis of experiment 2, the regional energy was chosen as the active measure to compare the PET and CT images. The gradient ratio, energy ratio, signal intensity ratio, structural similarity, all of which are commonly used to describe the similarities of images, and the proposed histogram distance were used as match measures to compare their influence on the performance of image fusion. The experimental results are shown in Table 2.

The standard deviation (SD), average gradient (AG), spatial frequency (SF), peak signal-to-noise ratio (PSNR), and mutual information (MI) were the highest when histogram distance was used as the match measure. In comparison, the information entropy (IE) and the edge preserving quantity ($Q^{AB/F}$) ranked the highest when the signal intensity ratio was used as the match measure. Therefore, it could be concluded that the match measure based on the histogram distance has a better stability and a wide applicability with the same activity measure; this is because that the match measure has the adaptability, enabling the fusion image to better integrate the redundant and complementary information of the source image and strengthen the ability of extract information from the source image. Taken the results of experiments 2 and 3 together, compared with other active measures and match measures, the proposed image fusion method based on the combination of regional energy and the histogram distance improves the performance and quality of image fusion, and has certain practical value.

3.2.4. Experiment 4: PET/CT Fusion Results of Lung Cancer (20 Groups). To further validate the effectiveness of this algorithm, a simulation experiment of 20 patients' PET and CT images with lung cancer was performed. Image size is 256 $*$ 256 and compared with maximum method, minimum method, weighted average method, compressed sensing image fusion based on wavelet transform (W-CS), and compressed sensing image fusion based on contourlet transform (CT-CS). The fusion results are shown in Table 3.

For these six methods, the objective evaluation is further carried out, indicators including standard deviation (SD), average gradient (AG), peak signal-to-noise ratio (PSNR), information entropy (IE), and edge preserving quantity ($Q^{AB/F}$). The evaluation indicators of these six methods were compared, respectively.

As can be seen in Table 4 and Figure 8, compared with weighted average method, compressed sensing image fusion based on wavelet transform (W-CS), and compressed sensing image fusion based on contourlet transform (CT-CS) method, the standard deviation values of proposed algorithm are the largest. The fusion results of the proposed algorithm are as follows: the 16 groups' standard difference value of fused image is higher than the maximum method (the proportion is 80%); the 17 groups' standard difference of fused image is higher than the minimum method (the proportion is 85%).

As can be seen in Table 5 and Figure 9, compared with minimum method, weighted average method, and compressed sensing image fusion based on wavelet transform (W-CS), average gradient value of proposed algorithm is the biggest. The fusion results of the proposed algorithm are as follows: the 14 groups' average gradient value of fused image is higher than compressed sensing image fusion based on contourlet transform (CT-CS); the proportion is 70%. The 19 groups' average gradient value of fused image is higher than the maximum method; the proportion is 95%.

As can be seen in Table 6 and Figure 10, compared with minimum and maximum method, weighted average method, compressed sensing image fusion based on wavelet transform (W-CS), and compressed sensing image fusion based on contourlet transform (CT-CS) method, PSNR results of fused image of proposed algorithm were the highest; next is the minimum method; the PSNR of maximum is the least.

As can be seen in Table 7 and Figure 11, compared with weighted average method, compressive sensing image fusion based on wavelet transform(W-CS), and compressed sensing image fusion based on contourlet transform (CT-CS)

Table 3: Fusion results of proposed algorithm and other algorithms (20 groups).

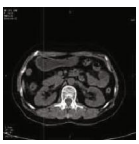
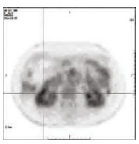
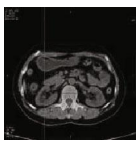
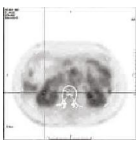
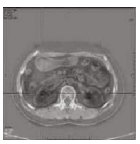
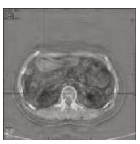
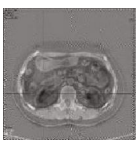
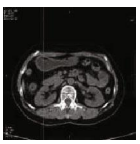
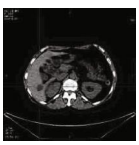
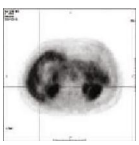
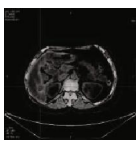
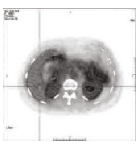
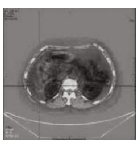
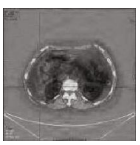
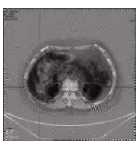
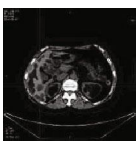
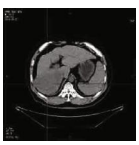
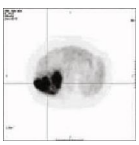
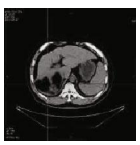
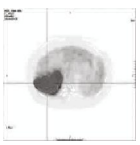
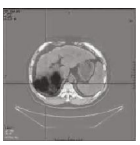
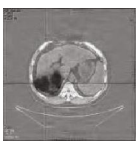
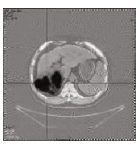
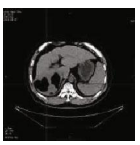
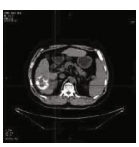
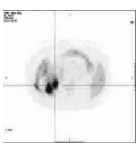
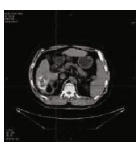
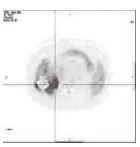
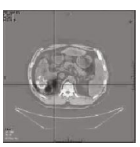
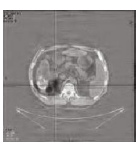
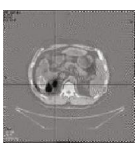
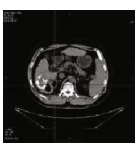
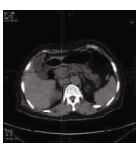
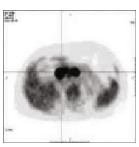
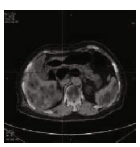
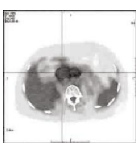
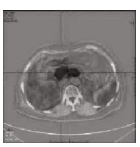
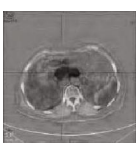
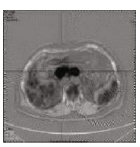
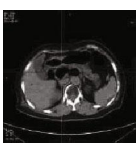
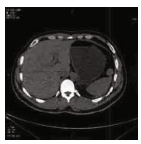
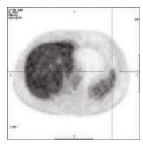
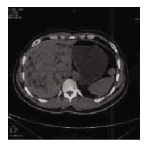
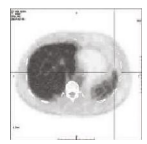
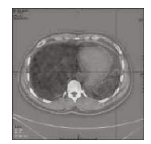
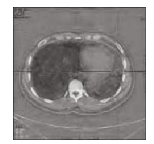
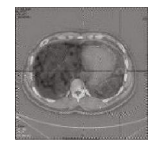
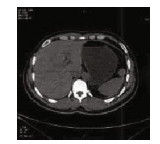
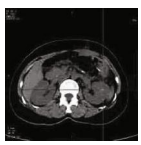
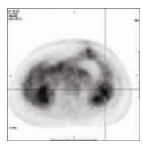
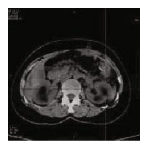
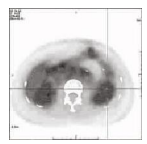
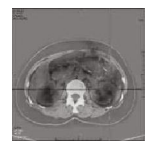
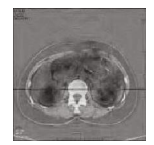
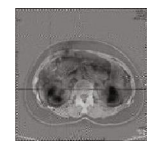
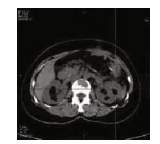
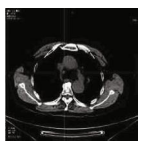
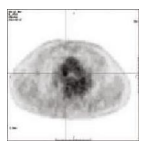
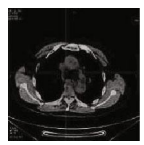
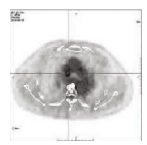
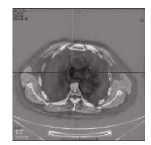
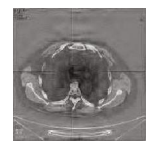
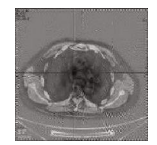
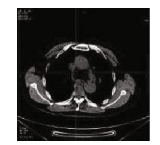
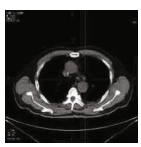
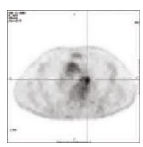
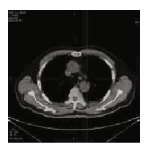
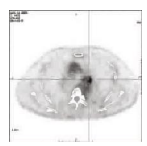
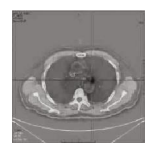
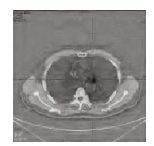
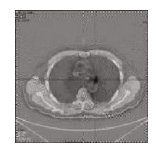
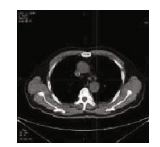
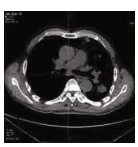
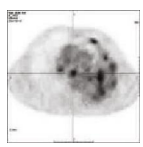
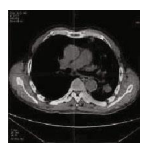
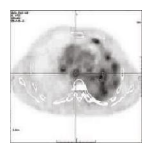
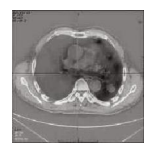
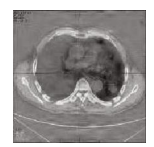
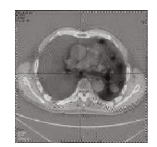
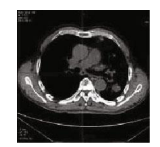
No.	CT	PET	Minimum	Maximum	Weighted average	W-CS	CT-CS	Proposed algorithm
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

Table 3: Continued.

No.	CT	PET	Minimum	Maximum	Weighted average	W-CS	CT-CS	Proposed algorithm
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								

method, information entropy of proposed algorithm is the largest, and it is close to minimum method; the information entropy of maximum method is greater than the proposed algorithm.

As can be seen in Table 8 and Figure 12, compared with weighted average method, compressed sensing image fusion based on wavelet transform (W-CS) method, and compressed sensing image fusion based on contourlet

Table 4: SD results of six fusion methods.

No.	Minimum	Maximum	Weighted average	W-CS	CT-CS	Proposed algorithm
1	40.8197	36.9042	21.6015	21.5099	25.4149	43.9313
2	35.4407	40.1216	28.3254	28.2984	33.3080	42.6002
3	47.9880	38.7795	24.2483	24.2095	31.2767	49.5474
4	44.8688	34.0856	23.1505	23.1364	29.3985	46.1418
5	39.2653	42.1417	22.5138	22.3139	28.7261	45.9070
6	32.4256	43.7024	25.5027	25.3936	31.6841	37.2577
7	44.0279	39.5197	28.1370	27.8942	30.1352	51.9443
8	38.9748	37.6799	28.2563	28.0749	33.8124	46.1748
9	39.2371	47.6241	33.3346	33.223	41.9339	51.0566
10	49.4623	46.8517	25.7300	25.5751	28.5237	53.7019
11	41.9787	46.8397	24.5689	24.4434	28.2317	47.4076
12	43.3735	43.1237	24.6357	24.5633	28.3111	49.9585
13	42.1685	48.7261	27.2208	27.0723	31.8086	47.8887
14	44.6154	38.9140	23.7810	23.5821	26.9285	49.3080
15	53.2571	46.9942	31.1363	30.7504	34.2505	58.8691
16	46.4683	40.8302	23.1941	23.0332	29.8259	38.6864
17	43.4063	42.2500	21.5156	21.5181	26.6296	34.9590
18	46.4432	41.1815	23.2101	23.1058	29.3805	38.6287
19	36.1625	35.5986	21.4938	21.3850	29.0136	38.3435
20	39.6568	36.0124	20.4559	20.3050	25.1195	42.8999

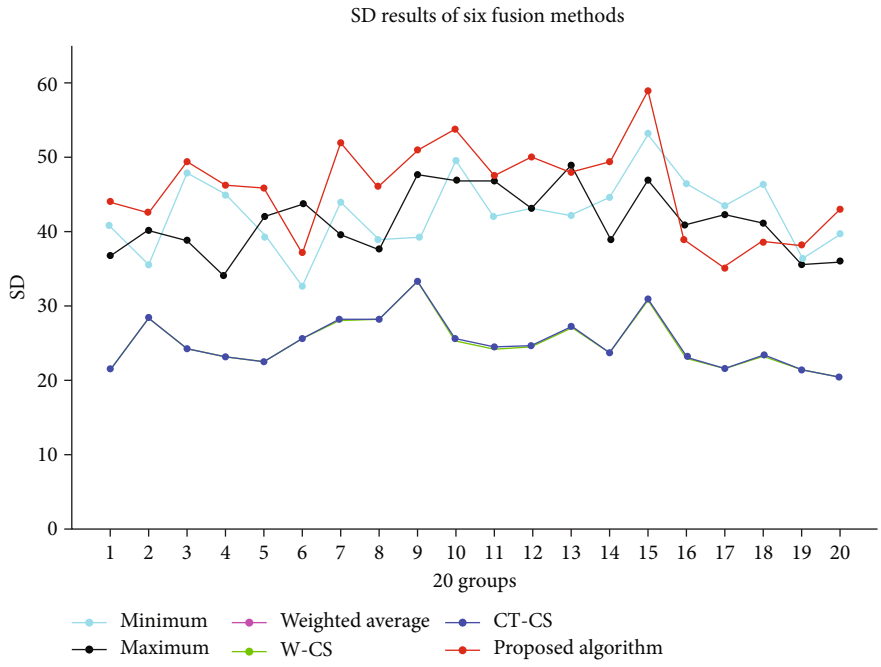


Figure 8: SD results of six fusion methods.

transform (CT-CS) method, the edge preserving quantity of fused images is the largest. The fusion results of the proposed algorithm are as follows: the 15 groups' $Q^{AB/F}$ value of fused image is higher than maximum method; the proportion is 75%. The 19 groups' $Q^{AB/F}$ value of fused image is higher than minimum method; the proportion is 95%.

In summary, proposed algorithm got a better effect both from subjective evaluation and objective evaluation, already obtained higher amount of information of fusion image,

Table 5: AG results of six fusion methods.

No.	Minimum	Maximum	Weighted average	W-CS	CT-CS	Proposed algorithm
1	6.3568	6.1624	4.9747	5.4682	5.0510	6.7764
2	6.4543	6.0410	5.1052	5.8527	6.0453	7.4259
3	6.8936	5.0006	5.0906	5.8554	7.7507	7.2631
4	6.9435	4.6529	5.1536	5.9005	8.3495	7.1439
5	6.1762	6.2140	5.0389	2.7049	5.6455	6.6749
6	4.9021	7.8187	5.1719	5.7303	6.1396	5.8516
7	7.4819	6.1103	5.6800	6.5694	7.7932	8.7691
8	7.3826	6.3881	5.6352	6.4652	7.7400	8.3470
9	8.8585	8.6682	7.1002	7.9744	9.9618	11.7203
10	7.0904	5.7411	5.5165	6.4633	6.6269	7.5915
11	6.8790	6.0677	5.4230	6.3439	5.9420	7.2829
12	6.6091	6.0912	5.1921	5.9674	5.9810	7.2460
13	8.1346	6.6117	6.3075	7.3008	6.6988	9.2933
14	6.6621	6.0377	5.4031	6.2970	6.8243	7.3030
15	8.5520	6.4565	6.3934	7.4013	9.2418	9.6747
16	5.8153	5.3448	4.7786	5.4066	8.8269	6.2349
17	5.0234	5.3312	4.3953	5.0137	6.246	5.3011
18	5.7055	5.4326	4.5938	5.3087	7.9593	6.1365
19	5.703	4.1691	4.0767	4.7882	7.6893	6.3831
20	6.2097	5.4733	4.5821	5.1453	4.6855	6.7737

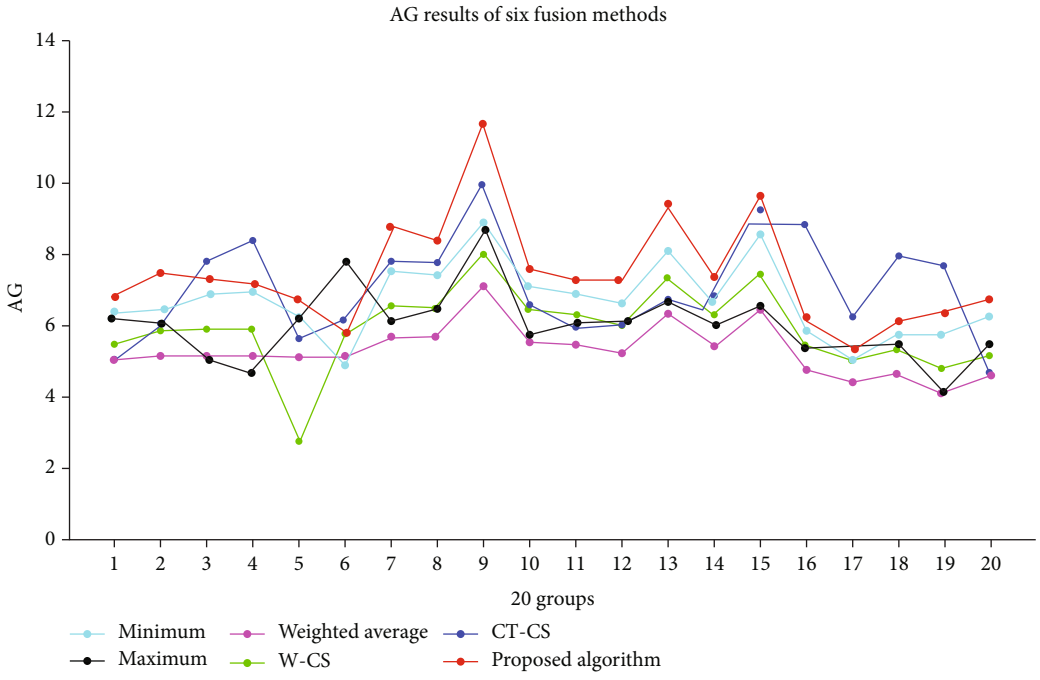


Figure 9: AG results of six fusion methods.

extracted the useful information in original image, and showed a significant comprehensive advantage; the extraction and synthesis of useful information in original images showed a significant advantage, which fully reflects the advantages of compressed sensing and nonsubsampled contourlet transform. Fusion images effectively combine the functional information and anatomical structure of CT image and the physiological and pathological information of PET image in patients with lung cancer. It is good for doctors to analyze and judge the lesions, and provide effective imaging information for clinical work, surgery, and disease diagnosis.

Table 6: PSNR results of six fusion methods.

No.	Minimum	Maximum	Weighted average	W-CS	CT-CS	Proposed algorithm
1	67.6778	4.2274	16.9469	16.7103	25.9103	98.4371
2	51.3925	4.0218	17.5258	18.1958	25.8573	58.6880
3	57.8488	3.1679	16.7534	17.3400	26.3616	61.6637
4	68.9252	2.7377	15.9110	16.2908	30.5160	74.8556
5	49.1341	4.3197	17.3668	18.5973	25.9304	56.4141
6	37.8793	4.9918	17.0994	16.8266	25.6844	41.7599
7	48.6666	5.5057	19.1978	18.8378	28.1369	64.9162
8	51.7864	5.5254	19.2524	18.5608	30.6913	60.3129
9	38.8599	8.3020	21.7048	21.4392	29.9463	43.9237
10	63.0813	4.0567	16.8100	18.1006	27.2135	76.2280
11	61.1094	5.1122	17.398	17.9735	31.5651	80.3592
12	54.2517	4.9864	17.4909	17.5970	27.5333	65.8369
13	55.5879	4.2598	18.064	19.1143	27.9871	63.5656
14	63.3288	3.6572	16.3681	16.8296	22.7251	79.6963
15	58.5403	5.3442	19.1583	19.9471	28.8951	79.7845
16	64.2377	2.2901	16.1327	17.0276	24.9213	71.2614
17	70.716	2.3219	15.7789	15.8686	26.5409	78.4282
18	66.8253	2.2940	16.1412	16.5010	24.9676	73.1706
19	55.7279	2.2588	16.0742	15.9996	27.0434	63.3388
20	65.7483	4.2700	16.6521	17.3850	31.7650	91.9760

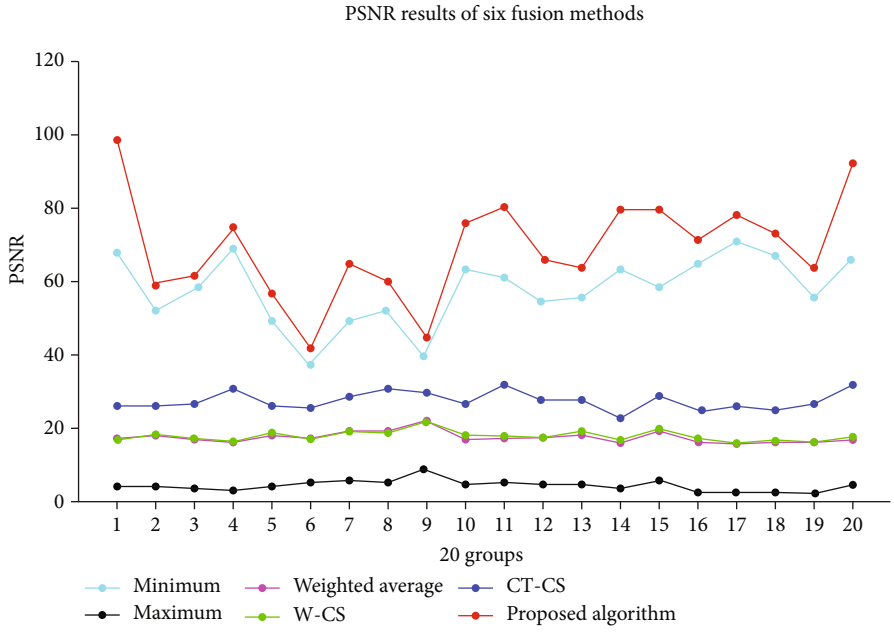


Figure 10: PSNR results of six fusion.

4. Conclusions

In this paper, a fusion rule, which is a self-adaptive fusion algorithm of PET/CT based on compressed sensing and histogram distance, is proposed. Firstly, the NSCT transform is performed on the PET and CT images. The fusion rule of the PCNN, which has a higher sensitivity to low-frequency image, is then used to highlight the lesions of the image. Secondly, the Gauss random matrix is used to obtain the measured values of the high-frequency subbands, the histogram distance of the high-frequency subblocks is used as the match measure, and the regional energy of the high-frequency subbands is used

Table 7: IE results of six fusion methods.

No.	Minimum	Maximum	Weighted average	W-CS	CT-CS	Proposed algorithm
1	5.7490	6.5975	3.4647	2.1908	1.3935	5.8255
2	4.6736	5.9679	3.7436	2.3225	1.8385	4.4064
3	4.7806	5.0008	3.7177	2.1721	1.5721	4.6331
4	4.6065	4.6319	3.3908	1.9372	1.2853	4.5862
5	5.5333	6.2812	3.5865	2.0250	1.5522	5.4658
6	4.1414	6.7461	4.0673	2.4312	1.8106	3.8784
7	4.9011	6.3873	4.1083	2.2561	1.5276	4.7905
8	5.4872	6.9503	3.8959	2.2320	1.6782	5.2839
9	5.9661	7.3766	3.8357	2.4102	2.1632	5.1696
10	5.1174	5.8507	3.7360	2.0422	1.4869	5.1696
11	6.0299	6.5942	3.8602	2.3184	1.6808	6.2032
12	5.4917	6.7467	3.7625	2.3607	1.6099	5.5316
13	4.5776	5.8927	3.9867	2.0982	1.4397	4.4849
14	4.8837	5.6078	3.6012	2.0416	1.4039	4.9171
15	5.5988	6.3457	4.1377	2.2427	1.7230	5.5535
16	4.0555	4.4984	3.4031	1.8996	1.3138	3.9313
17	4.0548	4.6739	3.2521	1.8294	1.2491	4.0369
18	4.1442	4.5384	3.3886	1.9017	1.3587	4.0917
19	4.3144	4.5056	3.2706	1.8754	1.3255	4.1776
20	5.7146	6.7371	3.5198	2.2422	1.3879	5.7593

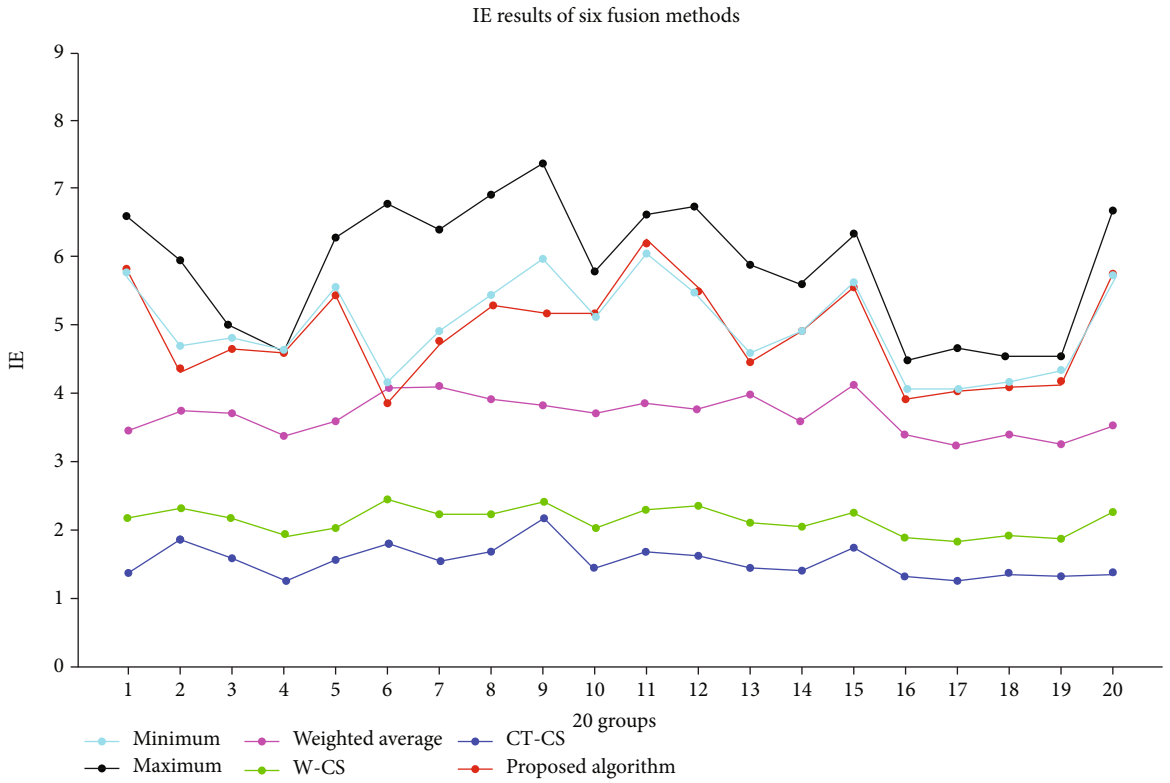


Figure 11: IE results of six fusion.

Table 8: $Q^{AB/F}$ results of six fusion methods.

No.	Minimum	Maximum	Weighted average	W-CS	CT-CS	Proposed algorithm
1	0.4906	0.5260	0.3378	0.2408	0.1899	0.5256
2	0.4650	0.5164	0.3411	0.2474	0.1960	0.5002
3	0.5168	0.4404	0.3694	0.2552	0.2027	0.5359
4	0.5420	0.4557	0.3885	0.2749	0.2085	0.5527
5	0.4649	0.5347	0.3385	0.2424	0.1945	0.4866
6	0.3623	0.6141	0.3347	0.2431	0.1641	0.3925
7	0.4974	0.4715	0.3558	0.2543	0.1660	0.5836
8	0.4856	0.4882	0.3434	0.2300	0.1827	0.5249
9	0.4689	0.5099	0.3358	0.2205	0.1725	0.5662
10	0.5378	0.4542	0.3750	0.2612	0.1934	0.5662
11	0.4962	0.5030	0.3644	0.2535	0.1925	0.5252
12	0.4953	0.5143	0.3546	0.2499	0.1913	0.5306
13	0.4831	0.4622	0.3678	0.2469	0.1594	0.5439
14	0.4954	0.4986	0.3607	0.2549	0.1736	0.5416
15	0.5194	0.4292	0.3655	0.2405	0.1774	0.5839
16	0.4555	0.4885	0.3665	0.2605	0.1765	0.4862
17	0.4436	0.5474	0.3720	0.2633	0.1812	0.4169
18	0.4854	0.5289	0.3653	0.2592	0.1715	0.5113
19	0.4552	0.4303	0.3294	0.2359	0.1873	0.4924
20	0.4875	0.5215	0.3340	0.2438	0.2010	0.5362

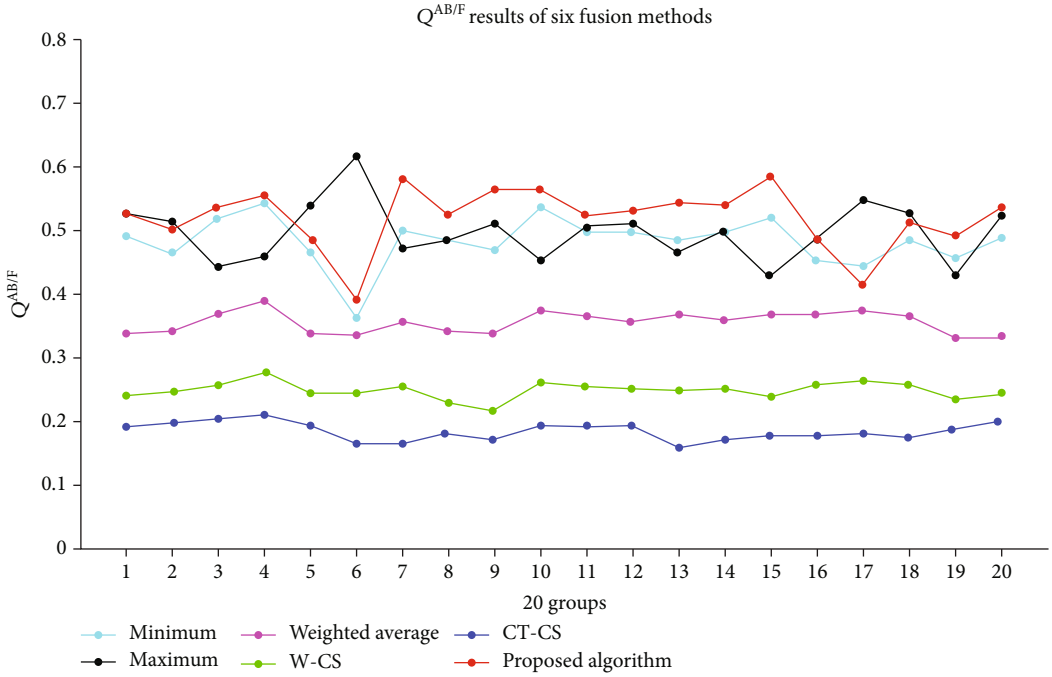


Figure 12: $Q^{AB/F}$ results of six fusion methods.

as the activity measure. The fusion factor d is calculated by using the match measure and the activity measure; the high-frequency measurement value is fused according to the fusion factor, and the high-frequency fusion image is reconstructed by using the orthogonal matching pursuit algorithm of the high-frequency measurement. Thirdly, the final fusion image is acquired through the NSCT inverse transformation of low-frequency fusion image and high-frequency fusion image. Finally, four experiments' results show that the algorithms are better than other algorithms.

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 they have no conflicts of interest.

Acknowledgments

This work made an oral report in the Workshop 1: Adversarial Machine Learning and IEEE WCCI2020. It is supported by the Natural Science Foundation of China (Grant No. 62062003), the Key Research and Development Project of Ningxia (special projects for talents) under Grant No. 2020BEB04022, and the North Minzu University Research Project of Talent Introduction under Grant No. 2020KYQD08.

References

[1] B. Meher, S. Agrawal, R. Panda, and A. Abraham, "A survey on region based image fusion methods," *Information Fusion*, vol. 48, pp. 119–132, 2019.

[2] C. Li, G. Li, Y. Tan, and X. Xu, "Medical images fusion of nonsubsampled contourlet transform and regional feature," *Journal of Computer Applications*, vol. 33, no. 6, pp. 1727–1731, 2013.

[3] Z. Tao, L. Huiling, W. Xinyu, and X. Yong, "Self-adaption fusion algorithm of lung cancer PET/CT based on Piella frame and DT-CWT," *Journal of University of Science and Technology of China*, vol. 47, no. 1, pp. 10–17, 2017.

[4] Z. Zhang and R. S. Blum, "A categorization of multiscale decomposition based image fusion schemes with a performance study for a digital camera application," *Proceedings of the IEEE*, vol. 87, no. 8, pp. 1315–1326, 1999.

[5] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, no. 4, pp. 259–280, 2003.

[6] E. J. Candes, *Ridgelets: theory and applications, Technical Report No. 1998-17*, Stanford University, California, 1998.

[7] J. L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

[8] E. Le Pennec and S. Mallat, "Bandlet image approximation and compression," *Multiscale Model and Simulation*, vol. 43, pp. 992–1039, 2005.

[9] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multi-resolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[10] A. L. Da Cunha, J. P. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: theory, design and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089–3101, 2006.

[11] G. Easley, D. Labate, and W. Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, 2008.

[12] Z. Xin and C. Weibin, "Medical image fusion based on weighted contourlet transformation coefficients," *Journal of Image and Graphics*, vol. 19, no. 1, pp. 133–140, 2014.

[13] T. Xiuhua and X. Wang, "Research on NSCT-based medical image fusion," *Computer Applications and Software*, vol. 30, no. 4, pp. 287–289, 2013.

[14] L. I. Guangxin, X. U. Shuyan, and W. U. Weiping, "Extension of Piella pixel-level multiresolution image fusion framework and its algorithm," *Precision Engineering*, vol. 20, no. 12, pp. 2773–2780, 2012.

[15] W. Da and B. Hongyu, "Sonar image fusion research based on Piella framework," *Journal of Sichuan University (Engineering Science Edition)*, vol. 47, no. 2, pp. 95–101, 2015.

[16] L. Yu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform," *Information Fusion*, vol. 24, pp. 147–164, 2015.

[17] W. Zhao, Z. Xua, and J. Zhao, "Gradient entropy metric and p-Laplace diffusion constraint-based algorithm for noisy multispectral image fusion," *Information Fusion*, vol. 27, pp. 138–149, 2016.

[18] L. Dong, Q. Yang, H. Wu, H. Xiao, and M. Xu, "High quality multi-spectral and panchromatic image fusion technologies based on curvelet transform," *Neurocomputing*, vol. 159, pp. 268–274, 2015.

[19] Q. Wang, S. Li, H. Qin, and A. Hao, "Robust multi-modal medical image fusion via anisotropic heat diffusion guided low-rank structural analysis," *Information Fusion*, vol. 26, pp. 103–121, 2015.

[20] T. Zhou, H. Lu, J. Zhang, and H. Shi, "Pulmonary nodule detection model based on SVM and CT image feature-level fusion with rough sets," *BioMed Research International*, vol. 2016, Article ID 8052436, 13 pages, 2016.

[21] Z. Tao, L. Huiling, W. Wenwen, and Y. Xia, "GA-SVM based feature selection and parameter optimization in hospitalization expense," *Applied Soft Computing*, vol. 75, pp. 323–332, 2019.

[22] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, and Y. Dong, "The ensemble deep learning model for novel COVID-19 on CT images," *Applied Soft Computing*, vol. 98, article 106885, 2021.

[23] Z. Tao, L. Huiling, F. Hu, S. Qiu, and W. Cuiying, "A model of high-dimensional feature reduction based on variable precision rough set and genetic algorithm in medical image," *Mathematical Problems in Engineering*, vol. 2020, Article ID 7653946, 18 pages, 2020.

[24] T. Wan, N. Canagarajah, and A. Achim, "Compressive image fusion," in *2008.15th IEEE International Conference on Image Processing,ICIP.IEEE*, pp. 1308–1311, San Diego, CA, USA, 2008.

[25] X. Luo, J. Zhang, J. Yang, and Q. Dai, "Classification-based image-fusion framework for compressive imaging," *Journal of Electronic Imaging*, vol. 19, no. 3, pp. 9–14, 2010.

[26] T. Wan and Q. Z. Chang, "An application of compressive sensing for image fusion," *International Journal of Computer Mathematics*, vol. 88, no. 18, pp. 3915–3930, 2011.

[27] B. C. Hui, H. X. Guang, and J. W. Guo, "An image fusion method based on combination of NSCT and PCNN," *Remote Sensing Information*, vol. 30, no. 2, pp. 50–56, 2015.

[28] X. Li and S. Y. Qin, "Efficient fusion for infrared and visible images based on compressive sensing principle," *IET Image Processing*, vol. 5, no. 2, pp. 141–147, 2011.

[29] N. Xue, H. L. Wang, L. Chen, and J. M. Wang, "Image compressed sensing based on sparse representation using

contourlet directional subbands," *Application Research of Computers*, vol. 30, no. 6, pp. 1889–1891, 2013.

[30] S. Yan-Fei, L. Jin-Tao, Z. Zhen-Min, Z. Yong-Dong, and D. Feng, "Image reconstruction algorithm of compressed sensing based on nonlocal similarity model," *Acta Automatica Sinica*, vol. 41, no. 2, pp. 261–272, 2012.

[31] J. Li, Q. Wang, Y. Shen, and B. Li, "Collaborative construction of measurement matrix and reconstruction algorithm in compressive sensing," *Acta Electronica Sinica*, vol. 41, no. 1, pp. 29–34, 2013.

Hindawi
BioMed Research International
Volume 2021, Article ID 6653879, 16 pages
https://doi.org/10.1155/2021/6653879

Research Article

QAIS-DSNN: Tumor Area Segmentation of MRI Image with Optimized Quantum Matched-Filter Technique and Deep Spiking Neural Network

Mohsen Ahmadi[1], **Abbas Sharifi**[2], **Shayan Hassantabar**[3] and **Saman Enayati**[4]

[1]*Department of Industrial Engineering, Urmia University of Technology (UUT), P.O. Box: 57166-419, Urmia, Iran*
[2]*Department of Mechanical Engineering, Urmia University of Technology (UUT), P.O. Box: 57166-419, Urmia, Iran*
[3]*Department of Electrical Engineering, Princeton University, P.O. Box: 08544, Princeton, NJ, USA*
[4]*Department of Computer and Information Science, Temple University, P.O. Box: 19122, Philadelphia, USA*

Correspondence should be addressed to Mohsen Ahmadi; mohsen.ahmadi@ine.uut.ac.ir

Received 27 November 2020; Revised 29 December 2020; Accepted 6 January 2021; Published 21 January 2021

Academic Editor: Huiyu Zhou

Tumor segmentation in brain MRI images is a noted process that can make the tumor easier to diagnose and lead to effective radiotherapy planning. Providing and building intelligent medical systems can be considered as an aid for physicians. In many cases, the presented methods' reliability is at a high level, and such systems are used directly. In recent decades, several methods of segmentation of various images, such as MRI, CT, and PET, have been proposed for brain tumors. Advanced brain tumor segmentation has been a challenging issue in the scientific community. The reason for this is the existence of various tumor dimensions with disproportionate boundaries in medical imaging. This research provides an optimized MRI segmentation method to diagnose tumors. It first offers a preprocessing approach to reduce noise with a new method called Quantum Matched-Filter Technique (QMFT). Then, the deep spiking neural network (DSNN) is implemented for segmentation using the conditional random field structure. However, a new algorithm called the Quantum Artificial Immune System (QAIS) is used in its SoftMax layer due to its slowness and nonsegmentation and the identification of suitable features for selection and extraction. The proposed approach, called QAIS-DSNN, has a high ability to segment and distinguish brain tumors from MRI images. The simulation results using the BraTS2018 dataset show that the accuracy of the proposed approach is 98.21%, average error-squared rate is 0.006, signal-to-noise ratio is 97.79 dB, and lesion structure criteria including the tumor nucleus are 80.15%. The improved tumor is 74.50%, and the entire tumor is 91.92%, which shows a functional advantage over similar previous methods. Also, the execution time of this method is 2.58 seconds.

1. Introduction

Brain tumors, which are well known to be one of the most common diseases of the nervous system, can cause many damages to human health and can also result in death. In this matter, the most common type of brain tumor among adults is glioma [1]. These tumors can be classified based on their grades as follows: Low-Grade Gliomas (LGG) exhibit benign trends and provide better patient awareness, whereas High-Grade Gliomas (HGG) are malignant, which may lead to receiving worse patient awareness [2]. The medical image of brain tumors helps assess disease development before and after treatment. Several imaging techniques, such as MRI, CT, PET, and SPECT imaging, have been used to examine brain tumors. However, MRI imaging is now the main imaging technique that can be used for glioma's diagnosis and treatment, because it has advantages such as good soft-tissue disparity, multiplied parameters, shooting in the desired direction, noninvasive photography, and so on. It also has various sequences, such as T1 weight images, T1 or T1ce-enhanced contrast, T2 weight, and Fluid Attenuation Inversion Retrieval (FLAIR). These sequences offer additional

details about different parts of brain tumors [3]. For instance, the tumor area via peritumoral edema may be diagnosed in FLAIR and T2 images. Conversely, the tumor nucleus area without peritumoral edema is more prominent in images of T1 and T1ce. In this way, the different main MRI methods focus on detailed information of images, which describe the features of brain tumors under several sides.

For a medical diagnosis, accurate segmentation of these tumors is critical and needs therapeutic planning. Segmentation of brain tumors in an automatic way and existing infrastructures from medical imaging allow for accurate diagnosis of tumors. It can help plan surgery and the treatment of brain tumors by providing a more efficient and better diagnosis [4]. In particular, it is critical to divide these tumor tissues, such as enhancing core, necrosis, edema, and nonenhancing core in terms of the natural brain tissue, containing white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Nevertheless, the precise automatic segmentation of these tumors is a challenging issue due to several reasons. In image segmentation operations, the outlines between the normal tissues and brain tumor are blurred because of the partial size effects, the gradient filtering intensity, and the magnetic field artifacts. Moreover, brain tumors are very varied in terms of size, shape, and location in patients. It is recommended to utilize a novel, robust, and fast method with the utmost care in the field of image segmentation. The segmentation of different images is a separate issue, and the right method should be designed according to each structure that should be segmented with a specific purpose. Deep convolutional neural networks have done very well in recent years in brain tumor segmentation [5]. In this regard, the convolutional neural network (so-called CNN) is a popular deep learning model that can elicit some favorite features for the original data classification [6].

This article proposes a new optimal framework of the brain tumor segmentation of MRI images that uses the structure of an optimized deep spiking network with the Quantum Artificial Immune System (QAIS). This framework is fully integrated with the QAIS-DSNN and conditional random field (CRF) combination. In the first step, a multiplied level architecture network is proposed to consider interdependence segmentation among neighboring pixels and supplementary information in various layers and measures. The background textual information of the three-dimensional MRI images is essential for brain tumor segmentation that is not taken into account by the CNNs. The study also introduces connected CRFs to correct the mapping probability attained by QAIS-DSNN.

2. Literature Review

To date, several methods have been proposed for MRI imaging. This section examines an overview of several classified methods.

2.1. Research on Deep Learning-Based Methods. The importance of MRI imaging methods for brain tumors in recent years with deep learning principles and methods due to high applications and relevant results has been highly regarded. In [7], the design of different types of convolutional neural network architecture is proposed in the form of 3×3 windowing with a deep layer in different grades of gliomas specimens using small nuclei. A two-way convolutional neural network model has been proposed in [8], and one channel provides detailed features of local and the other provides universal feature extraction. In [9], a convolutional neural network architecture has been created as a cascaded CNN to obtain the local dependencies of tags, achieving better performance in segmentation. Besides, they selected a two-step training strategy to address label imbalance distribution. Recently, there are advantages of multiscale features of the convolutional neural network in segmentation work [10–16]. In general, there are two methods to elicit features of multiplied scales: the first method is to use feature mapping of different network levels to show multiscale features [10].

In this respect, a multiscale convolutional neural network has been suggested to divide the retinal vein in [17]. Scale images are identified at different stages of the convolutional neural network to obtain the retinal arteries' probability mappings. Also, in [18], the structure of the Fully Convolutional Neural Network (FCNN) was developed for training with CRF; however, the process of training them was extremely time-consuming and expensive in terms of memory consumption. The second case is the transfer of versions on a different scale from the input image using the same network [10]. Also, multiscale features have been obtained by the convolutional neural network in [18]. This paper adopted three-dimensional CRFs to process segmentation results, but configuring three-dimensional CRFs is a complex process. In [18], different sizes from a convolutional neural network architecture have been used as cascaded CNN to record multiscale features. Due to this research and, of course, many other types of research that are beyond the scope of this research, the convolutional neural network has achieved significant achievements. The ability to learn neural networks with architecture and fixed parameters is limited, and the useful information, for three-dimensional MRI data, may be overlooked.

Some researchers use two-dimensional [19] or three-dimensional convolutional neural network models [18, 20, 21] to deal with three-dimensional images. For brain tumor segmentation, a three-dimensional semantic segmentation network based on the encoder-decoder architecture was developed in this way [22]. A hierarchical segmentation system that has varied the segmentation into three binary tasks has been proposed [19, 23, 24]. They also taught models of segmentation from sagittal, coronal, and axial perspectives. In the practical step, to achieve the final results, they averaged the SoftMax outputs obtained in the mentioned perspectives. Even though these methods do work very well, they raise both memory consumption and fiscal complexity. Thus, fiscal models, such as conditional random fields (CRFs) and Markov Random Fields (MRFs), are mainly employed to investigate spatial text information. In [25], a neonatal structure of a deep neural network called the Growing Deep Convolutional Neural Network (GCNN) is presented to segment MRI images to diagnose brain tumors.

There is also another method combined with GCNN that is a Stationary Wavelet Transform (SWT). The hybrid deep

learning method is simulated with the use of BraTS2018 dataset and evaluated using the peak signal-to-noise ratio (PSNR), the average square error, and so on. In [26], a complete convolutional neural network with pyramidal features is presented as an Atrous convolution for brain tumor segmentation by MRI images. This research uses data sets from BraTS2013, BraTS2015, and BraTS2018, the results of which have a functional advantage over lesion structure, including tumor nucleus, improved tumor, and the whole tumor, compared to previous methods, especially the convolutional neural network. These results are based on the Dice criterion, 76.88% for the tumor nucleus, 74.43% for the optimized nucleus, and 86.58% for the entire tumor. Also, in [27], the convolutional neural network is used in three dimensions based on a method called Test-Time Augmentation. This research uses BraTS2018 data and shows the results of its evaluation with a lesion structure, including tumor nucleus, improved tumor, and whole tumor, with functional superiority over many convolutional methods and deep networks. These results were in two ways using the Dice criterion, which was 90.21% for the tumor nucleus, 79.72% for the optimized nucleus, and 85.83% for the entire tumor. In a similar study, in [28], the convolutional neural network is considered to be multicascaded (CNN) and conditional random field proposed as MCCNN. The results of this study, based on the lesion structure criteria, were 71.78% for the improved nucleus, 88.24% for the total tumors, and 74.81% for the tumor nucleus. For breast imaging monitoring and data system ranking, Kang et al. [29] indicated a dominant fuzzy full-connected layer. The aim of the model was to establish complementary scoring properties for semantic segmentation with fuzzy rules.

2.2. A Review of Deep Spiking Neural Network. First of all, it should be noted that sparks are the neurons of the neural network that use spikes instead of neurons in the spiking neural network, and a set of neurons in an input layer with spikes is called a spark. spiking neural networks (SNNs) are driven by the processing of biological knowledge, which communicates in parallel scattered and nonsynchronous binary signals. In neuromorphic hardware, SNNs indicate some appropriate features such as fast inference, low energy use, and event-dependent processing of information. It creates interesting applicants to apply deep learning (DL) networks effectively and a selection process for several learning tasks at a computer. Here, SNNs consider a wide range of training methods, including the conversion of convolutional deep networks to SNNs, limited preconversion training, and a variety of biological motivations [30].

Neural networks are usually read if they have at least two hidden layers of nonlinear input conversion. In this study, only feedback networks are considered to calculate mapping from input to output. Spiking neural networks were initially studied as biological information processing models in which neurons exchange information through spikes. Here, all spikes are expected to be stereotypical events; in this way, data processing is minimized to two main factors: First of all, the timing of spikes, for example, firing frequency, the relative timing of pre-/postsynapse spikes, and special patterns of movement. Secondly, the identification of the synapses used means it is possible to connect nerve cells, whether the synapse is stimulating or inhibitory. With regard to the degree of detail of the simulation neurons, the two neurons are the point at which the input spikes alter their (somatic) membrane potential immediately or are built together with complex (dendritic) spatial structures as multichamber models. Hence, the dendritic currents will communicate before that. There were also changes to physical capacity. Here, several models of spike neurons, such as Hodgkin's Huxley model, integrate-and-fire, and spike response, explain the evolution of membrane potential and the spike of different rates of detail in development. Essentially, the membrane potential of the stream merges with the entry of the spikes and generates a new spike since the threshold is crossed. After the spike is obtained, the small axon is sent to all the linked nerve cells by a delay via the axon, based on which the membrane potential is adjusted to a certain base. Figure 1 shows this.

Direct communication between spiking and analog neural networks is formed by assuming a stable state, by considering the activation of an analog neuron is equal to the firing rate of a spiking neuron. Many geometric models used those rate codes to describe brain computational processes. Nevertheless, more complex processes can also form the neural spike models, which depend on some reference signals or relative timing between spikes, such as network fluctuations. Temporary codes are very important in biology; even a spike or small time-consuming changes in neuron firing may cause different reactions, as most decisions must be calculated before a reliable estimation of the spike [30].

In addition to the biological definition of SNNs, they contain a pragmatic functional representation in the field of neural engineering; SNNs are commonly referred to as spikes and are event-based. An event here is a collection of digital information defined by a time marker's origin and destination address. Unlike biologically motivated SNNs, it may have several bits of load information. The source for this protocol is the address index or AER (Address Event Representation) protocol, which is used after processing to link to event-based sensors through digital connection to neural chips or digital hardware. Event-based visual sensors use the loading bit to differentiate between silent and visual events; however, the loading bit can also be used to send other types of information to postsynapse targets, potentially to calculate more advanced functions than the fire integration method or integrate and fire used. The reason for researching SNNs is that, in real-world activities, the brains display considerable cognitive function. With continuing efforts to enhance our perception of brain-like calculations, models closer to biology are closer to achieving human intelligence than more abstract, or at least more computationally effective, models [30].

In this way, SNN methods are ideally appropriate to process the space-time information based on neuro sensors, which are themselves energy-efficient. Sensors collect precise environmental information, and SNNs can use some useful time codes for their calculations. This information processing is also the focus of the event, which is denoted whenever

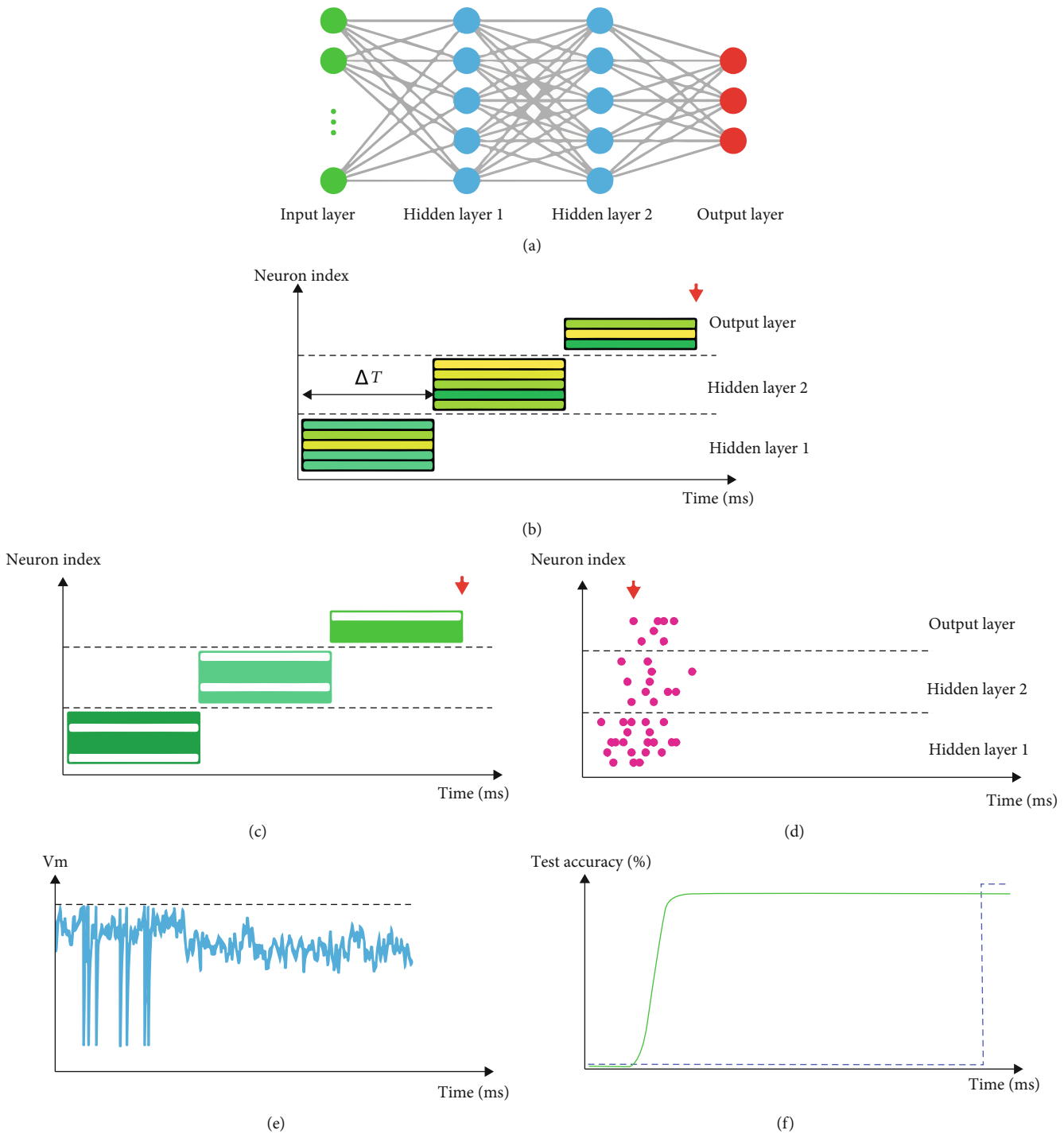


Figure 1: General structure and mechanism of a spiking neural network or SNN [30].

a small amount of information is not recorded in the SNN; it does not do much calculation, but the SSN creates more spikes when an activity explosion is recorded. It leads to a very efficient way of calculation, assuming that information from the outside world is usually scattered. Also, time-domain input is another precious piece of information compared to framework-based approaches, where an artificial timeline is introduced entered by the sensor. It can result in an effective calculation features such as optical current or stereo inequality combined with spike-sensitive learning rules. In deep SNNs, asynchronous axis-based computing mode results in the rapid dissemination of prominent information through multiple network layers. In practical terms, SNNs must be run on neuromorphic hardware to take advantage of this effect. This process is a quasisimultaneous data processing combined with an event-based sensor, which implies that after the first input spikes are registered, the first estimated output of the final layer is immediately available. Also,

for multilayered networks, it is right as the spikes extend immediately to the higher layers as soon as enough activity is generated by the bottom layer. You do not have to wait to complete of the complete input series, which is unlike traditional deep neural networks, where it is important to completely charge all layers until the final output is calculable. The primary performance spikes are inevitably based on incomplete data. It was thus concluded that deep SNNs would increase their efficiency in classification and decrease the processing time of the spike more than their input. To decrease the expected delays in inference, SNNs can also be specifically fitted. SNNs are the computational model chosen to run highly energy-efficient neuromorphic hardware devices supporting a data-driven processing mode and maintaining local calculations, thus prohibiting access to expensive memory [30].

Here, despite recent advances, one of the major deep SNN disadvantages is their accuracy in standard metrics such as MNIST, CIFAR, or ImageNet which is not as good as that of their machine learning counterparts. The existence of the benchmarks present in traditional frame-based images can perhaps be attributed to this, to some extent. A sort of conversion of the picture to the Spark sequence that is typically inefficient is therefore required. The lack of training algorithms that take advantage of Spark neurons' features, such as efficient timescales, is another limiting factor. In contrast, several approaches employ many approximations according to the rate of use of convolutional deep learning neural networks, denoting that no progress can be expected. Deep SNNs may be practical in these cases and maybe faster, in which they get more efficient than convolutional systems, where SNN runs on diagonal neural hardware. For SNNs, the training algorithms are difficult to analyze due to their noncomputational and discontinuous computational methods, which generated direct use of successful techniques behind the scenes, especially for deep neural networks be difficult [30]. In traditional AI standards, the performance of SNNs should only be considered as concept proof, but not as the ultimate research goal. If biology is the model of spike networks, it can be concluded that they are designed for behavioral tasks such as making decisions based on continuous current input when moving in the real world. Whereas brains may solve these things, they are certainly not optimal for it. Recently, the Internet environment lacks good metrics and evaluation metrics that can measure effective performance in the real world [30].

3. Proposed Approach

The preprocessing phase of the proposed approach is aimed at reducing the initial noise. In the following, the operation of segmentation and extraction of features is aimed at distinguishing tumor masses from the data set. The preprocessing section applies a method called Quantum Matched-Filter Technique, followed by a CRF-based QAIS-DSNN combination approach.

3.1. Preprocessing Phase. Initially, there will be a preprocessing phase involving noise reduction. Every single image is displayed in a combination of local threshold and active contouring using a two-dimensional array of pixels; their values are integers in the range of [0,255]. Local thresholds initialize images in two steps. First, the input noise image is considered the primary image to which image noise removal will be applied. This operation is mainly utilized as a local search operator to enhance the initial images, using the Quantum Matched-Filter Technique (QMFT). The use of local thresholds and active contours has been used in this paper because they are computationally faster than other methods in the literature. Thus, at the end of the first step, there will be a decomposed image. In the second step, thresholding is done on the detail coefficients, and one of these decomposed sections is randomly selected and sent to a reconstruction operation. The reconstruction section can be defined:

(i) *Gaussian blur*: uses a Gaussian filter to filter the image. Between 3×3 pixels and 5×5 pixels, the filter size is accidentally selected

(ii) *Mean filter (averaging filter)*: filters the image using an average filter

(iii) *Intensity change*: all image pixels are multiplied by a similar criterion randomly selected in the range [0.7, 1.3]

(iv) Implement light-intensive sections in quantum and reverse processing that performs the QMFT

Then, the following operations are performed:

(i) *One-point row*: a pixel row is chosen randomly

(ii) *One-point column*: it is identical to the preceding form, but instead of a row, the column is considered

(iii) *Point-to-point random*: accidentally, every pixel is selected from the decomposition until a new image is created

(iv) Identify all points in a row and column in the image to reduce the majority of noise as QMFT

After analysis, when the selected range value [0.1] is less than the local search rate in the QMFT, a new image of the local search operator may pass. As the decomposition is complete, the entire image is sorted by its pixel value. Then the best aspect ratio in the image is considered as a quantum value in the sequel. A signal in MRI images may be broken down into multiple displaced or resized displays of features known at the feature extraction stage. Local thresholds and active contours can be used to analyze an image into its components. It is possible to perform image segmentation operations after applying QMFT along with local and active contouring thresholds. In this case, the local threshold coefficients and the active contour based on QMFT can be destroyed to eliminate some details. Local thresholds and QMFT-based active contours have a tremendous advantage in separating fine detail in an image. Active contour can be used to isolate very fine details of an image. At the same time, local thresholds can detect large details, combining fine and

large details, and reading all rows and columns linearly and diagonally. Quantum satisfies QMFT to minimize the noise in the MRI image. QMFT based on local thresholds and active contours can create a sparse display. A local and active contouring threshold function with QMFT has two main features, the first of which is a function of oscillation or wave appearance, such as

$$\int_{-\infty}^{0} \left| \Psi(t) \right|^2 dt < \infty. \tag{1}$$

In this case, most of the energy in Ψ (t) is limited to a limited time, which is in the form of

$$\int_{-\infty}^{0} \Psi(t) dt = 0. \tag{2}$$

The proposed method is generally calculated to reduce the noise in

$$\Omega(I) = \left(\sum_{\Omega} \sqrt{1 + \beta^2 |\nabla I|^2} \right) + \frac{\lambda}{2} (I - I_0)^2. \tag{3}$$

In Equation (3), the term $(I - I_0)^2$ ensures a certain degree of validity and accuracy between the rated and original image, in which I denotes the rated image while I_0 means the noisy image. The ∇I parameter is defined as the sum of the variable adjustment periods, β and λ are the balancing parameters, and Ω is the sum of the image's points. The purpose of minimizing Equation (3) is to decrease total image diversity while maintaining accuracy and validity. The balancing values are changed from 1 to the size of the image for both β and λ to minimize Equation (3).

3.2. Segmentation with QAIS-DSNN Combined Approach. The deep spiking neural network presented in this study, due to its high flexibility, can use a linear and nonlinear functions such as sigmoid or sinusoidal in hidden layers. Use nonderivative as well as intermittent activation. By default, DSNN has

$$y(p) = \sum_{j=1}^{m} \beta_i \beta_j g \left(\sum_{i=1}^{n} w_{i,j} x_i + b_j \right). \tag{4}$$

According to Equation (4), β_i displays the weights between the input and the hidden layers, and β_j displays the weights between the output and the input layers (b_j). The value of the neuron threshold is in the hidden layer or the bias. $g(.)$ is an activator or stimulus function. The weights of the input layer, $w\,(i, j)$, and bias, b_j, are randomly assigned. The beginning of the neuron number on the input layer n and the neuron number on the hidden layer m is assigned to activation function $g(\cdots)$. If the known parameters in the general equilibrium are combined and controlled on the basis of this information, the output layer will be similar to

$$H(w_{i,j}, b_j, x_i) = \begin{bmatrix} g(w_{1,1}x_1 + b_1) & \cdots & g(w_{1,m}x_m + b_m) \\ g(w_{n,1}x_n + b_1) & \cdots & g(w_{n,m}x_m + b_m) \end{bmatrix}, y = H\beta. \tag{5}$$

The main goal is to minimize errors as much as possible in all models of training-based algorithms. The y_p the output error function is obtained by the actual y_{main} output in DSNN, which can be done with two training sections, $\sum_k^s (y_{\text{main}} - y_p)$ and the test section, $\| \sum_k^s (y_{\text{main}} - y_p)^2 \|$. The output y_p generated by the real output, y_{main}, must be identical with the same y_p for both functions. An unknown parameter is specified when this equation is performed and the results are satisfying. While spikes have been used to understand local label dependencies, for medical images such as MRI, they are not appropriate. Typically, that is because anatomical forms have complex shapes for models that are distinct.

Moreover, either the temporal or the spatial relationship of MRI data also plays a critical role in classification, which should be paid attention to with regard to the method. Therefore, it is better to modify the mapping of the probability achieved by DSNN. A rather low-probability matrix may be the H-matrix, which means that the amount of data in the training process will not be identical to the total number of data characteristics. But it would be a big challenge to reverse $[H]$ and find weights or β. A fully connected CRF matrix is used to overcome this challenge in DSNN, which can develop an approximate reversal of the matrix that cannot be reversed. It can reduce the size, selection, and extraction of features at the segmentation with high precision and incredible speed compared to other methods. Currently, CRFs have been implemented in many medical imaging applications because they perform well when modeling some complex spatial data dependencies. In this way, to segment brain tumors, CRFs can be used not only to model the relationship between an image pixel and poster properties but also to make local pixel properties and their labels dependent. As discussed earlier, in [11] and [26], CRFs were employed to visualize images through image formulation as neural networks. Nevertheless, the process of training their method is cumbersome and mathematically complex. In contrast, CRFs will be utilized as a suitable hash method. Using the fully connected matrix and CRF layer, the output matrix β^* and the matrix H^* are all inverted and generalized by H. Therefore, due to the improvement of DSNN as CRF-DSNN in this section, the problem of the output weights in DSNN has been resolved and converted to $B^* = H^*$. In general, CRF-DSNN becomes a series of repeating units over time in the training phase. CRF-DSNN will be able to act as a belt conveyor and add or subtract information to neurons. Unlike deep learning structures and other classification models, such as backup conveyor machines or nanoscale works, no weight update is performed during training. CRF-DSNN can define features

at the segmentation. By reducing CRF energy performance, a suitable model is taught that can be modeled as

$$E(Y) = \sum_{i}^{N} \Psi_u(y_i) + \sum_{\forall i,j,i\neq j}^{N} \Psi_p\left(y_i, y_j\right), \tag{6}$$

where $u, p \in \{1, 2, \cdots, C_n\}$ are the designations of the segmentation and $i, j \in \{1, 2, \cdots, N\}$ properties are specific pixels of the original image or I. $\Psi_p(y_i) = -\log P(y_i \mid I)$ is the negative logarithmic probability where $P(y_i \mid I)$ is a probability obtained by DSNN per pixel i. While measuring the capabilities of a matrix pair of CRFs in a fully connected layer, it deals with the relationship between each pixel that is defined as

$$\Psi_p\left(y_i, y_j\right) = \mu\left(y_i, y_j\right) \sum_{m=1}^{M} w^{(m)} k^{(m)}\left(f_i, f_j\right), \tag{7}$$

where $M = 2$, the number of Gaussian nuclei and $w^{(m)}$ indicate a weight for the Gaussian nucleus mth, and $\mu(y_i, y_j) = [y_i \neq y_j]$ is the label of consistent function. $k^{(1)}$ displays the core appearance, which tries to assign the same class labels to neighboring and adjacent pixels with the same intensity. $k^{(2)}$ displays the kernel smoothness, which is associated with eliminating unnecessary areas. These two steps are shown as

$$k^{(1)}\left(f_i, f_j\right) = \exp\left(-\frac{\left|s_i - s_j\right|}{2\theta_\alpha^2} - \frac{\left|e_i - e_j\right|}{2\theta_\beta^2}\right), \tag{8}$$

$$k^{(2)}\left(f_i, f_j\right) = \exp\left(-\frac{\left|s_i - s_j\right|}{2\theta_\gamma^2}\right). \tag{9}$$

e_i and e_j are the light intensities of the pixel i and j and s_i and s_j are the corresponding spatial coordinates. f_i and f_j mean the characteristics of each pixel pair, i.e., the brightness intensity and spatial information. θ_α, θ_β, and θ_γ show the parameters of the Gaussian nucleus, respectively. However, some points in the mass may not be segmented in this way, so this algorithm optimization will be done in layers. In general, the DSNN method's layers are the use of the input layer with the number of neurons (spikes). Then, the structure of the training and testing layer used convolution, pooling, and fully connected layers along with CRF. Then, a SoftMax layer is embedded for it and then an output layer to display the work. The training layer window is in the form of a matrix, 9×9 in the convolution layer, 7×7 in the pooling layer, and 5×5 in the maximum section (Maxpool). The structure of the fully connected layer is 9×9. The SoftMax layer is also 7×7.

The Quantum Artificial Immune System (QAIS) is used to optimize the segmentation process during high-altitude neural network training in the SoftMax layer section. The QAIS uses a factor called an antigen. In an MRI image, all antigens are detected through a memory-based adult detection system, which has a fault tolerance experiment with a choice of the colon and immune mutations. Colonial choices and immune mutations are the other two factors of the QAIS algorithm. The more MRI data, the more copies are duplicated. In this algorithm, reproduction is plural, especially like a crossover in the genetic algorithm [12, 13]. Antibodies focus on modern quantum memory detection systems in mass segmentation in real time and examine detection and cross-sectional states against the MRI image structure.

The display of MRI image data is performed by a set of antigens $\text{Ag} = \{\text{ad} \mid \text{ad} \subset S\}$, in which the antigens determine the ad. They display one bit of binary string bits' properties that are represented by MRI image data antigens. These bits contain Trait codes. Also, S is the spatial state in the QAIS that is presented by $S = \{0, 1\}^l$ and displays all the activities of the primary population in the image in segmentation. l is the natural state number in the QAIS algorithm, which is considered as a constant value. There are two states of self-adjusting and non-self-adjusting in the artificial immune system algorithm. The self-adjusting state ($\text{self} \subset \text{Ag}$) displays all MRI image data and the non-self-adjusting state ($\text{nonself} \subset \text{Ag}$) displays all the segmented data. Therefore, there is a relationship between self-adjusting and non-self-adjusting states, represented by $\text{self} \cup \text{nonself} = \text{Ag}$ and $\text{self} \cap \text{nonself} = \varnothing$ equations.

The safety diagnostic set is also $D = \{\text{ab}, p, t, \text{age}, \text{cnt} \mid \text{ab} \in S, p \in R, t, \text{age}, \text{cnt} \in N\}$, where ab is the antibody, p is the concentration of the antibody, t is the tolerance of error, age indicates the age of memory and the maturity of genes, R is a set of real numbers, and N is the case number natural genes. Memory detection set $M_d = \{d \mid d \in D, d \cdot \text{cnt} > \beta\}$ and the gene recognition maturity group are shown as $T_d = \{d \mid d \in D, d \cdot \text{age} < \lambda, d \cdot \text{cnt} < \beta\}$. There is also immaturity, which is defined as the immaturity of genes that are expressed as $I_d = \{d \mid d \in D, d.t < \alpha\}$. In these relationships, $D = M_d \cup T_d \cup I_d$, where α represents the threshold for error in detecting immature status, λ represents the gene life cycle, and β represents the threshold value for detecting gene maturity.

In order to establish and evaluate the structure of diagnostic development, an immature gene detector becomes the mature state detector, which will be successful in the fault tolerance phase. When the adaptive time between the adult gene detector in the gene's life cycle and the antigens activated exceeds the β threshold, the adult detector clones or collects itself and then evolves into a memory detector. It means that genes and antigens will have a memory. Once the antigens are recognized by a specialist, he/she assembles the mature diagnostic compound. To ensure that antigens are effectively detected and that a variety of antibodies are detected in the reagent (mature or immature), they will detect known or unknown attacks. A total of three operators are used for the QAIS algorithm to improve the transverse distribution of MRI image data, which includes dependency assessment, reproduction selection, or safety and mutation combination, which are described separately.

Hamming distance is used to compute the correlation for antigen detection. For example, the error tolerance mode is considered to create a model of correlation assessment. An unsuccessful identifier can succeed, if the immature identifier has never been compared to all elements of the self-

organizing group in the α variable. On the contrary, it can lead to the death of genes and antigens. The $s \in$ self is assumed, and Equation (10) shows how id is determined by the s.

$$f_{\text{match}}(s, \text{id}) = \begin{cases} 1, & \dfrac{f_{\text{affinity}}}{l_d} > \gamma, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

According to Equation (10), 1, 0 indicate whether the id is compatible with s and l_d is the size of the id detector, so f_{affinity} is used to calculate the correlation between s and id. Likewise, γ as $\gamma(0 \le \gamma \le 1)$ represents the correlation threshold. Equation (11) is used to implement the mature error detector of the immature id, and Equation (12) is used to add the self-enforced identifier time when the results of the Equation (11) return to 1, and if $t \ge \alpha$, the immature identity must develop into a mature

$$f_{\text{tolerance}}(s, \text{id}) = \begin{cases} 0, & \exists s \in \text{self}, f_{\text{match}}(s, \text{id}) = 1, \\ 1, & \text{otherwise,} \end{cases} \tag{11}$$

$$\begin{cases} T_d = T_d \cup \text{id}, I_d = I_d - \text{id}, \text{if id} \in I_d, \text{id} \cdot t \ge \alpha, \\ \text{id} \cdot t = \text{id} \cdot t + 1, \text{if id} \cdot t < \alpha \wedge f_{\text{tolerance}}(s, \text{id}) = 1. \end{cases} \tag{12}$$

The colonial or combination choice operator performs cellular operators in mature and memory diagnosis. Equation (13) is used to detect cloning state and a mixture of genes and antigens.

$$C_{\text{num}}(d) = \left\lceil \xi \cdot \left(1 - \frac{n_d}{N_d}\right) \right\rceil. \tag{13}$$

According to Equation (13), ξ (>0) is a colonial or combination constant. $N_d = T_d \cup M_d$ shows all the combinations. The colonial determinant or combination factor is used to analyze the performance of cellular operators in mature and memory diagnosis. Equation (13) is used to detect the cloning state and a mixture of genes and antigens. In Equation (14), T_{cln} and M_{cln} display the colonic selection group or group of memory and mature detectors. After making a colonial selection or group of genes and antigens in a generation, the cloned or combined section is added to the adult diagnostic group, and the same detector $d_t(\in T_d)$, in the colonial selection group, or the T_{cln} and M_{cln} combination will be removed.

$$T_d = T_d \cup T_{\text{cln}} \cup M_{\text{cln}} - \{d_t \mid d_t \in T_{\text{cln}} \vee d_t \in M_{\text{cln}}\}. \tag{14}$$

The goal of the immune mutation operator is to enhance the detector's diversity with the mutation of the antibody generation in the corresponding detector, which is used to improve the ability to detect antigens. Considering the $(l_d - f_{\text{affinity}}(d, \text{ag}))$ bit, the $d(\in N_d)$ detector set is matched by the ag($\in$Ag) antigen; these bits are used by 0.1 instead of randomly. l_d displays the size of d. The mutant detector is used as an immature detector by the self-regulating set. To detect the adult mode, if the adaptive time is greater than the activated threshold β, the stimulus operation is performed using Equation (13) and then combined with the memory detector according to

$$M_d = M_d \cup \{d \mid d \in T_d, d \cdot p = \eta_1, d \cdot \text{age} = 0\}. \tag{15}$$

Equation (15) is assumed to represent the arranged numbers of the reagent that can be matched with antigens. Therefore, the memory diagnosis segment is combined with Equation (16), but this occurs when the memory diagnosis segment can be successfully matched with antigens.

$$M_d = \{d \mid d \in M_d, d \cdot p(t) = \eta_1 + \eta_2 \cdot d \cdot p(t-1), d \cdot \text{age} = 0\}. \tag{16}$$

Equation (17) also illustrates a different type of antigen removed in the MRI image data for display.

$$d \cdot p = \begin{cases} d \cdot p\left(1 - \dfrac{1}{\theta - d \cdot \text{age}}\right), & d \cdot \text{age++} < \theta, \\ 0, & d \cdot age\text{++} \ge \theta. \end{cases} \tag{17}$$

Intensity and variety are two important features of swarm intelligence algorithms. The intensity is in the search of the best-obtained solutions and choosing the best candidate points. It is worthwhile to mention that the diversification procedure can allow the optimizer to explore the search space more efficiently. Inertial weight parameters (w_n, w_f) indicate changes in optimal global attractiveness that affect the convergence rate and update each mass's position in the combination algorithm QAIS-DSNN. In the proposed QAIS-DSNN hybrid algorithm, the inertial weights (w_n, w_f) are set to a large value to emphasize exploration, i.e., 0.9, which are set in the initial search mode, finally reduced to 0.1 linearly for the importance of linear optimization. Inspired by the classic artificial immune system, it is guaranteed that quantitatively, global characteristics for optimal segmentation can be determined when using the spiking neural network. As the number of repetitions increases, the initial population is encouraged to local search. Finally, the population should only carefully search for a local area without discovery to find out if there are any other masses. As a result, the first quantum combination strategy is to provide a linear weight reduction of the new frequency. The model (w_n, w_f) is created as

$$w_{\text{total}} = 0.9 - \left(\frac{0.9 - 0.1}{\text{MI}}\right) \times J, \tag{18}$$

where $[0.1, 0.9]$ is the inertial weight range and MI is the maximum number of repetitions, in which J denotes the number of repetitions. As such, both w_n and w_f are linearly decreased from 0.9 to 0.1 in the repeat cycle. The developed mixed QAIS-DSNN algorithm may be trapped in local improvement due to the presence of different iterative cycles

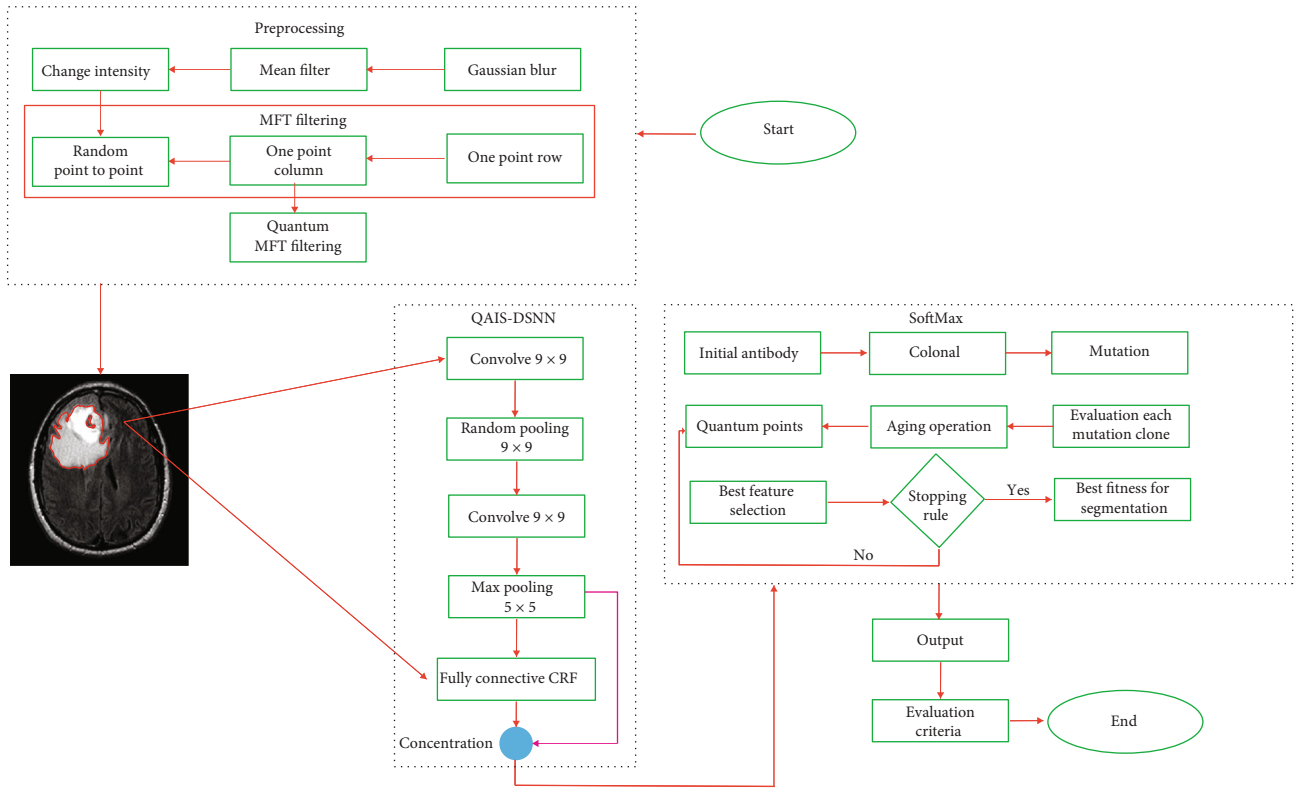


Figure 2: Flowchart of the proposed approach.

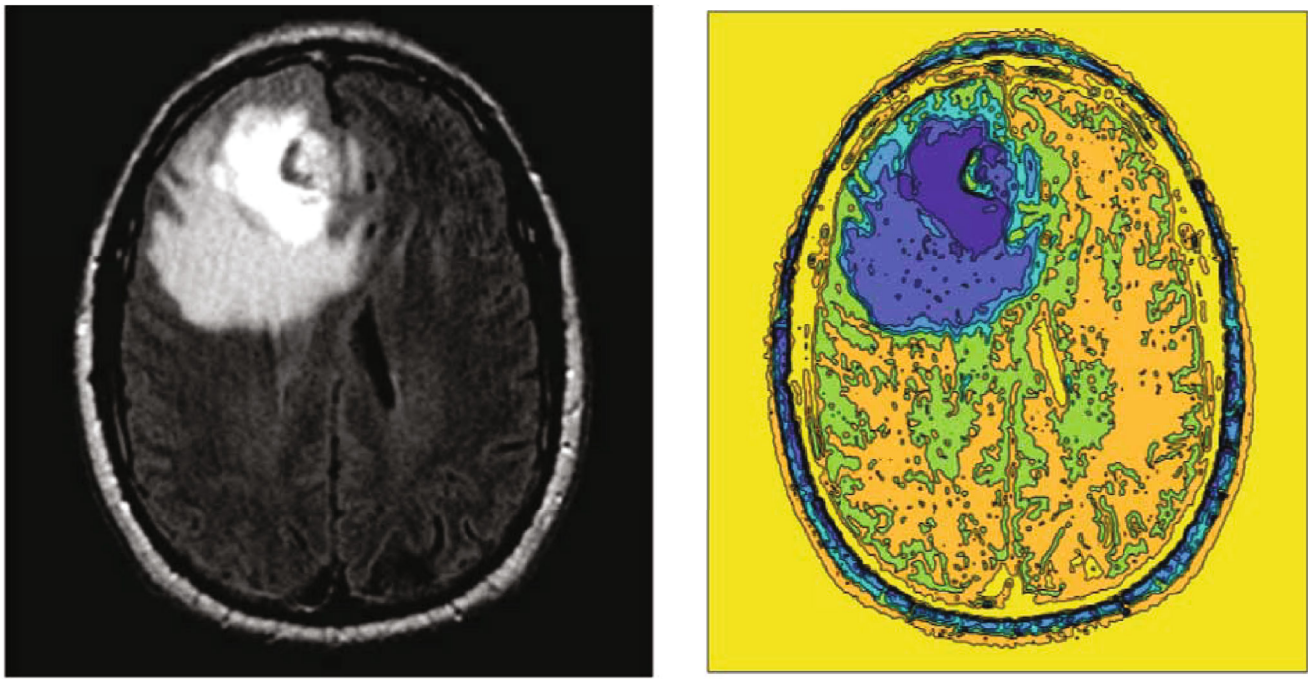


Figure 3: Input image.

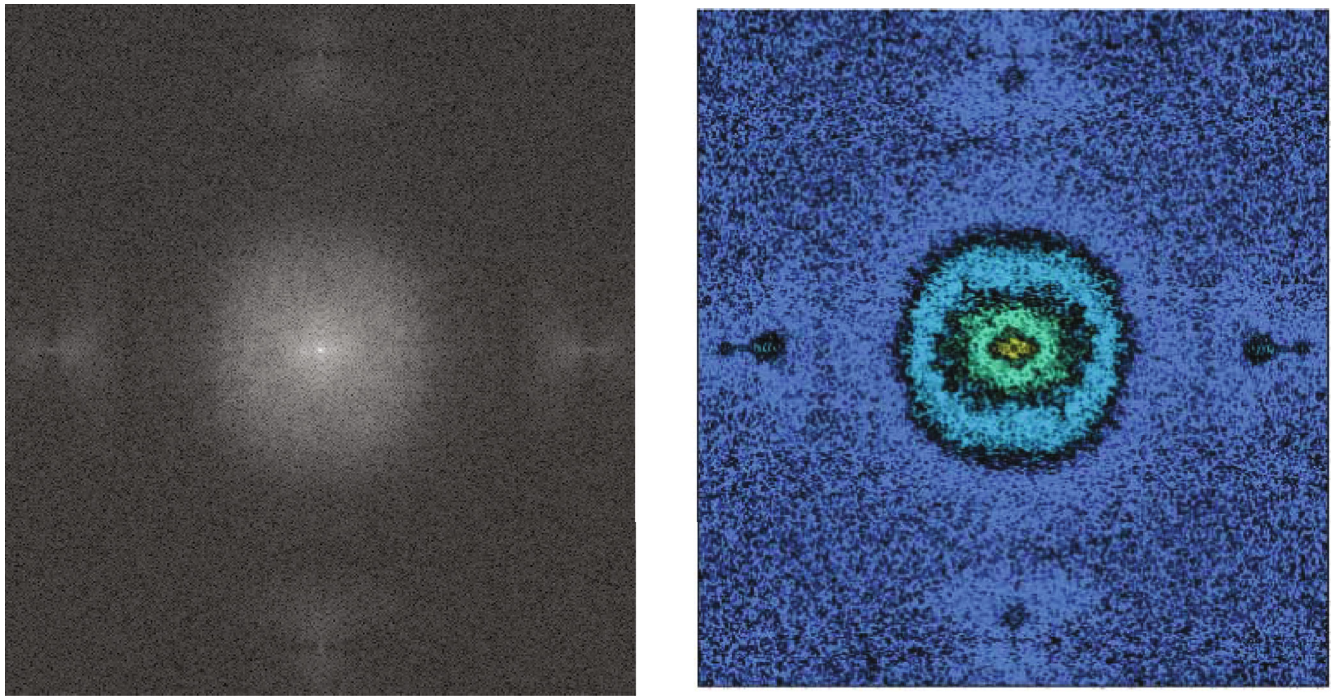


Figure 4: QMFT noise reduction algorithm applied in a row, column, and diagonal without repetition.

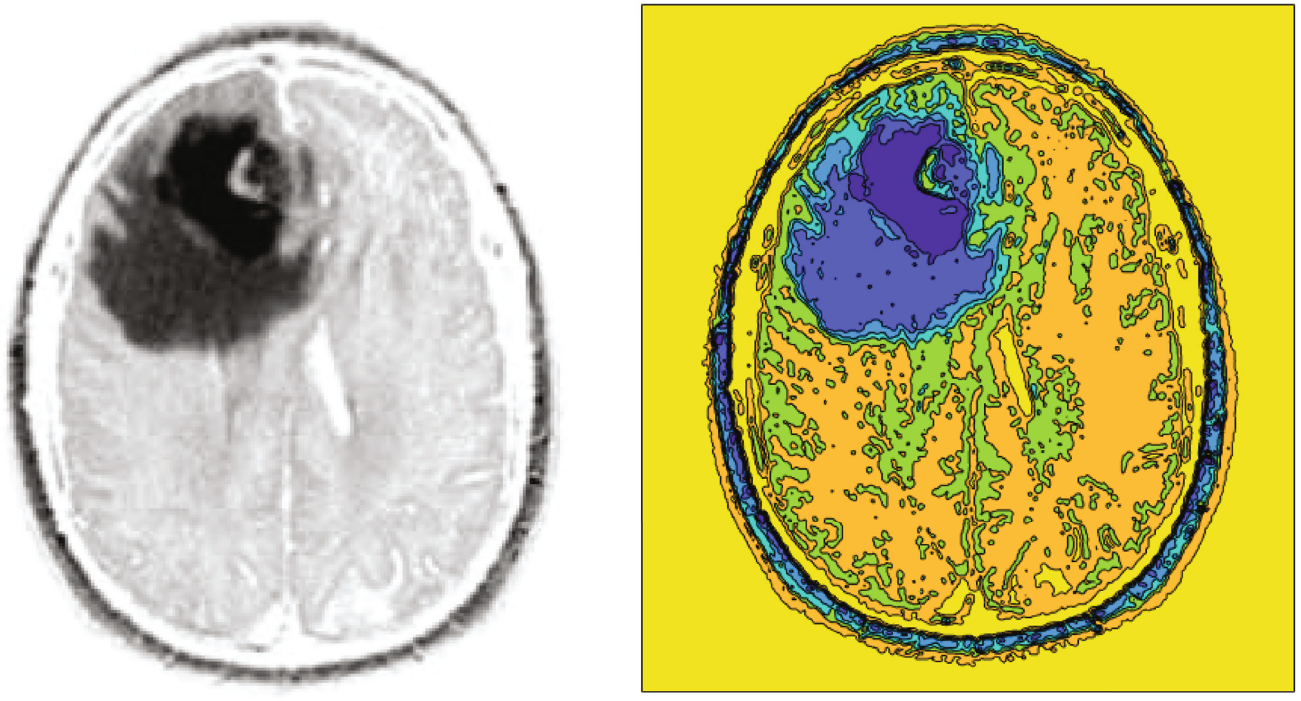


Figure 5: The result of image noise reduction and highlighting.

in the tumor fractionation improvement process, in addition to the high research capacity. Therefore, to solve this problem, it is possible to provide a comparative update strategy for the C^{best} parameter that is best to assist neurons and primary residents of the proposed algorithm out of the optimal local areas. In this strategy, C^{best} is considered the best at a great value in the initial phase of finding the optimum value of the QAIS-DSNN algorithm with strong exploration ability (global search) and gradually decreasing with increasing frequency for accurate searching.

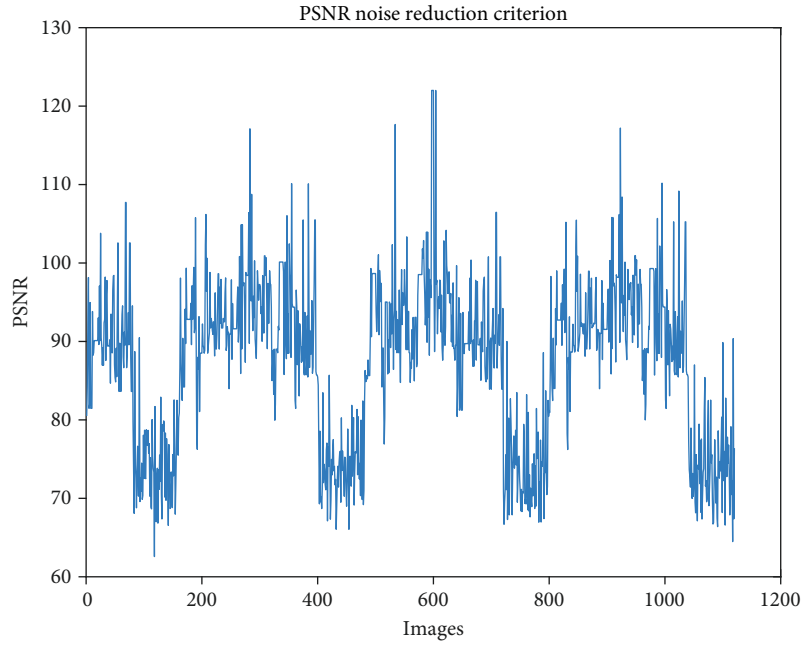


Figure 6: The PSNR criterion for noise reduction measurement.

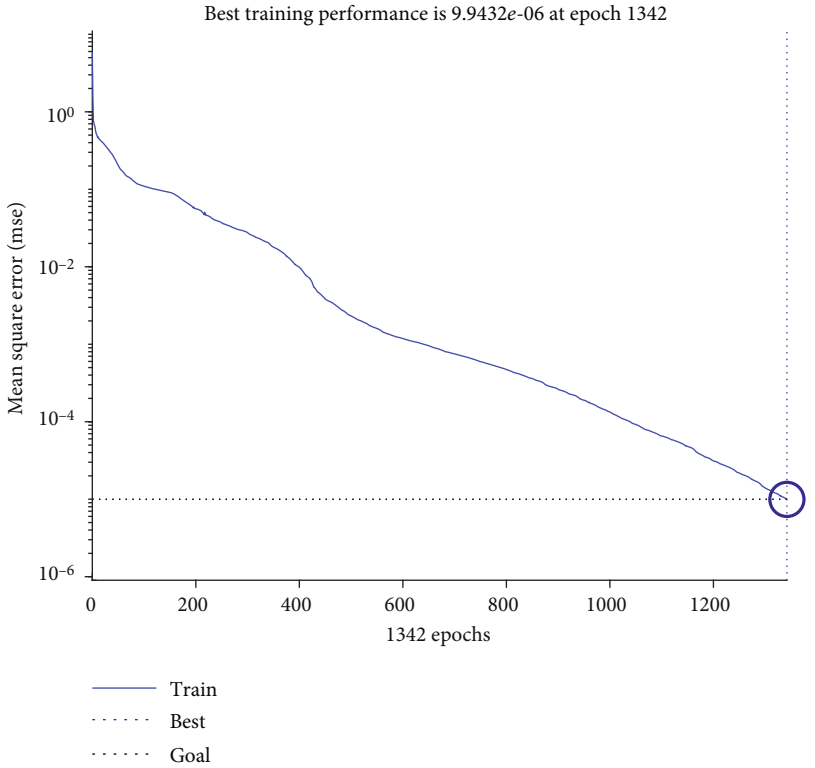


Figure 7: Mean square error of the training process for 1342 epochs.

Equation (19) displays the optimal value for the better adaptive update scheme C^{best} used here.

$$C^{\text{best}} = 2 \times 1 - \frac{J}{\text{MI}}, \tag{19}$$

where J means the number of repetitions, whereas MI denotes the maximum number of repetitions. The next step is to introduce a novel method for updating the neurons and the initial population to accelerate global convergence. Initially, the status vector is updated by

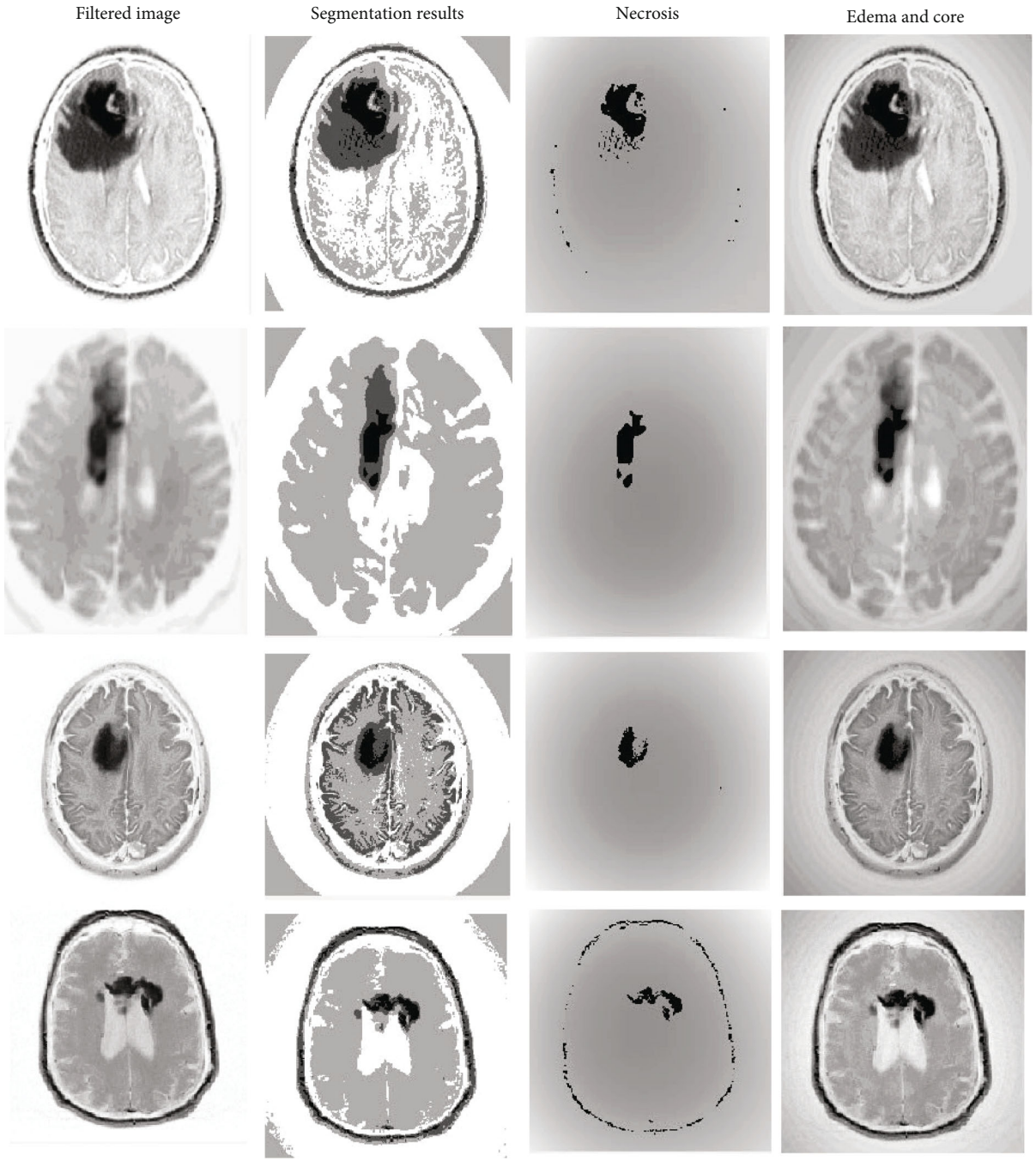


Figure 8: Display of training operations to test the segmentation of the image and identify the masses. From left to right: the overall result of noise reduction as input in QAIS-DSNN and QAIS-DSNN approach segmentation, finding necrosis, and finding edema (part completely black) and nucleus (the almost white part inside the edema).

$$\Delta t = C_t \sum_{j-1}^{\mathrm{NV}} \left(\mathrm{UB}_j - \mathrm{LB}_j\right), \tag{20}$$

where NV is the total number of variables, UB_j is the upper limit, and LB_j is the lower limit of the variable in the j^{th} variable. C_t is the search environment, the same as the main input image. Randomization is then performed that prevents trapping in the optimal local solution, so randomization is introduced in Equation (21) with the value of α_{rand}, which is a randomization parameter.

$$X_i(t+\Delta t) = X_i(t+\Delta t) + \left(X_{\mathrm{gbets}} - X_i(t+\Delta t)\right) + \alpha_{\mathrm{rand}} \times \left(\mathrm{rand} - \frac{1}{2}\right), \tag{21}$$

where X_{gbets} is the position of each neuron, and the initial population of the combination approach and rand is a

random number generated, represented as a uniform distribution in the range of [0, 1]. In general, the flowchart of the proposed approach is shown in Figure 2.

3.3. Investigating the Computational Complexity of the QAIS-DSNN Method. Here, the computational complexity of the developed QAIS-DSNN algorithm is investigated. Computational complexity includes temporal and spatial complexity. The time complexity of the QAIS-DSNN algorithm depends on two steps including calculation of the motion and updating of the positions of the neurons and the initial population. Therefore, the complexity of time can be defined in

$$O(\text{QAIS-DSNN}) = O(t(O(\text{intensity neighbor edges}) + O(\text{position update}))), \quad (22)$$

$$O(\text{QAIS-DSNN}) = O\left(t\left(n^2 \times d + n \times d\right)\right) = O\left(tn^2 d + tnx\right) = O\left(tn^2 d\right), \quad (23)$$

where t is the maximum repetition cycle, n denotes the number of neurons or initial population, and d is the dimensions of the problem.

4. Simulation and Results

BraTS data is a collection of brain tumor MRI images, including 145 folders for patients under different conditions. The dataset consists of 4 versions from 2012 to 2018. Database versions are getting better every year. The primary data is in DICOM format which have been converted to JPEG format for easier use through DICOM Viewer software. The input images are three-dimensional. Due to the large size of the images in the BraTS, we used 1000 video input samples to study the proposed approach. The simulation will be done in MATLAB 2015b environment and a system with 7-core processor specifications with 6 MB of cache and 3.6 MHz and 6 GB of memory in Windows 10. When the simulation is performed, all BraTS2018 data are trained and tested by the proposed method. For visualization, an example of images is shown to examine the proposed approach's results, step by step. Initially, the input image is given to the system, as shown in Figure 3.

In all BraTS2018 data images, an initial noise reduction is required, using the QMFT algorithm, in which the image is read linearly, columnar, and diagonally without any repetition to reduce noise. The schematic of this output is in the form of Figure 4, and the result of the image that the noise reduction operation and its initial highlighting is in Figure 5.

The value of peak signal-to-noise ratio (PSNR) is illustrated in Figure 6. The analysis is done for 1120 MRI images. The mean value of PSNR is 87.35. It can be seen that the noise reduction provides an interesting picture in which a good segmentation can be applied. For this purpose, a deep neural network spiking or DSNN method is applied to the noise reduction operation output, and the BraTS2018 video data set is trained, which will be 75% training and 25% test. But the combined DSNN approach with the QAIS algorithm is

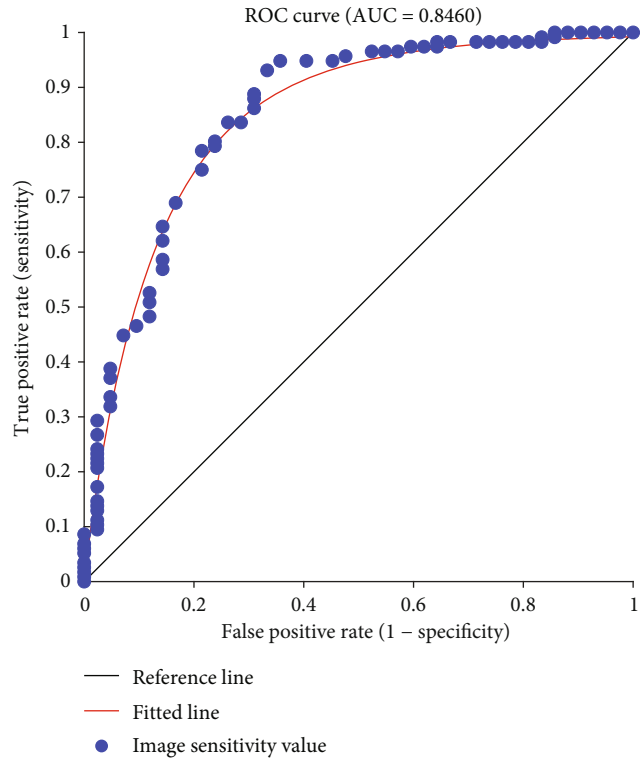


Figure 9: ROC diagram and AUC rate.

made in this section so that the overall result is visible. According to the DSNN structure, it is observed that five input layers are considered, in which all BraTS2018 video data are placed. Then, there are three rows of training layers, the first of which is the training deep layer. In this row, from the deep layer, one by one, the convolution layer with 9×9 windowing, and then the random polarizing layer with 7×7 windowing, again the convolution layer with 9×9 windowing, and then the maximum polarization layer as 5×5 is located. The stimulus function of this layer is a zygomatic logarithm in that the number of general layers is 20. Then, the fully connected layer is associated with CRF, which is considered as 10 layers.

Then, there is a SoftMax layer with the QAIS algorithm designed to optimize DSNN segmentation during training and testing, more accurate mass detection, and feature selection operations. Its drive function is linear. There are separate settings for the QAIS algorithm. The initial population of this algorithm is considered to be 200. The colonial rate is 0.04, and its repetition rate is 10 cycles for optimizing the DSNN algorithm segmentation and selecting features in the SoftMax layer. In the end, there is an output layer that is a layer to display the output. The number of raw data training and testing rounds in QAIS-DSNN is equal to 7000 rounds. The QAIS-DSNN core is resilient back-propagation, and its performance is measurable with average error squares. The training process is illustrated in Figure 7. The termination criterion for the training process is mean square error as 10^{-5}. Regrading Figure 7, 1342 epochs lead to converge the training process. When the proposed approach is applied to multiple images, the overall result will be Figure 8.

Table 1: The comparison of the proposed approach with previous ones in terms of mean square error, peak signal-to-noise ratio, and accuracy.

The peak signal-to-noise ratio (decibels)	Mean square error	Method	Reference
72.1	0.23	KNN	Mittal et al. [25]
88.4	0.045	Genetic algorithm	Mittal et al. [25]
74.2	0.021	SVM	Mittal et al. [25]
73.5	0.256	SOM	Mittal et al. [25]
94.2	0.012	CNN	Mittal et al. [25]
96.64	0.001	GCNN	Mittal et al. [25]
77.7	0.001	BAT-IT2FCM	Alagarsamy et al. [31]
97.79	0.006	QIAS-DSNN	Proposed approach

Table 2: The comparison of the proposed approach with previous ones in terms of Dice evaluation criteria.

Tumor improvement or ET areas	Tumor nucleus or TC	Total tumor or WT	Method	Reference
81.84%	88.34%	91.2%	ADNN-PSO	Irfan Sharif et al. [32]
85.83%	79.72%	90.21%	3D cascaded CNN-TTA	Wang et al. [27]
79.19%	85.40%	90.31%	Cascaded CNN	Wang et al. [27]
77.07%	73.04%	89.56%	Multiclass WNet+TTA	Wang et al. [27]
71.78%	74.81%	88.24%	MCCNN	Hu et al. [28]
72.29%	76.75%	86.23%	Two-stage	Zhou et al. [26]
70.9%	75.1%	85.1%	Ordinary fusion	Zhou et al. [26]
73.44%	76.58%	86.38%	3D UNet	Zhou et al. [26]
72.55%	75%	84.94%	APFNet	Zhou et al. [26]
74.43%	76.88%	86.56%	APF+3D-CRF	Zhou et al. [26]
74.50%	80.15%	91.92%	QAIS-DSNN	Proposed approach

The ROC chart and the AUC rate are the proposed approach in Figure 9. This curve is known as one of the most important evaluation criteria, which measures the efficiency of classification operations in a system. In general, in a binary classification system in which the differentiation threshold differs, the ROC curve is a graphical representation of the degree of sensitivity or correct prediction versus false prediction. The ROC curve is also shown by plotting the correct positives against the predicted false positives. A number which measures and evaluates an aspect of performance is the area below the ROC curve. This area below the curve is called the AUC. A value above 0.7 to 1 indicates an excellent level of prediction and classification performance. According to Figure 9, it is observed that the value of AUC is a number below one, which shows the optimization of the proposed approach as much as possible. The presence of some similar sections with cancerous masses in the available data and presented method led to the creation of a series of minor errors that have not been adapted to the fitting line. The blue circles are the criterion values, and the red line is the ROC diagram on which the data is fitted. In some areas where the data is a bit far away, an error occurs and leads to a decrease in inaccuracy. Also in the middle line is regression called the ROC peak relative to regression, and the area below it is AUC. After applying the proposed approach, it is necessary to compare the proposed approach with other proposed methods, which are examined in terms of different evaluation criteria to determine the guarantee of the proposed approach. For this purpose, Table 1 shows a comparison in terms of average error squares, signal-to-noise ratio, and accuracy. Also, a comparison has been made in terms of Dice evaluation criteria for tumor nucleus, total tumor, and tumor areas, the results of which can be seen in Table 2.

Table 3: Comparison of the proposed approach with previous methods in terms of accuracy in terms of percentage.

Accuracy (%)	Method	Reference
98.20%	GCNN	Mittal et al. [25]
95%	3D cascaded CNN-TTA	Wang et al. [27]
88.50%	MCCNN	Hu et al. [28]
96.12%	BAT-IT2FCM	Alagarsamy et al. [31]
92%	ADNN-PSO	Irfan Sharif et al. [32]
98%	PSO-LDA-GA-ANN	Sharif et al. [33]
98.21%	QAIS-DSNN	Proposed approach

The next comparison is the percentage-based accuracy for segmentation to distinguish the mass region from the images, which are averaged from the BraTS data set. The results are reported in Table 3.

Finally, a comparison is made in terms of computational complexity in terms of time between the method presented in

Table 4: Comparison of the proposed approach with previous methods in terms of computational complexity over time.

Computational complexity in terms of time (seconds)	Method	Reference
0.84 seconds	GCNN	Mittal et al. [25]
1.6 seconds	CNN	Mittal et al. [25]
3.2 seconds	Genetic algorithm	Mittal et al. [25]
2.7 seconds	SVM	Mittal et al. [25]
4.2 seconds	SOM	Mittal et al. [25]
3.8 seconds	KNN	Mittal et al. [25]
2.58 seconds	QAIS-DSNN	Proposed approach

this study and the methods in reference [25], the results of which are shown in Table 4. It is noteworthy that this study has listed the system used during the processing of the proposed method, and this comparison is made on a case-by-case basis with reference [25].

Based on the results of the comparisons in terms of evaluation, it is observed that the proposed approach is optimal in terms of the mean error squares of most methods, but the GCNN [25] and BAT-IT2 FCM [38] algorithms have better results than the research approach. In terms of signal-to-noise ratio, the proposed method of compared algorithms has had better results. In terms of Dice evaluation criteria, most research is on the same level. There are differences in the parts of the whole tumor, the tumor nucleus, or the improved part of the tumor, depending on the different methods available. In terms of accuracy, the prediction approach has better results, but with the GCNN [25] algorithm, it is 0.01% more efficient. Also, the results were obtained at the level of convolutional methods, and the computational complexity of the proposed approach has been implemented in the system; however, the computational complexity can be seen by combining the existing algorithms.

5. Conclusion

This article is innovative in the field of noise reduction and segmentation of MRI images to detect the area of tumor masses. Also, we used the QMFT method to find noise and reconstruct it with adjacent pixels to process them horizontally, vertically, and diagonally. It is formed in the fastest time and has been able to move the noise by identifying and reviewing neighbors and matching the pixel data with neighbors based on the edge of the image. Then, the segmentation operation was performed with a QAIS-DSNN combination approach. In this approach, the deep neural network of spiking with CRF is considered, so that after the input layer—including neurons (spikes)—the training layer has convolution and polarizing layers. All of them are connected to the CRF format. Then, there is a SoftMax layer outside the training layer, which is optimized for segmentation and detection to accurately identify tumor features, in this method with the QAIS. The simulation results show that the proposed QAIS-DSNN approach has a functional advantage over the previous methods evaluation criteria. Among these evaluation results, we can point out the accuracy in segmentation and detection of the exact mass area in MRI images with an accuracy of 98.21%. Also, the average rate of error squares is 0.006, and the peak rate of the signal-to-noise ratio is 97.79 decibels. The use of lesion structural criteria includes a tumor nucleus of 80.15%, improved tumor of 74.50%, and a total tumor of 91.92%, which is a functional advantage over similar previous methods. Reducing computational complexity compared to previous methods and improving execution time by 2.58 seconds also confirms it. In the future, the plan is to use a huge dataset or transfer this system to the breast, lung, and some other tumor detection tasks. Moreover, we are going to add this automated segmentation method for CNN-based segmentation ground truth images. Furthermore, the presented filtering system can be added as a layer in the CNN method and change the resolution of the matrix in each iteration.

Data Availability

The data that support the findings of this study are openly available in BraTS2013, BraTS2015, and BraTS2018 at https://www.med.upenn.edu/sbia/brats2018/data.html

Disclosure

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

Conflicts of Interest

We declare no conflict of interest.

References

[1] S. Bauer, R. Wiest, L. P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Physics in Medicine and Biology*, vol. 58, no. 13, pp. R97–R129, 2013.

[2] D. N. Louis, A. Perry, G. Reifenberger et al., "The 2016 world health organization classification of tumors of the central nervous system: a summary," *Acta Neuropathologica*, vol. 131, no. 6, pp. 803–820, 2016.

[3] B. H. Menze, A. Jakab, S. Bauer et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[4] S. Bakas, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, article 170117, 2017.

[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] S. Hassantabar, M. Ahmadi, and A. Sharifi, "Diagnosis and detection of infected tissue of COVID-19 patients based on lung X-ray image using convolutional neural network

approaches," *Chaos, Solitons & Fractals*, vol. 140, article 110170, 2020.

[7] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[8] W. Mengqiao, Y. Jie, C. Yilei, and W. Hao, "The multimodal brain tumor image segmentation based on convolutional neural networks," in *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)*, pp. 336–339, Beijing, China, 2017.

[9] M. Havaei, A. Davy, D. Warde-Farley et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[10] G. Wang, M. A. Zuluaga, W. Li et al., "DeepIGeoS: a deep interactive geodesic framework for medical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1559–1572, 2019.

[11] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Medical Image Analysis*, vol. 43, pp. 98–111, 2018.

[12] M. Ahmadi, S. Jafarzadeh-Ghoushchi, R. Taghizadeh, and A. Sharifi, "Presentation of a new hybrid approach for forecasting economic growth using artificial intelligence approaches," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8661–8680, 2019.

[13] S. Dorosti, S. J. Ghoushchi, E. Sobhrakhshankhah, M. Ahmadi, and A. Sharifi, "Application of gene expression programming and sensitivity analyses in analyzing effective parameters in gastric cancer tumor size and location," *Soft Computing*, vol. 8, pp. 1–22, 2019.

[14] F. Isensee, P. Kickingereder, D. Bonekamp et al., "Brain tumor segmentation using large receptive field deep convolutional neural networks," in *Bildverar- beitung für die Medizin 2017*, pp. 86–91, Springer-Vieweg, Berlin, Germany, 2017.

[15] S. Chen, C. Ding, and M. Liu, "Dual-force convolutional neural networks for accurate brain tumor segmentation," *Pattern Recognition*, vol. 88, pp. 90–100, 2019.

[16] Z. Liu, L. Chen, L. Tong et al., *Deep Learning Based Brain Tumor Segmentation: A Survey*, 2020, http://arxiv.org/abs/2007.09479.

[17] K. Hu, Z. Zhang, X. Niu et al., "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocom- puting*, vol. 309, pp. 179–191, 2018.

[18] K. Kamnitsas, C. Ledig, V. F. J. Newcombe et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[19] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Proceedings International MICCAI Brainlesion Workshop*, pp. 178–190, Springer, Quebec City, QC, Canada, 2017.

[20] S. Zheng, S. Jayasumana, B. Romera-Paredes et al., "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 1529–1537, 2015.

[21] M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, and P. Büchler, "The virtual skeleton database: an open access repository for biomedical research and collaboration," *Journal of Medical Internet Research*, vol. 15, no. 11, article e245, 2013.

[22] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," *Proceedings International MICCAI Brainlesion Workshop*, , pp. 311–320, Springer, Granada, Spain, 2018.

[23] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, http://arxiv.org/abs/1412.7062.

[24] S. Zheng, S. Jayasumana, B. Romera-Paredes et al., "Conditional random fields as recurrent neural networks," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1529–1537, Santiago, Chile, 2015.

[25] M. Mittal, L. M. Goyal, S. Kaur, I. Kaur, and D. Jude Hemanth, "Deep learning based enhanced tumor segmentation approach for MR brain images," *Applied Soft Computing*, vol. 78, pp. 346–354, 2019.

[26] Z. Zhou, Z. He, and Y. Jia, "AFPNet: a 3D fully convolutional neural network with atrous-convolution feature pyramid for brain tumor segmentation via MRI images," *Neurocomputing*, vol. 402, pp. 235–244, 2020.

[27] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation," *International MICCAI Brainlesion Workshop, BrainLes 2018: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 61–72, 2018.

[28] H. Kai, G. Qinghai, Z. Yuan et al., "Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field," *IEEE Access*, vol. 7, no. 8, pp. 92615–92629, 2019.

[29] C. Kang, X. Yu, S. H. Wang et al., "A heuristic neural network structure relying on fuzzy logic for images scoring," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 34–45, 2020.

[30] M. Pfeiffer and T. Pfeil, "Deep learning with spiking neurons: opportunities and challenges," *Frontiers in neuroscience*, vol. 12, p. 774, 2018, https://www.frontiersin.org/articles/10.3389/fnins.2018.00774/full.

[31] S. Alagarsamy, K. Kamatchi, V. Govindaraj, Y.-D. Zhang, and A. Thiyagarajan, "Multi-channeled MR brain image segmentation: a new automated approach combining BAT and clustering technique for better identification of heterogeneous tumors," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 4, pp. 1005–1035, 2019.

[32] M. I. Sharif, J. P. Li, M. A. Khan, and M. A. Saleem, "Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images," *Pattern Recognition Letters*, vol. 129, pp. 181–189, 2020.

[33] M. Sharif, J. Amin, M. Raza, M. Yasmin, and S. C. Satapathy, "An integrated design of particle swarm optimization (PSO) with fusion of features for detection of brain tumor," *Pattern Recognition Letters*, vol. 129, pp. 150–157, 2020.

Hindawi
BioMed Research International
Volume 2021, Article ID 6685723, 8 pages
https://doi.org/10.1155/2021/6685723

Research Article

Radiomics Analysis of MR Imaging with Gd-EOB-DTPA for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma: Investigation and Comparison of Different Hepatobiliary Phase Delay Times

Shuai Zhang,[1] Guizhi Xu,[2] Chongfeng Duan,[1] Xiaoming Zhou,[1] Xin Wang,[1] Haiyang Yu,[1] Lan Yu,[1] Zhiming Li,[1] Yuanxiang Gao,[1] Ruirui Zhao,[3] Linlin Jiao,[4] and Gang Wang[1]

[1]*Department of Radiology, The Affiliated Hospital of Qingdao University, Qingdao, China*
[2]*Department of Radiology, Zhucheng People Hospital, Zhucheng Shandong, China*
[3]*Operating Room, The Affiliated Hospital of Qingdao University, Qingdao, China*
[4]*Interventional department, The Affiliated Hospital of Qingdao University, Qingdao, China*

Correspondence should be addressed to Xiaoming Zhou; zhouxm@qduhospital.cn and Gang Wang; 313682216@qq.com

Received 23 November 2020; Accepted 23 December 2020; Published 8 January 2021

Academic Editor: Yong Xia

Purpose. To investigate whether the radiomics analysis of MR imaging in the hepatobiliary phase (HBP) can be used to predict microvascular invasion (MVI) in patients with hepatocellular carcinoma (HCC). *Method.* A total of 130 patients with HCC, including 80 MVI-positive patients and 50 MVI-negative patients, who underwent MR imaging with Gd-EOB-DTPA were enrolled. Least absolute shrinkage and selection operator (LASSO) regression was applied to select radiomics parameters derived from MR images obtained in the HBP 5 min, 10 min, and 15 min images. The selected features at each phase were adopted into support vector machine (SVM) classifiers to establish models. Multiple comparisons of the AUCs at each phase were performed by the Delong test. The decision curve analysis (DCA) was used to analyze the classification of MVI-positive and MVI-negative patients. *Results.* The most predictive features between MVI-positive and MVI-negative patients included 9, 8, and 14 radiomics parameters on HBP 5 min, 10 min, and 15 min images, respectively. A model incorporating the selected features produced an AUC of 0.685, 0.718, and 0.795 on HBP 5 min, 10 min, and 15 min images, respectively. The predictive model for HBP 5 min, 10 min and 15 min showed no significant difference by the Delong test. DCA indicated that the predictive model for HBP 15 min outperformed the models for HBP 5 min and 10 min. *Conclusions.* Radiomics parameters in the HBP can be used to predict MVI, with the HBP 15 min model having the best differential diagnosis ability.

1. Introduction

Hepatocellular carcinoma (HCC) is one of the most common malignant tumors in the liver [1]. Surgery is regarded as the first choice for eligible patients [2]. Microvascular invasion (MVI) is a vital predictor of HCC recurrence, especially in the early stage after surgical resection [3, 4]. Previous studies have identified MVI as a major risk factor for early recurrence within two years after hepatectomy and transplantation [5]. The application of preoperative imaging methods to predict MVI has important clinical significance. Therefore, it is necessary to predict MVI to identify tumor invasion and predict tumor recurrence after hepatectomy and transplantation.

Previous studies found that some imaging features, such as the tumor size, shape, capsule, margin, apparent diffusion coefficient (ADC) values, and enhancement pattern, may contribute to the diagnosis of MVI before surgery [6–8]. However, these qualitative findings can be affected by many factors, including the variability between observers and the lack of external validation, and there is still debate about the predictive value of MVI in HCC. Recently, radiomics

analysis has become an emerging quantitative image processing method. It can quantify tissue heterogeneity by evaluating the distribution of radiomics roughness and irregularity within lesions. Different from tissue biomarkers, which can assess the microheterogeneity of regional tumors, radioactive biomarkers can noninvasively examine the whole tumor at the millimeter level [9]. Therefore, this method is expected to quantitatively evaluate lesion characteristics in more detail and with better repeatability than visual analysis by human observers. Some published studies have evaluated the potential of radiomics in predicting MVI in hepatocellular carcinoma [4, 10–12]. To the best of our knowledge, no research on predicting MVI or comparing imaging at different hepatobiliary phase (HBP) times using radiomics analysis of gadolinium-ethoxybenzyl-diethylenetriaminepentaacetic acid- (Gd-EOB-DTPA-) enhanced MR has been reported.

Thus, the aim of this study was to investigate whether radiomics analysis of MR imaging with Gd-EOB-DTPA in HBP can be used to predict MVI in patients with HCC and compare the prediction of MVI on different HBP delay times.

2. Materials and Methods

2.1. Patients. This retrospective study was approved by our institutional review committee, and patient informed consent was waived. By searching our institution's database, 294 consecutive liver cancer patients were selected between January 2015 and May 2020. The inclusion criteria were as follows: (1) MR images showing liver tumors larger than 1 cm in diameter; (2) Gd-EOB-DTPA-enhanced MRI scan including complete examination recordings at HBP 5 min, 10 min, and 15 min; and (3) HCC diagnosed by postoperative pathology. The exclusion criteria were as follows: (1) patients who underwent MRI examination more than one month before surgery; (2) patients who had received liver cancer treatment before surgery; and (3) insufficient image quality for radiomics analysis. Finally, 130 HCC patients, including 80 MVI-positive patients and 50 MVI-negative patients, were included in this study. The MVI information was obtained from the HIS system at our hospital and was diagnosed by the same pathologist. According to the date of MRI, the cohort was divided into a training set ($n = 91$; 60 men and 31 women; mean age 57.8 ± 12.6 years) and a time-independent validation set ($n = 39$; 29 men and 10 women; average age 58.6 ± 11.6 years).

2.2. MR Techniques. All study patients underwent MR imaging using a 3.0T scanner (GEHCGEHC, GE medical systems, Waukesha, WI). A dose of 0.1 mL/kg (0.025 mmol/kg) Gd-EOB-DTPA (Primovist, Bayer HealthCare, Berlin, Germany)) was administered at a flow rate of 1.0 mL/s followed by 25 mL of saline. A 3D fat-suppressed Liver Acquisition with Volumetric Acceleration (LAVA, GE Healthcare) sequence was performed in the axial plane at 5, 10, and 15 min after contrast agent injection (HBP 5 min, 10 min, and 15 min, respectively). The imaging parameters of the LAVA sequence were as follows: TR/TE, 2.5/1.1; inversion time, 5.0 milliseconds; flip angle, 9°; thickness, 5 mm; slice spacing, 2.5 mm; FOV, 380–450 mm; 256 × 256 matrix; number of signals acquired, 0.70; and bandwidth, 976.6 kHz. The comparison of dynamic T1-weighted and T2-weighted imaging was not the focus of this study and was not conducted.

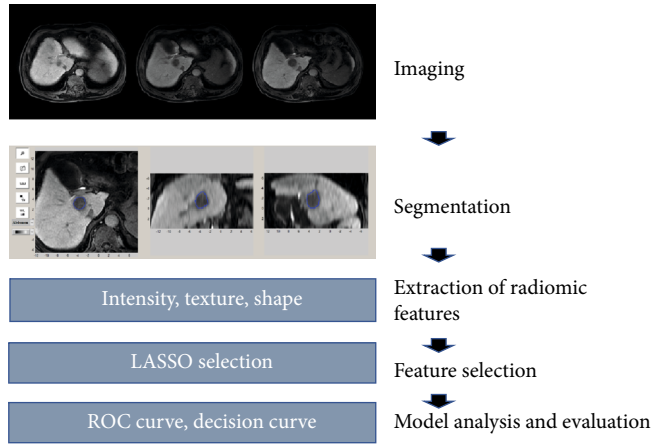


Figure 1: Workflow of radiomics analysis.

2.3. MR Radiomics Analysis. The workflow of the radiomics analysis included tumor segmentation, feature extraction, feature selection, and model construction and evaluation (Figure 1).

Three-dimensional segmentation of HCC using the IBEX software (http://bit.ly/IBEX) was performed by two radiologists in abdominal diagnostics with 8-year and 10-year MR experience who were blinded to the MVI information. When patients had multiple tumors, the largest tumor was analyzed. The regions of interest were drawn manually on HBP 5 min, 10 min, and 15 min images, covering the whole tumor. Radiomics parameters were selected using the IBEX software and included eight categories: Gradient Orient Histogram, Gray Level Cooccurrence Matrix 25, Gray Level Run Length Matrix 25, Intensity Direct, Intensity Histogram, Intensity Histogram Gauss Fit, Neighbor Intensity Difference 25, and Shape. Each category included different radiomics parameters. The intraclass correlation coefficient (ICC) of 30 randomly selected tumors was calculated to test the repeatability of features extracted by repeated segmentation, and features with an ICC less than 0.80 were excluded.

2.4. Statistical Analysis. Two independent sample t tests were used to compare the mean age between the MVI-positive and MVI-negative patients. The chi-square test was used to compare the sex distribution between the MVI-positive and MVI-negative patients. The least absolute shrinkage and selection operator (LASSO) regression method was used to select the most valuable parameter from all parameters obtained at HBP 5 min, 10 min, and 15 min. Receiver operating characteristic (ROC) curves and areas under the ROC curve (AUCs) of the radiomics parameters selected at each phase were calculated. The selected features at each phase were adopted into support vector machine (SVM) classifiers to establish models. SVM models were evaluated by 10-fold cross-validation to reduce overfitting. Multiple comparisons of the AUCs at each phase were performed by the Delong test

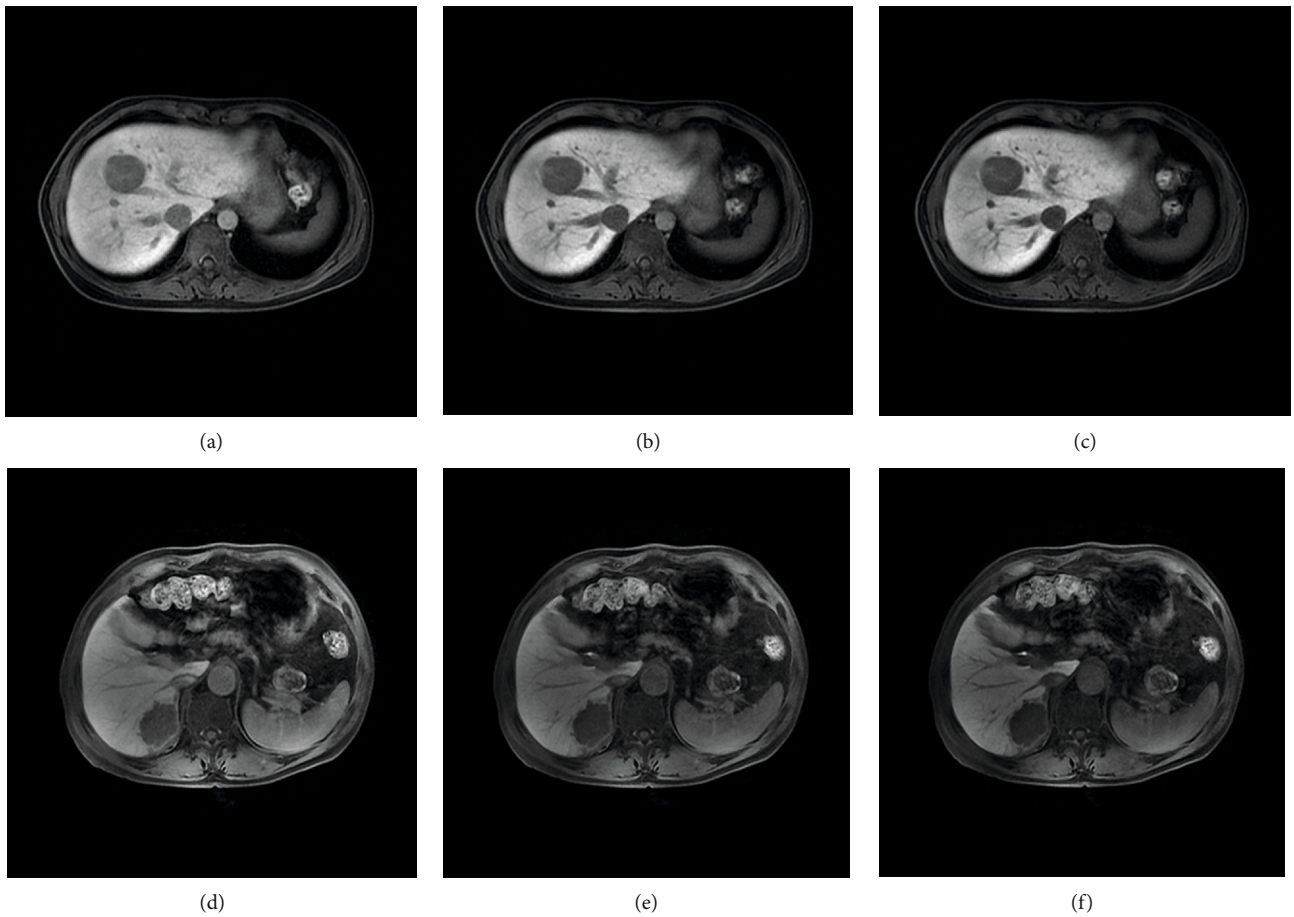


Figure 2: Axial MR imaging with Gd-EOB-DTPA on HBP in a HCC MVI-negative patient ((a) HBP 5 min, (b) HBP 10 min, and (c) HBP 15 min), and a MVI-positive patient ((d) HBP 5 min, (e) HBP 10 min, and (f) HBP 15 min). The imaging of MVI negative shows a smooth tumor margin, while MVI-positive shows a nonsmooth tumor margin. However, other tumor features between MVI positive and negative are difficult to identify by visual inspection.

with Bonferroni-adjusted p values. To present the distribution of the radiomics parameters in which HBP imaging can best differentiate MVI-positive and MVI-negative patients, a heat map was created. Decision curve analysis (DCA) was used to analyze the classification. The interobserver reproducibility of the selected valuable radiomics parameter was evaluated by ICC. SPSS 22.0 (Chicago, Illinois, USA) was used for statistical analysis. LASSO regression, ROC curves, the Delong test, and DCA were performed by using R (https://www.r-project.org/). $p < 0.05$ was considered statistically significant.

3. Results

3.1. Demographics. Eighty MVI-positive HCC patients and 50 MVI-negative patients were included. There was no significant difference in age or sex between MVI-positive and MVI-negative patients. Examples of HCCs in MVI-positive and MVI-negative patients are shown in Figure 2.

3.2. Comparison of MR Radiomics Analyses with LASSO Regression. Each ROI has 8 categories and 1768 radiomics parameters. The most predictive features between MVI-positive and MVI-negative patients included 9 radiomics parameters at HBP 5 min, 8 radiomics parameters at HBP 10 min, and 14 radiomics parameters at HBP 15 min (Table 1). The two radiomics parameters with the top two AUC values were X0.7 Homogeneity (AUC = 0.641) and Compactness2 (AUC = 0.615) in the hepatobiliary phase (HBP) at 5 min, X1.7 Contrast (AUC = 0.625) and X4.7 Auto Correlation (AUC = 0.605) in the hepatobiliary phase (HBP) at 10 min, and X6.1 Difference Entropy (AUC = 0.645) and X4.7 Dissimilarity (AUC = 0.638) in the hepatobiliary phase (HBP) at 15 min. A model incorporating all radiomics parameters selected by LASSO in each phase produced AUCs of 0.685, 0.718, and 0.795 at HBP 5 min, 10 min, and 15 min, respectively (Figure 3).

3.3. Comparison of the 3 HBP Delays in Differentiating MVI. The results of the Delong test used to differentiate MVI-positive and MVI-negative patients for the 3 HBP delays are shown in Table 2. The predictive model for HBP 5 min, 10 min, and 15 min showed no significant difference (HBP 5 min vs. HBP 10 min, $p = 0.751$; HBP 5 min vs. HBP 15 min, $p = 0.362$; HBP 10 min vs. HBP 15 min, $p = 0.440$). The radiomics parameter distribution at HBP 15 min is

Table 1: The most predictive features between MVI-positive and MVI-negative selected by LASSO regression.

Phase	Radiomics parameter	Which category belongs to	Regression coefficient
HBP 5 min	Compactness2	Shape	0.30551482
	Mass	Shape	0.03586056
	VoxelSize	Shape	0.08370105
	TextureStrength	NeighborIntensityDifference25	-0.4756972
	7.7Energy	GrayLevelCooccurenceMatrix3	-0.22185
	0.7Homogeneity	GrayLevelCooccurenceMatrix3	0.17812152
	4.1InformationMeasureCorr1	GrayLevelCooccurenceMatrix3	0.10846544
	1.4InverseDiffMomentNorm	GrayLevelCooccurenceMatrix3	0.0502487
	X10.4InverseDiffNorm	GrayLevelCooccurenceMatrix3	-0.10842471
HBP 10 min	.333ShortRunHighGrayLevelEmpha	GrayLevelRunLengthMatrix25	-0.025523039
	0ShortRunHighGrayLevelEmpha -	GrayLevelRunLengthMatrix25	-0.037606409
	NumberOfObjects	Shape	-0.150641355
	SurfaceArea	SurfaceArea	0.04154656
	1.7Contrast	GrayLevelCooccurenceMatrix3	-0.327361605
	.333.7Dissimilarity	7Dissimilarity	-0.159344876
	7.7Energy	GrayLevelCooccurenceMatrix3	-0.247759079
	6.7MaxProbability	GrayLevelCooccurenceMatrix3	-0.301320203
HBP 15 min	MedianAbsoluteDeviation	GradientOrientHistogram	0.012259179
	5Percentile	GradientOrientHistogram	-0.064760131
	Mass	Shape	0.052851135
	SphericalDisproportion	Shape	-0.067929609
	4.7AutoCorrelation	GrayLevelCooccurenceMatrix3	-0.182481531
	1.7Contrast	GrayLevelCooccurenceMatrix3	-0.049683698
	9.4Contrast	GrayLevelCooccurenceMatrix3	-0.300911891
	6.1DifferenceEntropy	GrayLevelCooccurenceMatrix3	-0.105303216
	4.7Dissimilarity	GrayLevelCooccurenceMatrix3	-0.205762541
	8.4InverseDiffNorm	GrayLevelCooccurenceMatrix3	-0.002903426
	1.1InverseVariance	GrayLevelCooccurenceMatrix3	0.046277632
	11.4InverseVariance	GrayLevelCooccurenceMatrix3	-0.001397914
	12.4InverseVariance	GrayLevelCooccurenceMatrix3	-0.188965726
	8.4MaxProbability	GrayLevelCooccurenceMatrix3	-0.251615445

HBP, hepatobiliary phases.

demonstrated with a heat map in Figure 4. The results of DCA at HBP 5 min, 10 min, and 15 min are shown in Figure 5. There was no net benefit of HBP 5 min when the threshold probability was less than approximately 0.5 and no net benefit of HBP 10 min within almost the same threshold probability range. HBP 15 min had a larger net benefit than HBP 5 min when the threshold probability was less than approximately 0.7, and there was a slightly lesser net benefit when the threshold probability was between approximately 0.7 and 0.8.

3.4. Interobserver Agreement for the Selected Valuable Radiomics Parameter at HBP 15 Min. The interobserver agreement between the 2 radiologists was good for the selected valuable radiomics parameter at HBP 15 min (ICC range: 0801–0.997) (Table 3).

4. Discussion

MVI is a vital independent predictor of early recurrence in HCC patients [13, 14]. Gd-EOB-DTPA is a biphasic T1-weighted MRI contrast agent which enters hepatocytes in an ATP-dependent manner through the organic anion transport polypeptide (OATP1B1/B3) and is finally excreted through the biliary tract. It is used for dynamically contrast-enhanced MRI of the liver, as well as the specific imaging process during the HBP after injection. A previous study indicated that radiomics signatures on HBP 20 min images could assess MVI in patients with HCC [15]. However, few studies have been conducted to assess MVI and compare HBP 5 min, 10 min, and 15 min images using radiomics from Gd-EOB-DTPA-enhanced MR. In the present study, after recruiting patients with HCC, we employed radiomics to

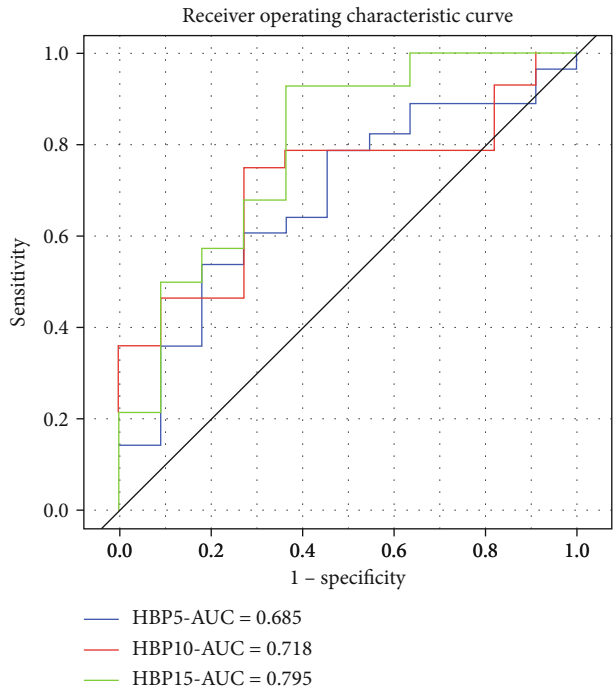


Figure 3: The ROC curves and AUC values of HBP 5 min, 10 min, and 15 min to differentiate MVI-positive and MVI-negative patients. The HBP 15 min produced the highest AUC of 0.795.

Table 2: The results of multiple comparisons of the AUCs by the Delong test.

	Z statistic	p
HBP 5 min-HBP 10 min	-0.3173	0.751
HBP 5 min-HBP 15 min	-0.9121	0.362
HBP10min-HBP 15 min	-0.7725	0.440

HBP, hepatobiliary phases.

assess MVI in HCC with Gd-EOB-DTPA on HBP 5 min, 10 min, and 15 min images. We verified the capability of the radiomics model for preoperative prediction of MVI status in a verification cohort.

Manifestations on the HBP images of Gd-EOB-DTPA-enhanced MRI indicate the functions of hepatocytes. HCC cells, relative to hepatocytes, fail to carry out the absorption of Gd-EOB-DTPA in the HBP. This can lead to low intensity within the tumor at this stage. However, previous studies [7, 16] reported that the occurrence of MVI cannot be predicted by assessing the difference in the occurrence of intratumoral hypointensity on HBP. In the present study, a model incorporating the radiomics parameters on HBP 5 min, 10 min, and 15 min images produced AUCs of 0.685, 0.718, and 0.795, indicating that the HBP model can assess MVI in HCC. This is because radiomics has the advantages of stable calculation, high repeatability, indefatigability, and being free from human subjective initiative interference [17, 18]. Tumor heterogeneity is likely to be difficult to identify and quantify by conventional imaging tools, the subjective assessment of images, or random sampling biopsy [19], whereas the mentioned techniques have been shown to be tightly associated with the pathophysiology of cancer. Existing studies have reported that the characteristics of radiomics show tight associations with the microstructure and biological behavior of tumors [20, 21]. In the present study, 14 quantitative characteristics on HBP 15 min images were found, which were not presented previously. Radiomics characteristics are important markers of intratumoral homogeneity. Of the 14 radiomics characteristics related to MVI in the present study, 2 were histogram-related characteristics (Median Absolute Deviation, 5th Percentile), 2 were shape-related characteristics (Mass, Spherical Disproportion), and others were matrix-related characteristics (4.7 Auto Correlation, 1.7 Contrast, 9.4 Contrast, 6.1 Difference Entropy, 4.7 Dissimilarity, 8.4 Inverse Diff Norm, 1.1 Inverse Variance, 11.4 Inverse Variance, 12.4 Inverse Variance, and 8.4 Max Probability). The features based on the histogram are first-order statistics, primarily determined by the statistics of intensity information (or brightness information) in and around the tumor. Subsequently, the overall distribution of intensity information in and around the tumor was explored. The signal intensity of MVI-positive HCC was lower than that of MVI-negative HCC, and differences in histogram characteristics were more frequent [22]. Shape-related characteristics were adopted to express the complexity of the lesion shape. Given histological studies, MVI-positive HCC exhibited an aggressive tendency, invading the tumor envelope and extending into the noncancerous substance, thereby causing a higher incidence of irregular tumor margins [23]. Matrix-based characteristics are second-order statistics applied to express lesions complex characteristics, the variation of hierarchical structure, and the thickness of texture. The difference in the mentioned parameters may indicate the heterogeneity of the tumor that is difficult to identify by the subjective assessment of images. Although radiomics has already been applied, it can effectively mark images, which can facilitate the assessment and quantification of processes of tumor space-related heterogeneity [24]. Nevertheless, the radiomics characteristics are acquired and determined with a PC. It is very challenging to explain the relationships between the radiomics characteristics, and pathology-related manifesting data are a challenge to develop [25]. First, the pathophysiological process involves several interacting parts; second, the maximum data acquired by the PC image study are significantly greater that acquired by visual examination.

The predictive models for HBP 5 min, 10 min, and 15 min had no significant differences according to the Delong test (HBP 5 min vs. HBP 10 min, $p = 0.751$; HBP 5 min vs. HBP 15 min, $p = 0.362$; and HBP 10 min vs. HBP 15 min, $p = 0.440$). To further compare the models for HBP 5 min, 10 min, and 15 min, this study applied DCA, i.e., a method to assess the models in terms of the clinical consequences and calculate the benefit and the loss of the assessed models for respective individuals [26]. This method attempts to overcome the limitations of traditional statistical indicators and complete decision analysis methods, which cannot directly provide clinical value information, nor can they be used in routine biostatistics practice [27]. The present study

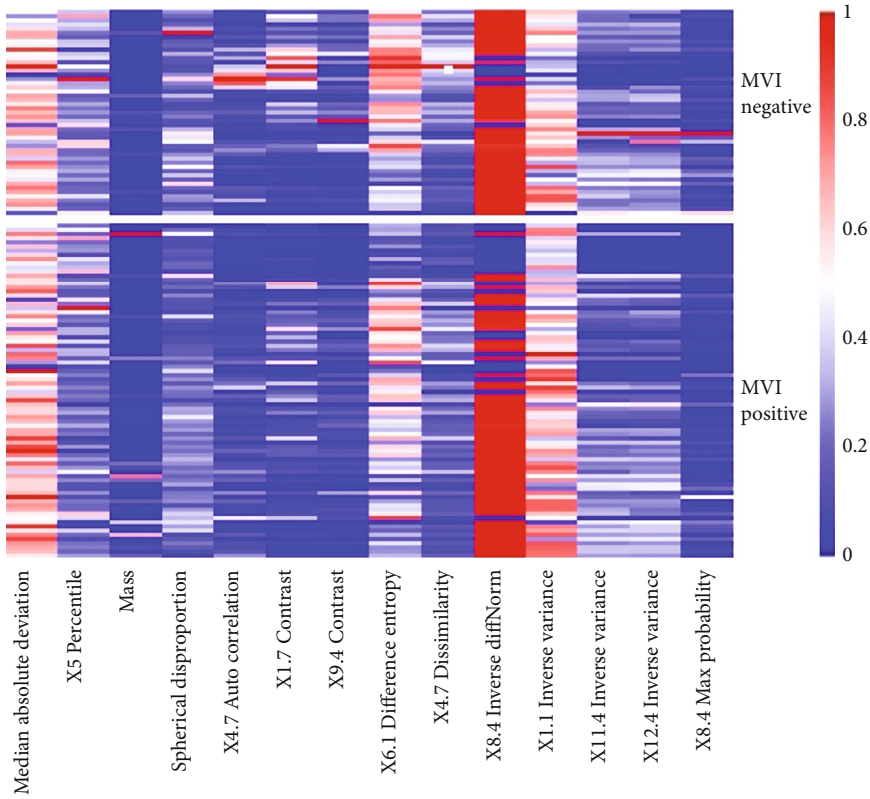


Figure 4: The heat map of HBP 15 min shows the distribution of the most predictive texture parameters between MVI-positive and MVI-negative patients. Difference in colors means different values of radiomics parameter.

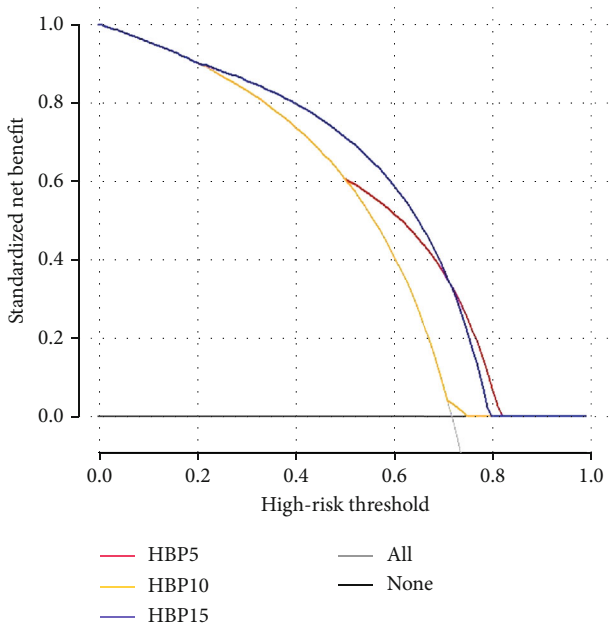


Figure 5: Decision curve analysis of HBP 5 min, 10 min, and 15 min. The X axis represents the threshold probability, and the Y axis represents the net benefit. HBP 5 min had no net benefit when the threshold probability is less than about 0.5. HBP 10 min had no net benefit almost across the range of the threshold probability. HBP 15 min had a larger net benefit than HBP 5 min when the threshold probability is less than about 0.7, and a little lesser net benefit when the threshold probability is about between 0.7 and 0.8.

Table 3: The interobserver reproducibility of the most predictive features on HBP 15 min.

Radiomics parameter	ICC
MedianAbsoluteDeviation	0.924
5Percentile	0.903
Mass	0.997
SphericalDisproportion	0.847
4.7AutoCorrelation	0.898
1.7Contrast	0.850
9.4Contrast	0.801
6.1DifferenceEntropy	0.948
4.7Dissimilarity	0.939
8.4InverseDiffNorm	0.832
1.1InverseVariance	0.975
11.4InverseVariance	0.812
12.4InverseVariance	0.838
8.4MaxProbability	0.929

ICC, intraclass correlation coefficient.

revealed that HBP 15 min images achieved the largest net benefit under the threshold probability, only with a slightly lesser net benefit when the threshold probability was between approximately 0.7 and 0.8. Gd-EOB-DTPA enters hepatocytes through organic anion transport polypeptides and is finally excreted through the biliary tract; this process takes

some time to complete. Therefore, we predicted that this, in theory, is why the HBP 15 min model outperformed the HBP 5 min and 10 min model. Wu et al. [28] found that the severity of liver cirrhosis had a significant negative effect on the detection of HCC by HBP. For patients with severe cirrhosis, HBP 15 min or longer seems to be more suitable for HCC than BHP 5 min and 10 min. Nakamura [29] reported that more focal liver lesions could be assessed on HBP 15 min images compared with HBP 5 and 10 min images. HCC patients often have a background of cirrhosis, leading to varying degrees of damage to liver function. The present study showed that the predictive model for HBP 15 min outperformed the HBP 5 min and 10 min models, which is in line with the proposed theory and previous research. Feng et al. [4] reported that the AUC of the HBP 20 min model for predicting MVI in the training and validation cohorts was 0.85 and 0.83, respectively; the diagnostic efficiency of this model was slightly higher than that of our study. However, their model combined intratumoral and peritumoral radiomics information. Liang et al. [30] reported that HBP 15 min was sufficient for lesion characterization in cirrhosis patients with mild liver dysfunction when compared with HBP 20 min. Though the present study did not include data for HBP 20 min, we predicted that the model for HBP 15 min was sufficient for MVI prediction in HCC. Additionally, other features and biomarkers could be incorporated in HBP 15 min to improve diagnostic efficiency.

Several limitations are revealed in this study. First, this study was a retrospective study, which may have caused inevitable selection bias, and lacks external validation. Second, compared with the relatively large number of variables, the sample size remained limited. Third, our verification cohort and training cohort were from the same center, and the radiology analysis conducted for the stability assessment will be further optimized in future multicenter studies. Fourth, in the present study, only MR images of HBP at 5 min, 10 min, and 15 min were explored. There are no data for HBP at 20 min on account of daily busy clinical work pressure. A multicenter and prospective study with a longer delay time and a larger population is needed to validate these results in the future. Ideally, the characterization of MVI should involve both intratumoral and peritumoral areas; therefore, it was another limitation for only analyzing intratumoral area in this study.

5. Conclusion

In conclusion, radiomics parameters on the HBP 5 min, 10 min, and 15 min images after Gd-EOB-DTPA injection can be used to predict MVI for HCC, with the HBP 15 min model having the best differential diagnosis ability; this model has potential clinical value for preoperative noninvasive prediction of MVI in HCC patients.

Data Availability

All data used during the study are available in the article and can be solicited from the corresponding author.

Conflicts of Interest

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

Authors' Contributions

Shuai Zhang and Guizhi Xu contributed equally to this work.

Acknowledgments

This research was supported by the General Project of Start-up Fund 2018, Affiliated Hospital of Qingdao University (3251).

References

[1] J. D. Yang and L. R. Roberts, "Epidemiology and management of hepatocellular carcinoma," *Infectious Disease Clinics of North America*, vol. 24, no. 4, pp. 899–919, 2010.

[2] K. Hao, J. M. Luk, N. P. Lee et al., "Predicting prognosis in hepatocellular carcinoma after curative surgery with common clinicopathologic parameters," *BMC Cancer*, vol. 9, no. 1, 2009.

[3] S. Lee, K. W. Kim, W. K. Jeong et al., "Gadoxetic acid-enhanced MRI as a predictor of recurrence of HCC after liver transplantation," *European Radiology*, vol. 30, no. 2, pp. 987–995, 2020.

[4] S. T. Feng, Y. Jia, B. Liao et al., "Preoperative prediction of microvascular invasion in hepatocellular cancer: a radiomics model using Gd-EOB-DTPA-enhanced MRI," *European Radiology*, vol. 29, no. 9, pp. 4648–4659, 2019.

[5] M. Renzulli, F. Buonfiglioli, F. Conti et al., "Imaging features of microvascular invasion in hepatocellular carcinoma developed after direct-acting antiviral therapy in HCV-related cirrhosis," *European Radiology*, vol. 28, no. 2, pp. 506–513, 2018.

[6] C. An, D. W. Kim, Y. N. Park, Y. E. Chung, H. Rhee, and M. J. Kim, "Single hepatocellular carcinoma: preoperative MR imaging to predict early recurrence after curative resection," *Radiology*, vol. 276, no. 2, pp. 433–443, 2015.

[7] S. Lee, S. H. Kim, J. E. Lee, D. H. Sinn, and C. K. Park, "Preoperative gadoxetic acid-enhanced MRI for predicting microvascular invasion in patients with single hepatocellular carcinoma," *Journal of Hepatology*, vol. 67, no. 3, pp. 526–534, 2017.

[8] P. Xu, M. Zeng, K. Liu, Y. Shan, C. Xu, and J. Lin, "Microvascular invasion in small hepatocellular carcinoma: is it predictable with preoperative diffusion-weighted imaging?," *Journal of Gastroenterology and Hepatology*, vol. 29, no. 2, pp. 330–336, 2014.

[9] S. R. Bowen, W. T. C. Yuh, D. S. Hippe et al., "Tumor radiomic heterogeneity: multiparametric functional imaging to characterize variability and predict response following cervical cancer radiation therapy," *Journal of Magnetic Resonance Imaging*, vol. 47, no. 5, pp. 1388–1396, 2018.

[10] J. Peng, J. Zhang, Q. Zhang, Y. Xu, J. Zhou, and L. Liu, "A radiomics nomogram for preoperative prediction of microvascular invasion risk in hepatitis B virus-related hepatocellular carcinoma," *Diagnostic and Interventional Radiology*, vol. 24, no. 3, pp. 121–127, 2018.

[11] X. Xu, H. L. Zhang, Q. P. Liu et al., "Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and

outcome in hepatocellular carcinoma," *Journal of Hepatology*, vol. 70, no. 6, pp. 1133–1144, 2019.

[12] S. Bakr, S. Echegaray, R. Shah et al., "Noninvasive radiomics signature based on quantitative analysis of computed tomography images as a surrogate for microvascular invasion in hepatocellular carcinoma: a pilot study," *Journal of Medical Imaging*, vol. 4, no. 4, article 041303, 2017.

[13] K.-C. Lim, P. K.-H. Chow, J. C. Allen et al., "Microvascular invasion is a better predictor of tumor recurrence and overall survival following surgical resection for hepatocellular carcinoma compared to the Milan criteria," *Annals of Surgery*, vol. 254, no. 1, pp. 108–113, 2011.

[14] V. Mazzaferro, J. M. Llovet, R. Miceli et al., "Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis," *The Lancet Oncology*, vol. 10, no. 1, pp. 35–43, 2009.

[15] L. Yang, D. Gu, J. Wei et al., "A radiomics nomogram for preoperative prediction of microvascular invasion in hepatocellular carcinoma," *Liver Cancer*, vol. 8, no. 5, pp. 373–386, 2019.

[16] L. Zhang, X. Yu, W. X. Wei et al., "Prediction of HCC microvascular invasion with gadobenate-enhanced MRI: correlation with pathology," *European Radiology*, vol. 30, no. 10, pp. 5327–5336, 2020.

[17] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[18] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar et al., "Correction: Corrigendum: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, no. 1, 2014.

[19] F. Davnall, C. S. Yip, G. Ljungqvist et al., "Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice?," *Insights Into Imaging*, vol. 3, no. 6, pp. 573–589, 2012.

[20] P. Grossmann, O. Stringfield, N. el-Hachem et al., "Defining the biological basis of radiomic phenotypes in lung cancer," *eLife*, vol. 6, article e23421, 2017.

[21] K. Hayano, F. Tian, A. R. Kambadakone et al., "Texture analysis of non-contrast-enhanced computed tomography for assessing angiogenesis and survival of soft tissue sarcoma," *Journal of Computer Assisted Tomography*, vol. 39, no. 4, pp. 607–612, 2015.

[22] J. Y. Kim, M. J. Kim, K. A. Kim, H. T. Jeong, and Y. N. Park, "Hyperintense HCC on hepatobiliary phase images of gadoxetic acid-enhanced MRI: correlation with clinical and pathological features," *European Journal of Radiology*, vol. 81, no. 12, pp. 3877–3882, 2012.

[23] H. Hu, Q. Zheng, Y. Huang et al., "A non-smooth tumor margin on preoperative imaging assesses microvascular invasion of hepatocellular carcinoma: a systematic review and meta-analysis," *Scientific Reports*, vol. 7, no. 1, p. 15375, 2017.

[24] M. G. Lubner, A. D. Smith, K. Sandrasegaran, D. V. Sahani, and P. J. Pickhardt, "CT texture analysis: definitions, applications, biologic correlates, and challenges," *Radiographics*, vol. 37, no. 5, pp. 1483–1503, 2017.

[25] X. Ma, J. Wei, D. Gu et al., "Preoperative radiomics nomogram for microvascular invasion prediction in hepatocellular carcinoma using contrast-enhanced CT," *European Radiology*, vol. 29, no. 7, pp. 3595–3605, 2019.

[26] V. Rousson and T. Zumbrunn, "Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies," *BMC Medical Informatics and Decision Making*, vol. 11, no. 1, p. 45, 2011.

[27] A. J. Vickers, B. van Calster, and E. W. Steyerberg, "A simple, step-by-step guide to interpreting decision curve analysis," *Diagnostic and Prognostic Research*, vol. 3, no. 1, p. 18, 2019.

[28] J. W. Wu, Y. C. Yu, X. L. Qu, Y. Zhang, and H. Gao, "Optimization of hepatobiliary phase delay time of Gd-EOB-DTPA-enhanced magnetic resonance imaging for identification of hepatocellular carcinoma in patients with cirrhosis of different degrees of severity," *World Journal of Gastroenterology*, vol. 24, no. 3, pp. 415–423, 2018.

[29] S. Nakamura, K. Awai, D. Utsunomiya et al., "Chronological evaluation of liver enhancement in patients with chronic liver disease at Gd-EOB-DTPA-enhanced 3-T MR imaging: does liver function correlate with enhancement?," *Japanese Journal of Radiology*, vol. 30, no. 1, pp. 25–33, 2012.

[30] M. Liang, J. Zhao, B. Xie et al., "MR liver imaging with Gd-EOB-DTPA: the need for different delay times of the hepatobiliary phase in patients with different liver function," *European Journal of Radiology*, vol. 85, no. 3, pp. 546–552, 2016.

Hindawi
BioMed Research International
Volume 2021, Article ID 6644071, 12 pages
https://doi.org/10.1155/2021/6644071

Research Article

FFU-Net: Feature Fusion U-Net for Lesion Segmentation of Diabetic Retinopathy

Yifei Xu,[1,2] **Zhuming Zhou**,[1,3] **Xiao Li**,[1] **Nuo Zhang**,[1] **Meizi Zhang**,[1] **and Pingping Wei**[4]

[1]*School of Software, Xi'an Jiaotong University, 710054, Xi'an, Shaanxi, China*
[2]*Huiyichen Inc. 1703, Block 1, No 1388, Jiulonghu Ave, 330038 Nanchang, Jiangxi, China*
[3]*Baidu Inc. Baidu Building, 10 Shangdi 10th Street, Haidian District, 100080 Beijing, China*
[4]*State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, 710054, Xi'an, China*

Correspondence should be addressed to Yifei Xu; belonxu_1@xjtu.edu.cn

Received 3 October 2020; Revised 25 November 2020; Accepted 21 December 2020; Published 4 January 2021

Academic Editor: Changming Sun

Diabetic retinopathy is one of the main causes of blindness in human eyes, and lesion segmentation is an important basic work for the diagnosis of diabetic retinopathy. Due to the small lesion areas scattered in fundus images, it is laborious to segment the lesion of diabetic retinopathy effectively with the existing U-Net model. In this paper, we proposed a new lesion segmentation model named FFU-Net (Feature Fusion U-Net) that enhances U-Net from the following points. Firstly, the pooling layer in the network is replaced with a convolutional layer to reduce spatial loss of the fundus image. Then, we integrate multiscale feature fusion (MSFF) block into the encoders which helps the network to learn multiscale features efficiently and enrich the information carried with skip connection and lower-resolution decoder by fusing contextual channel attention (CCA) models. Finally, in order to solve the problems of data imbalance and misclassification, we present a Balanced Focal Loss function. In the experiments on benchmark dataset IDRID, we make an ablation study to verify the effectiveness of each component and compare FFU-Net against several state-of-the-art models. In comparison with baseline U-Net, FFU-Net improves the segmentation performance by 11.97%, 10.68%, and 5.79% on metrics SEN, IOU, and DICE, respectively. The quantitative and qualitative results demonstrate the superiority of our FFU-Net in the task of lesion segmentation of diabetic retinopathy.

1. Introduction

Diabetic retinopathy is one of the main causes of blindness in human eyes, and regular fundus screening is an effective way to discover the location of disease [1–6]. At present, fundus screening is mainly diagnosed by analyzing fundus images manually, which requires ophthalmologists to have expert clinical experience. Therefore, the automatic screening and diagnosis of diabetic retinopathy have important practical significance. Moreover, the lesion segmentation of diabetic retinopathy is the prerequisite work for screening and diagnosing diabetic retinopathy, and it also lays a foundation for the subsequent grading of the severity of diabetic retinopathy. Generally, common diabetic retinopathy consists of microaneurysms (MA), hard exudates (EX), soft exudates (SE), and hemorrhage (HE).

In the past few decades, numerous researchers have devoted themselves to solving the segmentation of diabetic retinopathy. In early years, the researchers focused on traditional image processing methods, such as morphological operations and threshold segmentation [7–9]. Limited by the heavy dependence of the design level, the traditional methods of lesion segmentation are relatively infeasible in real-world application.

With the rapid development of deep learning technology, many researchers resort to deep learning methods to segment the lesion of diabetic retinopathy [3]. Although deep learning models can avoid handcrafted complex image features, it is difficult to segment tiny lesions composed of relatively macrostructures, such as microaneurysms and hemorrhage. As a classical medical semantic segmentation network, the symmetry-driven U-Net model [10] is weak in processing

tiny lesions. In order to achieve more accurate results, we propose a deep neural model called FFU-Net with an encoder-decoder structure. In detail, the pooling layer of U-Net is substituted with a convolutional layer to reduce the spatial loss of the fundus image. For the purpose of extracting multiscale lesion features, the MSFF block is embedded in the encoder by considering splitting operations and residual modules into account. For the decoders, contextual channel attention modules is integrated with the concatenation of skip connection and lower-resolution decoder. To alleviate the imbalance problem between lesion area and normal area in a fundus image, an improved Focal Loss named Balanced Focal Loss is proposed to train our model. In comparison with the state of the art, the experimental results on the public IDRID demonstrate that our model surpasses other models on metrics SEN, IOU, and DICE.

Our contributions are summarized as follows: (1) We replace the pooling layer of U-Net with a convolutional layer for downsampling, which helps to preserve spatial loss of fundus images as much as possible. (2) In the encoders, we integrate MSFF block with U-Net to extract multiscale lesion features by taking splitting operation and residual module into account, which is beneficial to representing informative features. (3) In the decoders, we propose the CCA module to fuse the information between skip connection and lower-resolution decoder, which share attentions and enhance their representative ability efficiently. (4) We propose a new loss to address the imbalance data problem when training our model, which facilitates the discrimination ability of our model. (5) We conduct several evaluations of the comparative methods on the benchmark dataset to figure out the superiority of our model.

The rest of this paper is organized as follows. Materials and Methods displays the related work, methodology, and experiment settings. The experimental results and the discussion are presented in Results and Discussion. Finally, Conclusion and Future Work concludes our work and suggests possible topics for future research.

2. Materials and Methods

2.1. Related Work. In the early years, the medical researchers focused on the segmentation of diabetic retinopathy based on traditional digital image processing methods, such as morphological operations and threshold segmentation. For example, Fleming et al. [7] used morphological operations and Gaussian matched filters to extract candidate regions of microaneurysms and then collected various statistical features to eliminate false positive points in blood vessels, yielding accurate segmentation of microaneurysms. Antal and Hajdu [11] adopted an ensemble learning strategy to integrate a series of image preprocessing approaches to improve final segmentation of microaneurysms. Kavitha and Duraiswamy [8] extracted exudate features using a multilayer threshold method, but this model has requirements for the input image quality. In conclusion, the traditional methods of lesion segmentation are relatively inefficient with poor generalization.

Recently, the development of deep learning has been widely concerned in the field of medical treatment. Medical image segmentation [12] has also become a hot topic. Most existing models with excellent performance in medical image segmentation tasks are reconstructed based on FCN or U-Net. In FCN [13, 14], the last full connection layer was replaced with a convolution layer. Rather than a fixed input size required by the classical CNN model, it allowed input image with arbitrary size. Also, skip connections were employed to combine local information learned from shallow layers and complex information learned from deeper layers. In U-Net, a contracting path was used for capturing context and a symmetric expanding path is designed for precise localization. With reference to the upsampling strategy, FCN applied upsampling operation to the last feature map while U-Net transformed high-level features to low-level features by deconvolution operations. References [15, 16] advanced in U-Net by using max-pooling indices and multipath input, respectively. Van Grinsven et al. [17] sped up the training by dramatically selecting misclassified negative samples. Sambyal et al. [18] presented a modified U-Net architecture based on the residual network and employ periodic shuffling with subpixel convolution initialized to convolution nearest-neighbor resize.

2.2. Methodology

2.2.1. Network Description. The overall pipeline of our proposed model is depicted in Figure 1. U-net was originally designed and developed for biomedical image segmentation. Its architecture is broadly regarded as an encoder network followed by a decoder network. For the encoder network, it is usually a pretrained classification network in which a downsampling pooling layer is appended at multiple different levels. For the decoder network, it includes upsampling and concatenation followed by regular convolution operations. The discriminative feature obtained by the encoder is projected onto pixel space to predict pixel-wise classification. As an extension of U-Net, our model makes the following three improvements adapted for lesion segmentation of diabetic retinopathy. (1) In the encoder stage, the maximum pooling layer of the original U-Net model for downsampling is substituted with a convolutional layer, in which the kernel size is 3×3 and stride = 2. The motivation behind this strategy could be explained as two points. (a) Compared with the pooling layer, downsampling with the convolution layer could keep structure information of diabetic retina images as much as possible. (b) It promotes the fusion of information between different channels, which is beneficial to the lesion segmentation task of diabetic retinopathy. Moreover, inspired by Inception block [19] and channel splitting idea [20], we design a new multiscale feature-fused block named MSFF to capture the features of the diabetic retinopathy image at different scales. As illustrated in Figure 2(a), the MSFF uses a series of multiscale residual splitting operations to extract different scale features. Firstly, as dilated convolution [21] could increase the receptive field under the condition that the resolution of the feature map is unchanged, we use a 3×3 dilated convolution followed by the RReLu layer

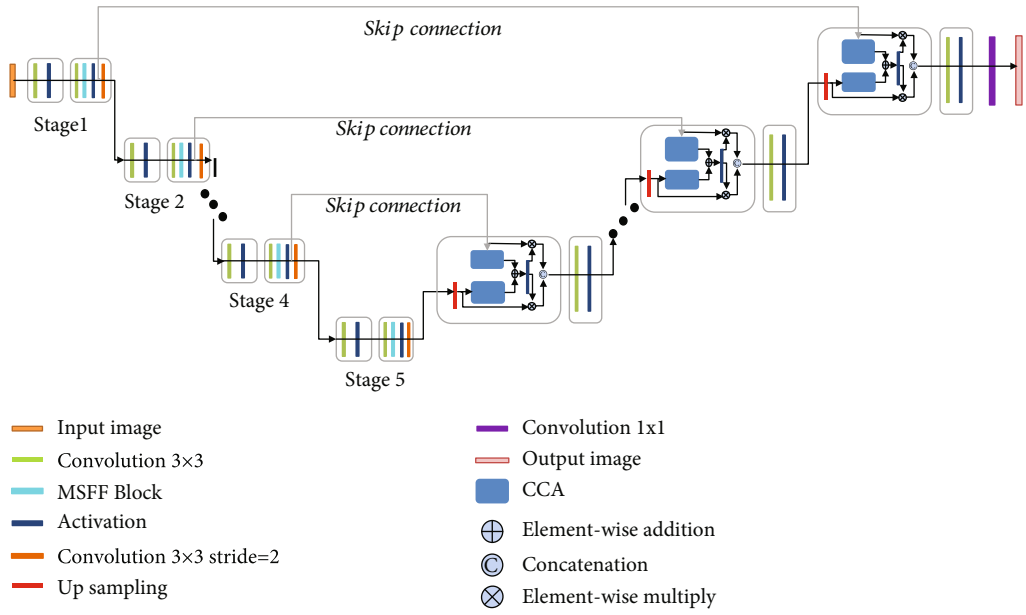


Figure 1: The overall architecture of the proposed FFU-Net model.

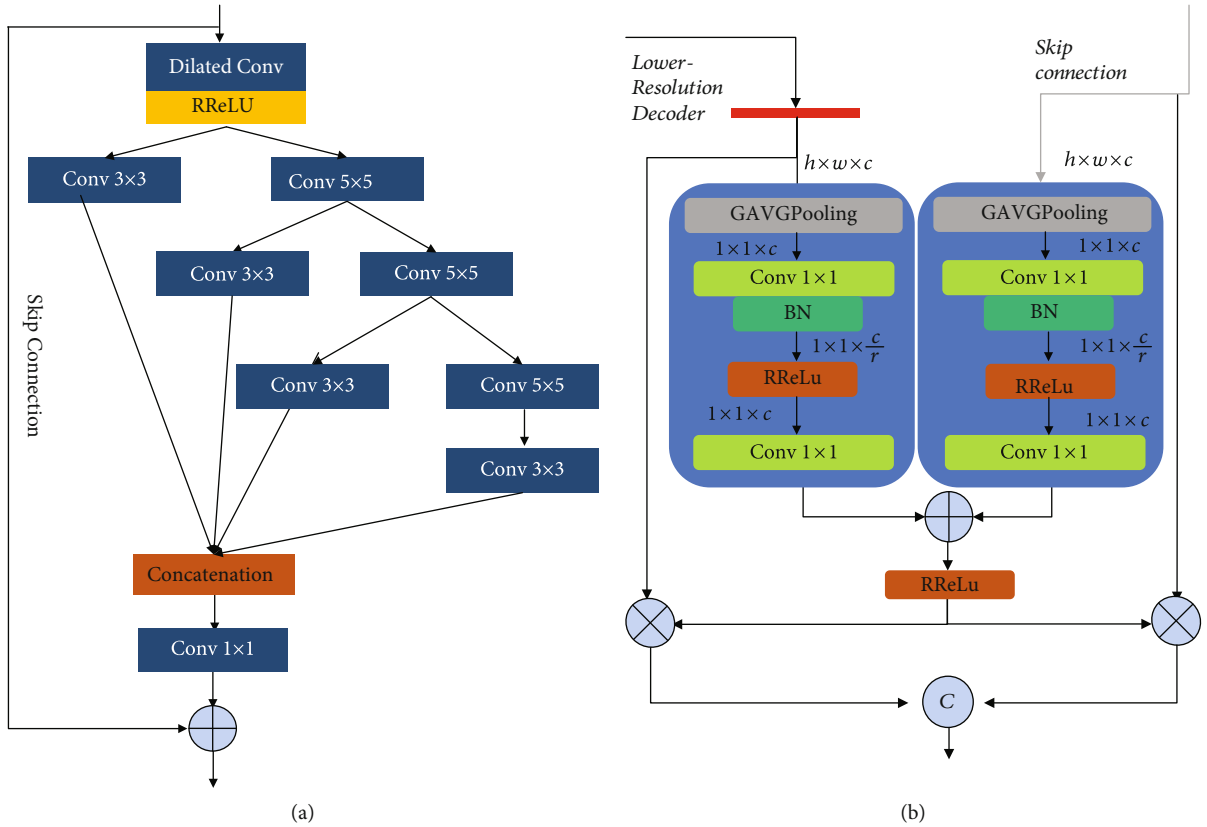


Figure 2: The improvements in the encoders and decoders of FFU-Net: (a) the structure of MSFF; (b) the CCA module in the decoders.

to perceive more information. Then, we put forward a series of splitting steps to produce multiscale features efficiently. For each step, MSFF employs 3×3 and 5×5 convolution layers to split the preceding features into two parts. One part is retained, and the other part is fed into the next step. After three splitting steps, all the distilled features are concatenated together and then fed into a 1×1 convolution to reduce the channels and parameters. In our implementation, only 1/3 channels in each splitting step are kept. (2) In the decoder stage, the concatenation procedure between skip connection

and lower-resolution decoder is improved with the contextual channel attention (CCA) module. We borrow the idea from SeNet [22] and depict the detail in Figure 2(b). Given lower-resolution decoder LD and skip connection SK with the size $h \times w \times c$, the proposed concatenation procedure with CCA can be described as

$$\begin{aligned} C_{\mathrm{LD}} &= \mathrm{Conv1}(\mathrm{RReLu}(\mathrm{Conv1\_BN}(\mathrm{GAPool}(\mathrm{UP}(\mathrm{LD}))))), \\ C_{\mathrm{SK}} &= \mathrm{Conv1}(\mathrm{RReLu}(\mathrm{Conv1BN}(\mathrm{GAPool}(\mathrm{SK})))), \\ \mathrm{CCA} &= \mathrm{RReLu}(C_{\mathrm{LD}} \oplus C_{\mathrm{SK}}), \\ F &= \mathrm{Concat}((\mathrm{CCA} \otimes \mathrm{LD}), \mathrm{CCA} \otimes \mathrm{SK}), \end{aligned} \tag{1}$$

where UP and GAPool denote upsampling operation and global average pooling. Conv1_BN is the 1×1 convolution followed by batch normalization while Conv1 is the common 1×1 convolution. RReLu and Concat represent the RReLu activation function and concatenation operation along the channel dimension. After the GAPooling operation, $1 \times 1 \times (c/r)$ $(r = 2)$ is employed to extract channel-wise statistics efficiently. As a contextual channel attention, CCA carries the channel-wise attentions from both LR and SK and then, respectively, multiply itself by LR and SK. Later, these two features are concatenated to replace the original concatenation procedure appearance in U-Net. In this way, LR and SK fully fuse the context information and share channel attention to provide more informative representation, which is conducive to the segmentation accuracy.

Besides, all the activation layers are replaced with nonlinear activation RReLu layers [23]. The reason why we prefer RReLu than other activation functions is that it could provide a random value from a uniform distribution to reduce overfitting during training. Herein, benefiting from the abovementioned improvements, our FFU-Net achieves segmentation accuracy of the four lesions of diabetic retinopathy effectively.

2.2.2. Loss Function. Apart from the network architecture, loss function also plays a key part in network design. In a diabetic retinopathy image, huge contrast could be found between the lesion and the normal from the perspective of appearance. Additionally, the size of the lesion area is always much smaller than the rest. Provided that we still insist on training our model to minimize the classification cross-entropy loss, the performance might not be like what it is supposed to be. This phenomenon can be ascribed to the imbalance problem occurring in the medical dataset. To address this issue, one can resort to data augmentation technology which duplicates samples to make the overall training set balanced. However, on account of the lack of diversity, the new dataset cannot provide clear improvement for our model. Alternatively, we turn to loss function according to the intrinsic distribution of data samples. Generally, the error penalties for the majority class and the minority class are different. Thus, we attempt to assign different weights to different classes and construct a Balanced Focal Loss for our model [24]. When training with this loss function, our model highlights the lesions of diabetic retinopathy. Different from original focal loss, in the task of medical segmentation in our application, the difference between easy and hard examples is more imperceptible. Mathematically, the loss function is formulated as follows:

$$\mathscr{L} = \sum_{i=1}^{n} -w\left(\left|y - \mathscr{Q}_i^y\right|\right)((1-y)\log(1-\mathscr{Q}_i) + y\log\mathscr{Q}_i), \tag{2}$$

where n represents the number of pixels in a diabetic retinopathy image and i denotes the ith sample. Here, $|\cdot|$ guarantees the nonnegativity. If the pixel is normal, its corresponding value is set to 0. If the pixel belongs to the lesion area, its corresponding value is set to 1. The parameter w represents the weight coefficient, which refers to the ratio between the pixels labeled as abnormal and the number of pixels in all samples. Q_i is the probability predicted by our proposed model; γ is the tunable focusing parameter which is always set to 2 in practice. As a comparison, we depict the values of Balanced Focal Loss and Focal Loss in Figure 3. As can be seen, when $Q_i \to 1$ and $y = 1$, the loss for well-classifier examples is downweighted. For instance, when $y = 1$, an example with $Q_i = 0.9$ and $w = 0.1$ would be 5x lower (0.002) than cross-entropy (0.010). Although the case with Focal Loss shows 100x lower (0.0001), the gap between Balanced Focal Loss and Focal Loss is 0.0019. Besides, another example with $Q_i = 0.1$ and $w = 0.1$ generates 0.227, which is closer to the result of cross-entropy (0.230). By this means, this proposed Balanced Focal Loss increases the importance of correcting misclassified examples.

2.3. Data Preparation and Processing

2.3.1. Data Preparation. The dataset we adopted is the Indian Diabetic Retinopathy Image Dataset (IDRID) [25], which is derived from a patient's fundus image during a real clinical examination at an ophthalmology clinic in India. All images in the dataset were taken by a Kowa VX-10α color fundus camera with a 50-degree field of view close to the macular area. All images have a resolution of 4288×2848 in JPG format. In our experiment, we select 81 color fundus images from 516 images along with pixel-level annotations. As illustrated in Figure 4(a), four typical diabetic retinopathy abnormalities appear in this dataset. The IDRID is split into the training set and testing set according to different lesion labels. Empirically, the distribution results are displayed in Table 1.

2.3.2. Fundus Image Preprocessing. A fundus image is taken with a color fundus camera. In most cases, influenced by uneven light intensity and camera lens contamination, the resultant fundus images are corrupted by uneven brightness, resulting in blurry and noisy areas. If the corrupted images are trained by the deep neural model directly, the noises will have adverse impact on the subsequent lesion segmentation of diabetic retinopathy.

To address the above problems, we take measures before feeding the fundus images into our network, such as image cropping, image denoising, image enhancement [26], image

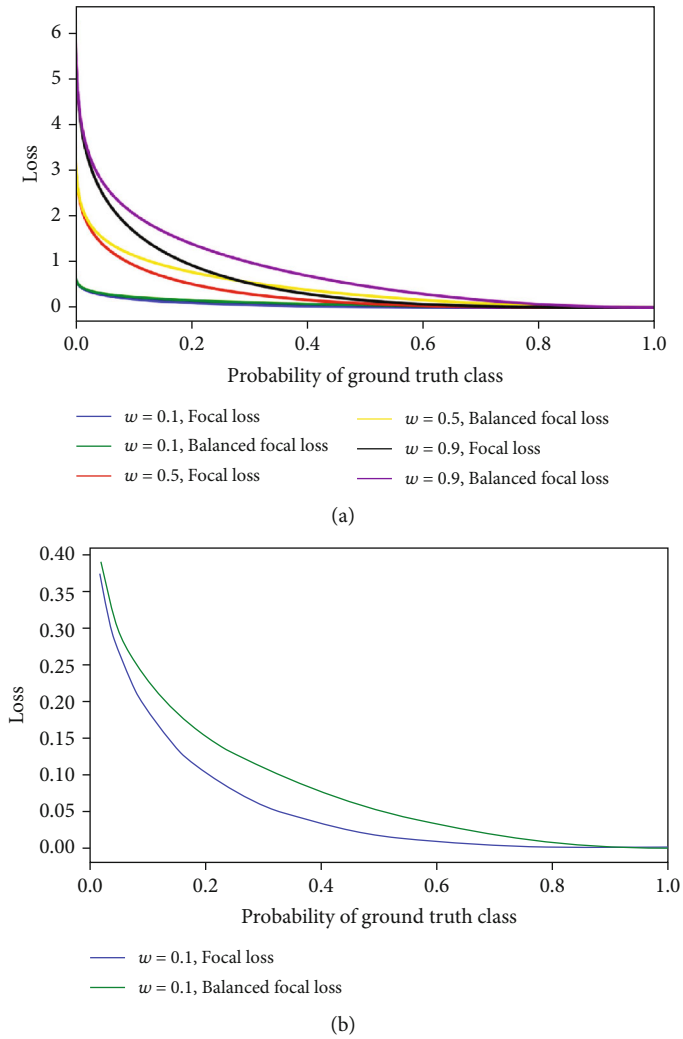
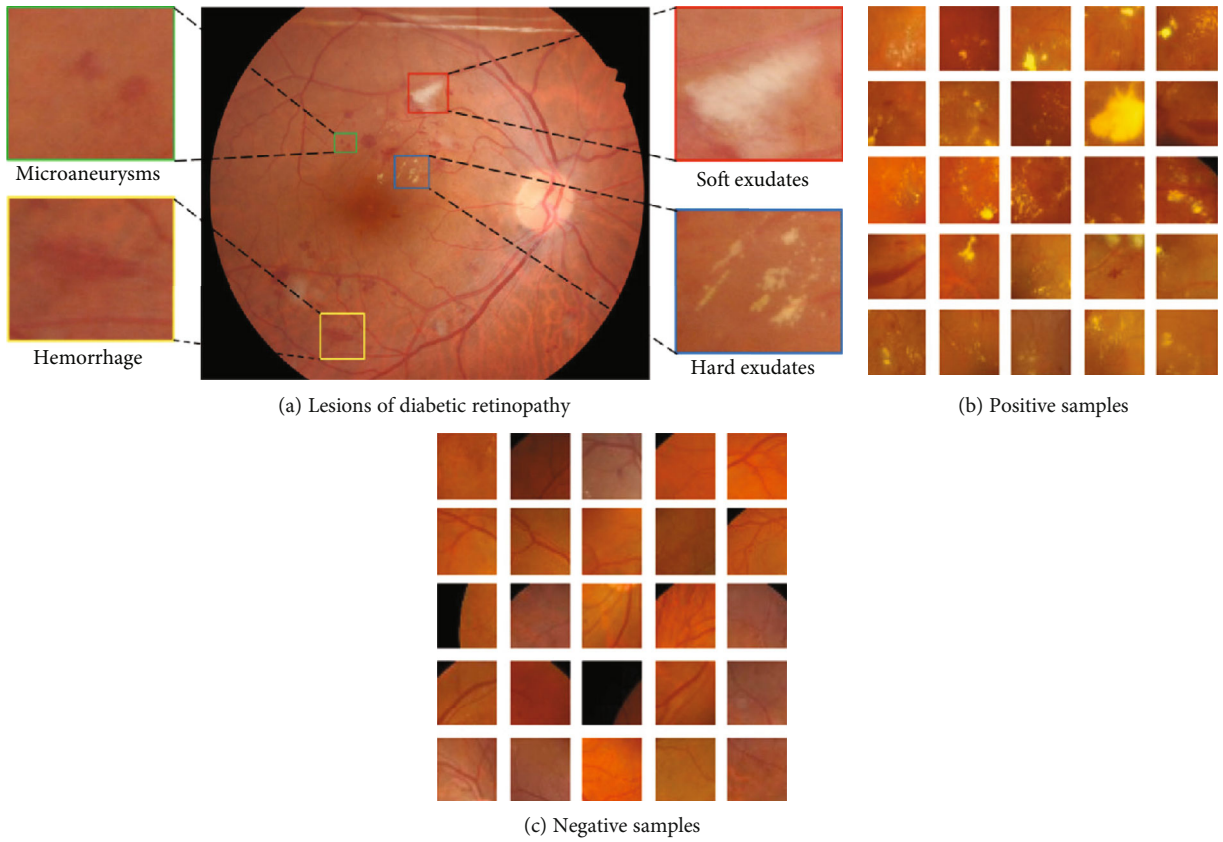


FIGURE 3: The comparative curves of Balanced Focal Loss and Focal Loss. (a) The results of Balanced Focal Loss and Focal Loss with different weights. (b) The zoom results of Balanced Focal Loss and Focal Loss when $w = 1$.

normalization, data augmentation, and image dicing. Here, we will illustrate the detail of preprocessing procedures. (1) Image cropping: the original samples are usually enclosed with a black border. To get the Region of Interest (ROI), OTSU and maximum connected components are used to obtain the optimal treatment threshold and remove outliers, respectively. (2). Image denoising: in the nature scenery, most photos are collected in Gaussian noise environment. To improve the robustness, Gaussian filter with 3×3 kernel is utilized to depress image noises. (3) Image enhancement: it can be observed that microaneurysms, hemorrhage, and blood vessel have indistinguishable appearance in color space. If one aims to enhance image quality towards the direction of color variance, it is in vain for recognizing the three objects. Therefore, CLAHE (Contrast Limited Adaptive Histogram Equalization) is applied to enhance images in contrast [27]. (4) Image normalization: considering that the color and brightness of fundus images are quite different, we need to confine some parameters in our network model to a reasonable range. Otherwise, the overlarge parameters will slow the convergence speed of our model. Thus, we use normalization operation to speed up and boost the performance of our model at the same time. Formally, the normalized image can be generated as follows:

$$x_{\text{norm}} = \frac{x - u}{\theta}, \tag{3}$$

where x and x_{norm} denote the original image and normalized image, respectively. μ and θ are the mean value and standard derivation of all the samples in dataset IDRID. (5) Data augmentation: in contrast with traditional RGB images, collecting medical images is arduous. However, the performance of a deep neural network relies heavy on the scale of training data. Hence, we resort to common-used data augmentation strategies: random horizon flips, rotation, random crop, shift, and rescaling. (6) Image dicing: as we can see, the resolution of the original image in dataset IDRID is 4288×2848, which hinders the deep model from running in low-capacity devices. Besides, the areas occupied by lesions are usually relatively small,



(a) Lesions of diabetic retinopathy

(b) Positive samples

(c) Negative samples

Figure 4: Color fundus samples of dataset IDRID: (a) the sample contains microaneurysms, hard exudates, soft exudates, and hemorrhage; (b) positive samples; (c) negative samples.

Table 1: The distribution of IDRID.

Lesion type	Training set	Testing set
Microaneurysms	54	27
Hard exudates	54	27
Soft exudates	26	14
Hemorrhage	53	27

and the locations of lesions are scattered. So, we resolve to improving the performance of our model via image dicing technology. In Figures 4(b) and 4(c), motivated by the sliding window method, the dataset is divided into positive samples (with lesions) and negative samples (without lesions). As depicted in Figure 5, the detailed characteristics of the lesion area are clear enough, which is conducive to the subsequent lesion segmentation.

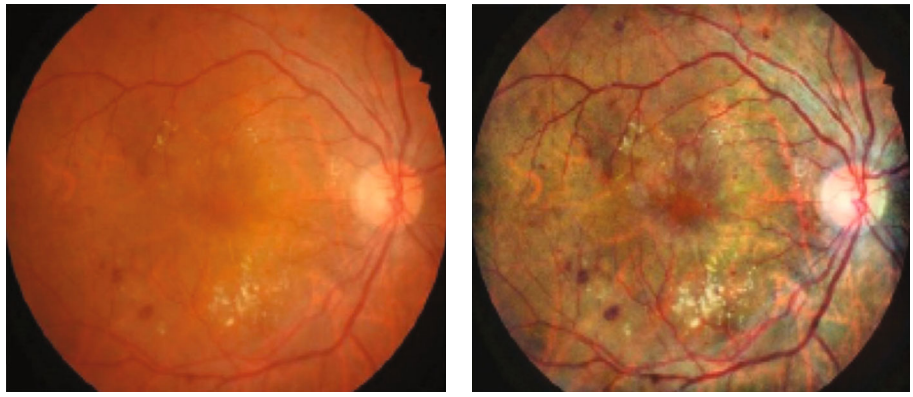
After the above image preprocessing operations, as displayed in Figures 4(b) and 4(c), the original high-resolution fundus images are transformed into several subimages with 256×256 pixels using the sliding window strategy with stride = 64. Then, the subimages with a black background are eliminated, and the remaining are treated as the valid input.

2.3.3. Fundus Image Postprocessing. After the above-mentioned image preprocessing, the whole image has been transformed into a group of subimages. For our trained model, the segmentation output has the same shape with the input subimage. Nevertheless, in real-world application, the pixels of the original image should be assigned with predicted labels in the final segmentation output. To achieve it, we attempt to merge these subimages to form the final segmentation result. The predicted label of a pixel is jointly determined by averaging the segmentation results of multiple subimages.

As mentioned in Fundus Image Preprocessing, the subimages are generated by the sliding window strategy. In this way, several subimages are overlapped inevitably. For the pixel inside the boundary, its final label will be assigned by averaging 16 subimage blocks. For the pixel on the boundary, it should be processed individually.

2.4. Experiments and Analysis

2.4.1. Training Parameters. All the experiments are executed on hardware devices with Intel Xeon CPU, 128 GB memory, and NVIDIA Tesla P100 GPU. The software environment is Ubuntu 16.04 operating system and PyTorch 1.0 framework. The input size is 256×256, and the batch size is set to 64. Since no pretrained model is provided, He initialization is used to initialize our model [28]. The network is trained by optimizing loss L for 100 epochs. As we all know, a higher and fixed learning rate cannot guarantee to bring better



(a) Original diabetic retina image (b) Diabetic retina image after preprocessing

Figure 5: The comparison between the original diabetic retinopathy image and its corresponding preprocessed result.

convergence to the deep neural network. Consequently, we adjust the learning rate as the training procedure goes on. The initial learning rate is set to 2×10^3. When the loss stops decreasing during training, the learning rate is reduced by a factor of 10. Also, the Adam optimizer with setting ($\beta_1 = 0.9$, $\beta_2 = 0.999$) is adopted. To be fair, all the comparative methods are implemented and line with the hyperparameters and parameters in their papers.

2.4.2. Evaluation Metrics. Evaluation metrics play an important role in measuring the performance of comparative models. In order to analyze the experimental results quantitatively, we use several specific metrics to evaluate the performance in the task of segmenting diabetic retinopathy image, including Sensitivity (SEN), Intersection-over-Union (IOU), and Dice coefficient (DICE). To implement them, we first calculated true positive (TP), false positive (FP), true negative (TN), and false negative (FN). TP here refers to the intersection of the true lesion area and predicted lesion area, FP denotes the intersection of the true normal area and predicted lesion area, TN is the intersection of the true normal area and predicted normal area, and FN is defined as the intersection of the true lesion area and predicted normal area. Based on the above concepts, we introduce the following metrics:

$$\begin{aligned} \text{SEN (sensitivity)} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{IOU} &= \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \\ \text{DICE} &= \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \end{aligned} \tag{4}$$

Let us take a close look at the three metrics. Sensitivity can be treated as the misdiagnosis rate of a disease. In our work, it refers to the proportion of TP and true lesion area, which is a critical and foremost factor for patients and doctors. In real-world application, we try to decrease the misdiagnosis rate to the best of our ability. IOU is an evaluation metric used to measure the accuracy of a semantic segmentation model, and it specifies the amount of overlap between the predicted results and the groundtruth. DICE is a measure of how similar the prediction and groundtruth are, which not only is a measure of how many positives the models predict but also penalizes for the false positives of the models. Regarding the above commonly used metrics, the closer they are to 1, the better the segmentation performance.

3. Results and Discussion

In this section, we conduct our experiments to evaluate the performance of our segmentation methods. The experiments include three parts: the first part makes ablation study of our method. It demonstrates the different performance brought by the components appearing in our methods. The second part makes user study to evaluate our method against several state-of-the-art methods on dataset IDRID. The last part describes the parameters and costs of all the methods to verify their efficiency.

3.1. Ablation Study. To better evaluate our proposed method, we design an ablation study by replacing each component and keeping the rest unchanged. We place particular emphasis on differences brought by four improvements discussed in Network Description. Thus, we conduct the following experiments.

Experiment 1: the original U-Net model is trained and tested on our testing samples.

Experiment 2: based on the original U-Net, the cross-entropy loss is replaced with Balanced Focal Loss function (denoted as U-Net-FL for convenience).

Experiment 3: based on Experiment 2, the pooling layers in encoders are replaced with 3×3 convolution layers, and all the activation functions are set to RReLu (denoted as U-Net V1 for convenience).

Experiment 4: on the basis of Experiment 3, the MSFF block to extract multiscale features is integrated into encoders (denoted as U-Net V2 for convenience).

Table 2: Ablation study of the proposed model against U-Net, U-Net V1, U-Net V2, and U-Net-FL on EX and SE.

Methods	EX			SE		
	SEN	IOU	DICE	SEN	IOU	DICE
FFU-Net	0.8755	0.8414	0.9138	0.7933	0.7876	0.8812
U-Net V2	0.8440	0.8159	0.8986	0.7547	0.7535	0.8594
U-Net V1	0.8033	0.7867	0.8769	0.6934	0.7028	0.8191
U-Net-FL	0.7929	0.7763	0.8704	0.6801	0.6893	0.8099
U-Net	0.7819	0.7602	0.8638	0.6713	0.6707	0.8029

Table 3: Ablation study of the proposed model against U-Net, U-Net V1, U-Net V2, and U-Net-FL on MA and HE.

Methods	MA			HE		
	SEN	IOU	DICE	SEN	IOU	DICE
FFU-Net	0.5933	0.5610	0.7188	0.7342	0.7365	0.8450
U-Net V2	0.5508	0.5267	0.6669	0.6936	0.6917	0.8177
U-Net V1	0.5172	0.4891	0.6334	0.6598	0.6562	0.7897
U-Net-FL	0.4968	0.4626	0.6255	0.6447	0.6425	0.7797
U-Net	0.4810	0.4490	0.6197	0.6366	0.6333	0.7755

Experiment 5: on the basis of Experiment 4, the CCA module is deployed to fuse skip connection and lower-resolution decoder (denoted as FFU-Net for convenience).

All the above experiments are performed on a preprocessed dataset, and the quantitative results are illustrated in Tables 2 and 3. Obviously, FFU-Net consistently outperforms U-Net on all metrics in the task of segmenting lesions. This improvement is mainly attributed to MSFF, CCA, and Balanced Focal Loss. Using Balanced Focal Loss, U-Net-FL increase IOU by up to an average 0.031 points on all lesion types, which proves that Balanced Focal Loss function is capable of coping with data imbalance and misclassification in the segmentation task. After the pooling layers are replaced with 3×3 convolution layers and all the activations are set to RReLu, U-Net V1 achieve slightly better than U-Net-FL. The introduction of MSFF brings more improvement on metrics SEN, IOU, and DICE, which verifies the effectiveness of MSFF block to lesion segmentation for diabetic retinopathy. With the help of CCA, FFU-Net achieves the DICE value increased by 0.0291 points, and the IOU value increased by 0.0347 points. Note that in the analyses of CCA and MSFF, we find that they surpass U-Net V1 by a large margin on all metrics. This indicates that the components of CCA and MSFF play more critical roles in segmentation of medical images.

In Figure 6, we visually present the segmentation results of different methods on dataset IDRID. It can be seen that U-Net and U-Net-FL cause too many defects with lower accuracy. Seen from the prediction results by U-Net V1, we observe that it can provide more clear boundaries than U-Net-FL. Since MSFF is utilized in the encoders, it appears that U-Net V2 produce clear and pleasing segmentation results. Nevertheless, we find that U-Net V2 fails to recognize the lesion with smaller size (MA). By incorporating the CCA module, FFU-Net aid in refining the details of lesions, leading to closer segmentation result to the groundtruth. Therefore, we can safely draw the conclusion that the improvements mentioned in FFU-Net are effective quantitatively and qualitatively.

3.2. User Study. To confirm the effectiveness and robustness of our proposed method, we conduct a user study against the state of the art on metrics SEN, IOU, and DICE. The comparative methods include the Dai et al. method [29], Zhang et al. method [30], Van Grinsven et al. method [17], M-Net [31], FC-DenseNet [32], Sambyal et al. method [18], and original U-Net. To further show our superiority, we, respectively, display the segmentation quantitative results on four lesion types in Tables 4 and 5. As can be seen, FFU-Net claims its superiority over the others on segmenting all the lesions. In comparison with the second best method (Sambyal et al.), FFU-Net achieves the DICE value increased by 2.0% and the IOU value increased by 3.5%. As reported in [29], the Dai et al. method is designed for timely detection and treatment of MA, which is consistent with our results in Table 5. However, it is unable to cope with the detection of other lesion types (EX and SE). Similarly, Zhang et al. aim to automatically detect exudates in color eye fundus images and perform better in segmenting EX and SE but work worse in segmenting MA and HE. Van Grinsven et al. solve the unbalanced problem by dynamically selecting misclassified negative samples and apply CNN to HE segmentation. The results reported in work [17] are verified in our experiment. Limited by the lack of generalization ability, Van Grinsven et al. are incapable of processing EX, SE, and MA perfectly. Although M-Net achieves state-of-the-art OD and OC segmentation results on the glaucoma dataset, it fails to transfer to our IDRID well. Besides, FC-DenseNet extends DenseNet to deal with the problem of semantic segmentation on natural images. When applying it to IDRID, it cannot show enough ability of presenting irregular microlesions. Sambyal et al. employ periodic shuffling with subpixel convolution initialized to convolution nearest neighbor resize. As we all know, the subpixel strategy is a common trick in the superresolution task. Whereas in Figure 7, we found more holes in the segmentation results, leading to unsatisfactory quantitative results on all metrics. Benefiting from the MSFF, CCA, and Balanced Focal Loss, our proposed FFU-Net achieves consistent improvement to all existing methods on all three performance metrics. Figure 7 shows some visual examples of four lesion types, where we observe that our method could generate closer results to the groundtruth without introducing additional artifacts. Apparently, we can see that Dai et al., Zhang et al., and Van Grinsven et al. suffer from inaccurate prediction for the boundaries of all lesion types. Also, the failure of M-Net and FC-DenseNet in transferring to all image samples is attributed to their poor generalization ability. Therefore, it can safely come to the conclusion that FFU-Net achieves comparable performance quantitatively and qualitatively.

3.3. The Overhead of Parameters and Computation. It is necessary to analyze the overhead of parameters and

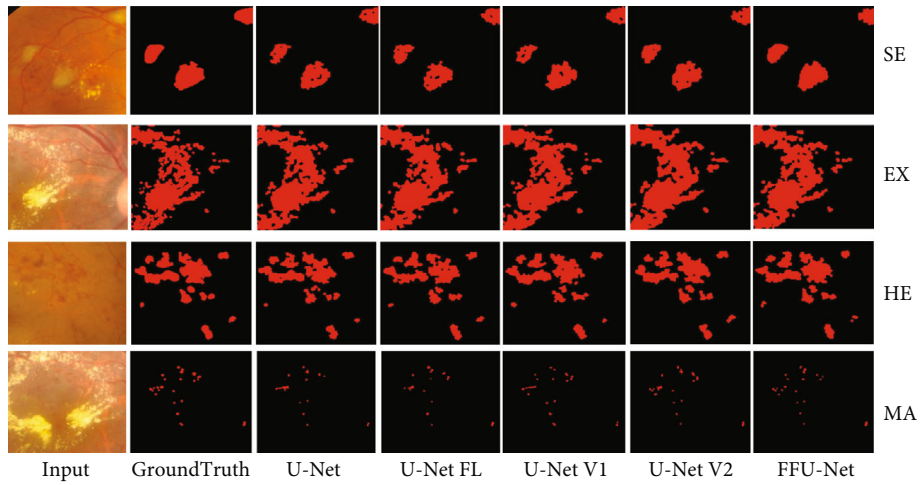


Figure 6: The visual segmentation results of U-Net, U-Net-FL, U-Net V1, U-Net V2, and FFU-Net. Zoom in to see the details.

Table 4: Comparative segmentation results of the proposed model against the state of the art on EX and SE.

Methods	EX SEN	EX IOU	EX Dice	SE SEN	SE IOU	SE DICE
Dai et al.	0.8074	0.7843	0.8791	0.7006	0.7071	0.8284
Zhang et al.	0.8418	0.8137	0.8973	0.7523	0.7505	0.8575
Van Grinsven et al.	0.8031	0.7749	0.8732	0.6988	0.692	0.818
M-Net	0.8327	0.8083	0.894	0.7297	0.7156	0.8343
FC-DenseNet	0.8414	0.8099	0.8949	0.7554	0.7623	0.8651
Sambyal et al.	0.8421	0.8183	0.9001	0.7563	0.763	0.8656
FFU-Net	0.8755	0.8414	0.9138	0.7933	0.7876	0.8812
U-Net	0.7819	0.7602	0.8638	0.6713	0.6707	0.8029

Table 5: Comparative segmentation results of the proposed model against the state of the art on MA and HE.

Methods	MA SEN	MA IOU	MA DICE	HE SEN	HE IOU	HE DICE
Dai et al.	0.5498	0.5237	0.6874	0.6895	0.6990	0.8228
Zhang et al.	0.4897	0.4723	0.6416	0.6418	0.6407	0.7810
Van Grinsven et al.	0.4832	0.4667	0.6364	0.6844	0.6761	0.8068
M-Net	0.5366	0.5097	0.6753	0.6872	0.6796	0.8093
FC-DenseNet	0.5521	0.5276	0.6908	0.6976	0.6960	0.8208
Sambyal et al.	0.5537	0.5438	0.7045	0.6998	0.7038	0.8261
FFU-Net	0.5933	0.5610	0.7188	0.7342	0.7365	0.8450
U-Net	0.4810	0.4490	0.6197	0.6366	0.6333	0.7755

computation of our comparative methods. Notably, all comparisons are evaluated on the same machine. Evidently, as seen in Table 6, the Dai et al. method and Zhang et al. method are significantly lighter than other models, but this comes at the price of an apparent performance drop. With respect to the Van Grinsven et al. method, it solves the segmentation task through a CNN pixel-wise classifier. Whereas without taking spatial relationship into account, Van Grinsven et al. cannot achieve pleasing results. Since FC-DenseNet has more dense residual modules and more than 100 layers, it needs more time and more parameters in the testing procedure. As another modified U-Net, the Sambyal et al. method employs periodic shuffling with subpixel convolution based on U-Net, so it will take more time to implement in our application. By introducing splitting operation into FFU-Net, we observe that FFU-Net elapses less time while making noticeable improvement on segmentation performance. From the above discussions, it is observed that perhaps FFU-Net is the best choice when considering the influences between various factors.

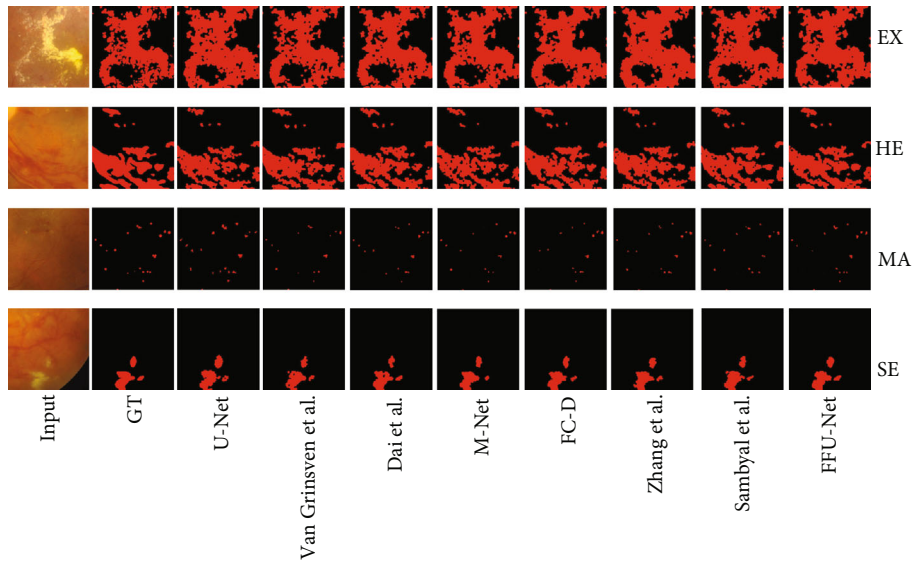


Figure 7: The visual comparative results for segmentation on dataset IDRID. GT: groundtruth; FC-D: FC-DenseNet. Zoom in to see the details.

Table 6: The overhead of parameters and computation of different comparative models.

Models	Running time	Parameters
Dai et al. method	616 ms	—
Zhang et al. method	688 ms	—
Van Grinsven et al. method	2598 ms	0.98 M
M-Net	3745 ms	1.67 M
FC-DenseNet	4361 ms	1.73 M
Sambyal et al. method	1535 ms	1.33 M
FFU-Net	695 ms	0.97 M
U-Net	780 ms	1.93 M

4. Conclusion and Future Work

Based on the original U-Net network, we propose a new model named FFU-Net which is suitable for lesion segmentation of diabetic retinopathy. The FFU-Net network model mainly has the following contributions: The original pooling layer is replaced with a convolutional layer to reduce the spatial loss of the fundus image. MSFF block is incorporated to extract multiscale features and speed up feature fusion with splitting operation. By virtue of the CCA module, FFU-Net fuses the information between skip connection and lower-resolution decoder with shared attention weights. Considering the data imbalance problem in diabetic retinopathy, we present a Balanced Focal Loss function. Finally, in order to verify the effectiveness of our proposed model, ablation study and user study are carried out on the public benchmark IDRID. The final experimental results demonstrate the effectiveness and advancement of our proposed FFU-Net in terms of almost all metrics.

In the future, we will investigate a more general and comprehensive segmentation method for diabetic retinopathy and put emphasis on the following points: (1) Few-shot learning: though we solve the overfitting problem caused by insufficient data by data slicing, the burden of collecting large-scale supervised data for real-world application is still challenging. Thus, we resort to few-shot learning to achieve better segmentation. (2) Contaminated labels: different from the benchmark that is refined and maintained by professionals, the practical images of diabetic retinopathy are vulnerable to be contaminated and damaged. Thus, we should learn how to segment the lesion images only with incomplete and contaminated labels. (3) Grading the severity of diabetic retinopathy: as a foundation work, we plan to expand our work to grade the severity of diabetic retinopathy and apply our achievements to real-world application.

Data Availability

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

Conflicts of Interest

There is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was funded in part by the Young Scientists Fund of the National Natural Science Foundation of China under Grant 61802300, China Postdoctoral Science Foundation Funded Project under grant 2018m643666, Xi'an Jiaotong University basic research foundation for Young Teachers under grant xjh012019043, and National Key Research and Development Project under grants 2019YFB2102501 and 2019YFB2103005.

References

[1] S. Wang, X. Wang, Y. Hu et al., "Diabetic Retinopathy Diagnosis Using Multichannel Generative Adversarial Network With Semisupervision," *IEEE Transactions on Automation Science and Engineering*, pp. 1–12, 2020.

[2] S. Feng, H. Zhao, F. Shi et al., "Cpfnet: context pyramid fusion network for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 3008–3018, 2020.

[3] A. Grzybowski, P. Brona, G. Lim et al., "Artificial intelligence for diabetic retinopathy screening: a review," *Eye*, vol. 34, no. 3, pp. 451–460, 2020.

[4] H. A. Nugroho, K. W. Oktoeberza, I. Ardiyanto, R. L. B. Buana, and M. B. Sasongko, "Automated segmentation of hard exudates based on matched filtering," in *2016 International Seminar on Sensors, Instrumentation, Measurement and Metrology (ISSIMM)*, pp. 84–87, Malang, Indonesia, 2016.

[5] I. A. Masoud Abdulhamid, A. Sahiner, and J. Rahebi, "New Auxiliary Function with Properties in Nonsmooth Global Optimization for Melanoma Skin Cancer Segmentation," *BioMed Research International*, vol. 2020, Article ID 5345923, 14 pages, 2020.

[6] H. Kayadibi, E. Sertoglu, and M. Uyanik, "Plasma Total Homocysteine Levels in Diabetic Retinopathy," *BioMed Research International*, vol. 2014, Article ID 758634, 2 pages, 2014.

[7] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1223–1232, 2006.

[8] S. Kavitha and K. Duraiswamy, "Automatic detection of hard and soft exudates in fundus images using color histogram thresholding," *European Journal of Scientific Research*, vol. 48, no. 3, pp. 493–504, 2011.

[9] M. U. Akram, I. Jamal, A. Tariq, and J. Imtiaz, "Automated segmentation of blood vessels for detection of proliferative diabetic retinopathy," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, vol. 429, pp. 232–235, Hong Kong, China, 2012.

[10] 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*, vol. 9351, pp. 234–241, Springer, Cham, 2015.

[11] B. Antal and A. Hajdu, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE transactions on biomedical engineering*, vol. 59, no. 6, pp. 1720–1726, 2012.

[12] 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.

[13] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 4, pp. 640–651, 2017.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, Boston, MA, USA, 2015.

[15] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[16] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: multi-path refinement networks for high-resolution semantic segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1925–1934, Honolulu, HI, USA, 2017.

[17] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sanchez, "Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1273–1284, 2016.

[18] N. Sambyal, P. Saini, R. Syal, and V. Gupta, "Modified u-net architecture for semantic segmentation of diabetic retinopathy images," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1094–1109, 2020.

[19] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, 2015.

[20] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 2024–2032, Nice, France, 2019.

[21] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 472–480, Honolulu, HI, USA, 2017.

[22] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, Salt Lake City, UT, USA, 2018.

[23] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, https://arxiv.org/abs/1505.00853.

[24] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, vol. 472, pp. 2980–2988, Venice, Italy, 2017.

[25] P. Porwal, S. Pachade, R. Kamble et al., "Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, 2018.

[26] L.-W. Wang, Z.-S. Liu, W.-C. Siu, and D. P. Lun, "Lightening network for low-light image enhancement," *IEEE Transactions on Image Processing*, vol. 29, pp. 7984–478 7996, 2020.

[27] E. D. Pisano, S. Zong, B. M. Hemminger et al., "Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, no. 4, pp. 193–200, 1998.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, Santiago, Chile, 2015.

[29] L. Dai, B. Sheng, Q. Wu et al., "Retinal microaneurysm detection using clinical report guided multi-sieving cnn," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, vol. 10435, pp. 525–532, Springer, Cham, 2017.

[30] X. Zhang, G. Thibault, E. Decencière et al., "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Medical image analysis*, vol. 18, no. 7, pp. 1026–1043, 2014.

[31] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1597–1605, 2018.

[32] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 11–19, Honolulu, HI, USA, 2017.

Hindawi
BioMed Research International
Volume 2020, Article ID 6687733, 12 pages
https://doi.org/10.1155/2020/6687733

Research Article

Medical Image Retrieval Using Empirical Mode Decomposition with Deep Convolutional Neural Network

Shaomin Zhang, Lijia Zhi, and Tao Zhou

School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China

Correspondence should be addressed to Lijia Zhi; ljzhi_paper@163.com

Received 8 November 2020; Revised 7 December 2020; Accepted 14 December 2020; Published 28 December 2020

Academic Editor: Lin Gu

Content-based medical image retrieval (CBMIR) systems attempt to search medical image database to narrow the semantic gap in medical image analysis. The efficacy of high-level medical information representation using features is a major challenge in CBMIR systems. Features play a vital role in the accuracy and speed of the search process. In this paper, we propose a deep convolutional neural network- (CNN-) based framework to learn concise feature vector for medical image retrieval. The medical images are decomposed into five components using empirical mode decomposition (EMD). The deep CNN is trained in a supervised way with multicomponent input, and the learned features are used to retrieve medical images. The IRMA dataset, containing 11,000 X-ray images, 116 classes, is used to validate the proposed method. We achieve a total IRMA error of 43.21 and a mean average precision of 0.86 for retrieval task and IRMA error of 68.48 and F1 measure of 0.66 on classification task, which is the best result compared with existing literature for this dataset.

1. Introduction

Imaging through different kinds of medical devices plays a fundamental role in clinical diagnosis [1], treatment planning [2], and treatment response assessing [3] in the process of medical care. In modern hospitals, different modalities and protocols of digital imaging techniques have been used to generate diagnostic images for each patient, including computed tomography (CT), X-ray, ultrasound, hybrid positron emission tomography and computed tomography (PET-CT), and magnetic resonance imaging (MRI). These medical images with multiple dimensions (e.g., 2D, volumetric 3D, and time series) reflect anatomic and functional aspects of organs and tissue types that require domain experts' analysis and interpretation. These volumes are usually formed in the Digital Imaging and Communications in Medicine (DICOM) format and stored in picture archiving and communication systems (PACS) [4]. A domain expert can search PACS through patient's ID, study ID, time range, or other textual keywords, which is labor intensive and time consuming. As an important part of computer-aided diagnostics (CAD), content-based medical image retrieval (CBMIR) [5–8] can retrieve medical images mainly via visual contents (e.g., same modality, same body orientation, same anatomical region, or same disease condition) in an existing dataset for more accurate comparative diagnosis.

In the CBMIR domain, there are two major directions in research works. One kind of methods focuses on automatic retrieving images from PACS-like databases, which search images of the same imaging modality, body orientation, body region, and the like [9–11]. Another kind of methods put their efforts into retrieving images that characterize the similar disease convenient for diagnostic comparing [12, 13]. In this study, we follow the former methods to propose an effective CBMIR system for 2D slice retrieval. That is because volumetric 3D medical images are formed by a series of 2D slices acquired from the target body organ, and physicians mainly rely on these 2D slices when they are analyzing and interpreting images on hand [8].

Unlike similarity defined in generic image retrieval domain, the retrieved medical images by directly comparing features using some similarity measure may not be in

accordance with what a physician would want for diagnosis, which formed a "semantic gap" in medical image retrieval [5]. To reduce this gap, CBMIR systems are generally designed under a classification-driven strategy. That is, a CBMIR system is trained using supervised approaches with labeled images. When a query image is submitted to the CBMIR system, the query image is classified first, and then, some visual features and similarity measures are used for similarity retrieval [9, 10, 14]. Deep learning is a breakthrough in machine learning research. Using artificial neural networks with many hidden layers to represent digital images has been proven to be a very effective method to describe low-level, mid-level, and high-level semantic features of an image for recognition and other purposes [15–17]. Among different deep learning architectures, deep convolutional neural networks (CNNs) have proven to be powerful tools that achieved very high precision results in many natural image classification contests [18–20]. In the medical field, deep CNNs are also quickly applied for different tasks, and promising results are emerging [9, 10, 21–24]. Training deep CNNs need a large number of labeled images to choose the huge number of parameters. Note that in the medical domain, such large image datasets are quite rare, due to the unbearable high cost of domain experts' manual image labeling and annotations [5, 21, 24]. And in contrast to generic image databases, medical image datasets usually are unbalanced because of uneven incidence rates of different malignancies. Dropout [25], data augmentation, and transfer learning [26] are the most common techniques used to prevent overfitting in the process of training deep CNNs on small and unbalanced image datasets. However, for medical image analysis tasks, these techniques meet various problems [5, 9, 24]; the requirement for a more effective and more robust CBMIR system is still urgent.

In this paper, inspired by pioneering research works [9, 10, 14], we focus on 2D medical image retrieval and put forth an effort to alleviate the two main difficulties in CBMIR (i.e., (1) the labeled medical image datasets are commonly not large enough for training deep CNNs and (2) the imbalance problem is naturally attached to medical image datasets from clinic diagnosis). A new deep CNN-based 2D medical slice retrieval method is proposed, which can be effectively trained on relatively small labeled and unbalanced medical image dataset and promote the retrieval precision. First, in addition to commonly used methods for training deep CNNs on small and unbalanced datasets, e.g., dropout [25] and data augmentation, we supplement nonlinear components by using empirical mode decomposition (EMD) on 2D medical images to enhance effective information and reduce the image noise for training deep CNNs. Second, as for deep CNN architecture in this work, we employ residual network (ResNet) [19] as the backbone network adapted for learning different level features from medical images, which is combined with an attention mechanism to focus on the most relevant features by integrating local and global features in different scales [27]. And center loss function is combined with softmax loss function as a supervision signal in a deep CNN training process to facilitate nearest-neighbor similarity retrieval performance. The contributions of this paper are given as follows:

(1) Nonlinear empirical mode decomposition on 2D medical images is proposed for supplementing effective information to original 2D medical images for better distinctively expressing 2D medical images

(2) A residual network-based deep CNN model with attention and center loss modules is employed and trained on publicly available medical image datasets. The learned concise feature vectors are suitable for both classification-based and nearest-neighbor similarity-based medical image retrieval and show the great potential to handle large-scale medical image retrieval

2. Related Work

Among CBMIR literatures, there are two crucial factors that determine the performance of systems: (1) *Feature vector construction*: medical image features such as texture, shape, etc., should be extracted and formed into a vector to represent the query image and the images in datasets. (2) *Retrieval strategy*: classification-based retrieval strategy, nearest-neighbor search strategy, or their combination should be carefully chosen for different medical retrieval task.

2.1. Hand-Crafted Features. Hand-crafted features including texture features, keypoint-based features, local features, and global features are commonly used in CBMIR systems [5, 6, 8, 28, 29]. Jiang et al. [30] proposed a retrieval strategy that used mammographic region of interest (ROI) as query input, then retrieve breast tumor based on SIFT features. Caicedo et al. [31] used SIFT features to retrieve basal-cell carcinoma. Haas et al. [32] used SURF to capture the local texture of lung CTs for retrieval. Local Binary Patterns (LBPs) as local texture features were successfully used in ImageCLEFmed, 2D-Hela, and brain MRI retrieval tasks [33–35]. Xu et al. [36] proposed a corner-guided partial shape matching method that can dramatically increase the matching speed for spine X-ray image retrieval. Holistic features such as global GIST, global HOG, global color histogram, and moments were also used in medical image retrieval [37–41].

2.2. Learned Features Using Deep CNNs. In recent years, using features get through deep CNNs has achieved impressive results in generic image classification, object recognition, detection, retrieval, and other related tasks. But in the medical field, there is not much attention on exploring deep neural networks CBMIR task, partially because the amount of labeled medical images is typically limited. Qayyum et al. [10] proposed a CNN framework and trained the CNN on the medical image set they collected. Khatami et al. [9, 14] tried two retrieval strategies for medical image retrieval: the first method used one CNN model with transferred weights to shrink the search space and then used Radon projection to do similarity search. The second method employed multiple CNN models trained in a parallel way to get the shrunk search space. Bar et al. [42] used a pretrained CNN model from natural images for chest X-ray retrieval. Semedo and Magalhães [43] trained their CNN models on provided medical images in ImageCLEFmed 2016; they employed dropout

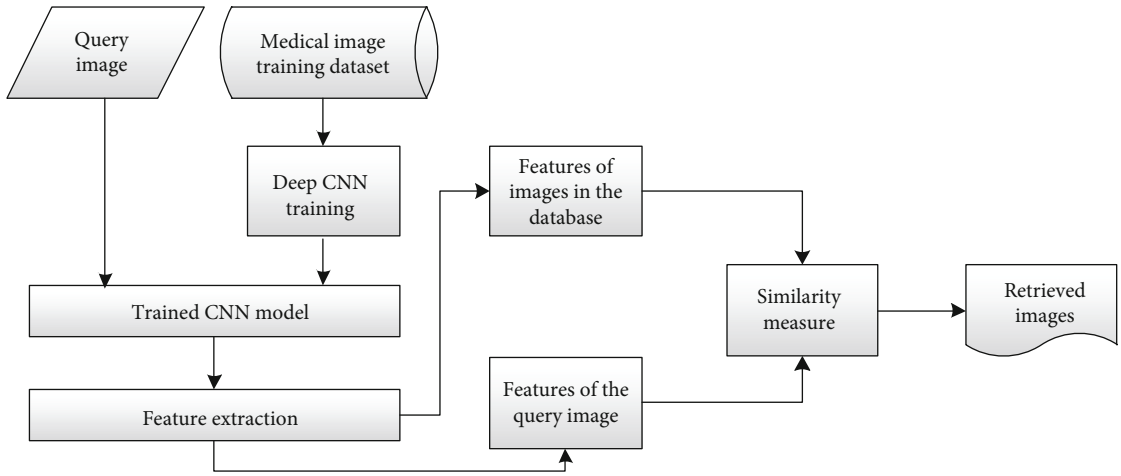


Figure 1: The proposed deep CNN-based content-based medical image retrieval flowchart.

and data augmentation to avoid overfitting. Hofmanninger and Langs [44] trained CNN using clinical routine images and radiology reports and carried out fine-tuning on current medical image retrieval task.

3. Methodology

There are pioneering studies that have been performed on deep CNNs for medical image retrieval and have shown promising results [9, 10, 14]; the problems of short of labeled images and highly imbalanced data distribution are still two main challenges for applying deep CNNs in medical image retrieval task [5]. There are also needs for more accurate and faster image retrieval methods for CBMIR [5]. To tackle these problems, in this work, we propose a multicomponent combined deep CNN framework for 2D medical image retrieval. The flowchart of content-based medical image retrieval is shown in Figure 1. This deep convolutional neural network is trained by a supervised learning way for classification and gets a concise feature vector for efficient nearest neighbor searching similar medical images. A brief description of the proposed framework is presented in the following sections.

3.1. Processing 2D Medical Image with Empirical Mode Decomposition (EMD). Empirical mode decomposition was originally introduced for the adaptive analysis of nonstationary and nonlinear time-domain signals and has become one of the most powerful tools for analyzing time-frequency (T-F) signal [45]. Then, EMD was extended to handle multidimensional data and acquired successful application in image tasks [46–49]. For image analysis, EMD is a fully data-adaptive multiresolution data analysis technique to decompose the multispatial resolution spatial-frequency-amplitude components of the image into a set of intrinsic mode functions (IMFs) [50, 51]. By advantage of the EMD principle, we can get multifrequency components (i.e., IMFs) of 2D medical images, and these frequencies are not predesigned; these frequencies can self-adapt to different content of an image. Thus, we acquire nonstationary and nonlinear multiresolution components of 2D medical images, which can provide supplementary information to commonly used spatial filter sets in image processing. EMD is implemented in an iterative process. First, a sifting process is used to find IMFs. Given a signal $x(t)$, Equation (1) is the process to get one IMF.

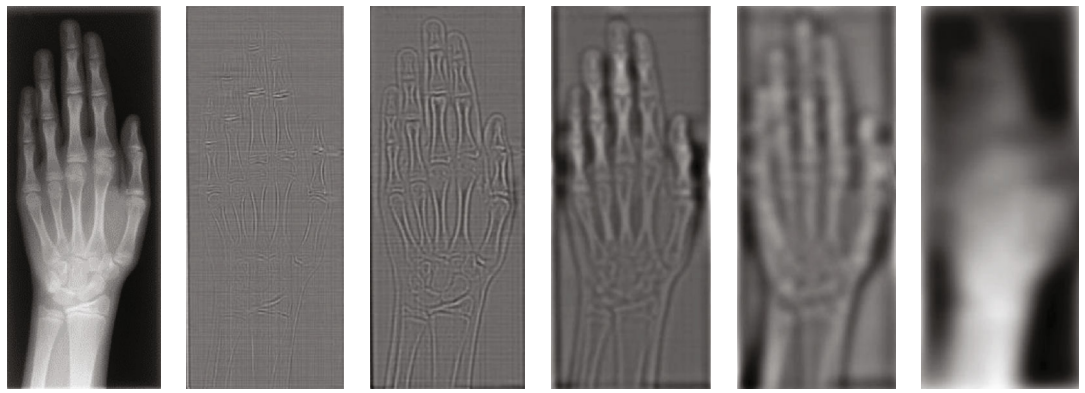
$$x(t) - \sum_{i=1}^{k} m_i = h_k \Rightarrow h_k = c_1, \tag{1}$$

where m_i is the local mean of the maxima and minima envelopes. These two envelopes are formed by connecting all local maxima or minima with a cubic spline. With the IMFs, the data $x(t)$ can be decomposed by another sifting process:

$$x(t) - \sum_{j=1}^{n} c_j + r_n, \tag{2}$$

where c_j ($j = 1$ to n) is the IMFs and r_n is the final residual component. Figure 2 shows an example of a 2D X-ray image decomposed using EMD.

3.2. The Proposed Medical Image Retrieval Method. In this section, we introduce a deep CNN framework for medical image retrieval on a rather small dataset and with highly imbalanced data distribution. First, we discuss the network architecture employed in this work. Second, the supervision signal combining softmax loss function with center loss function to train deep CNN is discussed. Third, the training process is detailed. The proposed deep CNN framework is illustrated in Figure 3. For the input of the network, we employ original image and its IMF2, IMF3, and IMF4 components, because IMF1 contains mainly noise with quite high spatial frequency, and IMF5 contains the overall image intensity trend with very low spatial frequency. For medical image classification, IMF1 and IMF5 cannot provide useful structure information.



(a) Original (b) IMF1 (c) IMF2 (d) IMF3 (e) IMF4 (f) IMF5

Figure 2: An example of a medical X-ray image is decomposed into five IMFs using empirical mode decomposition.

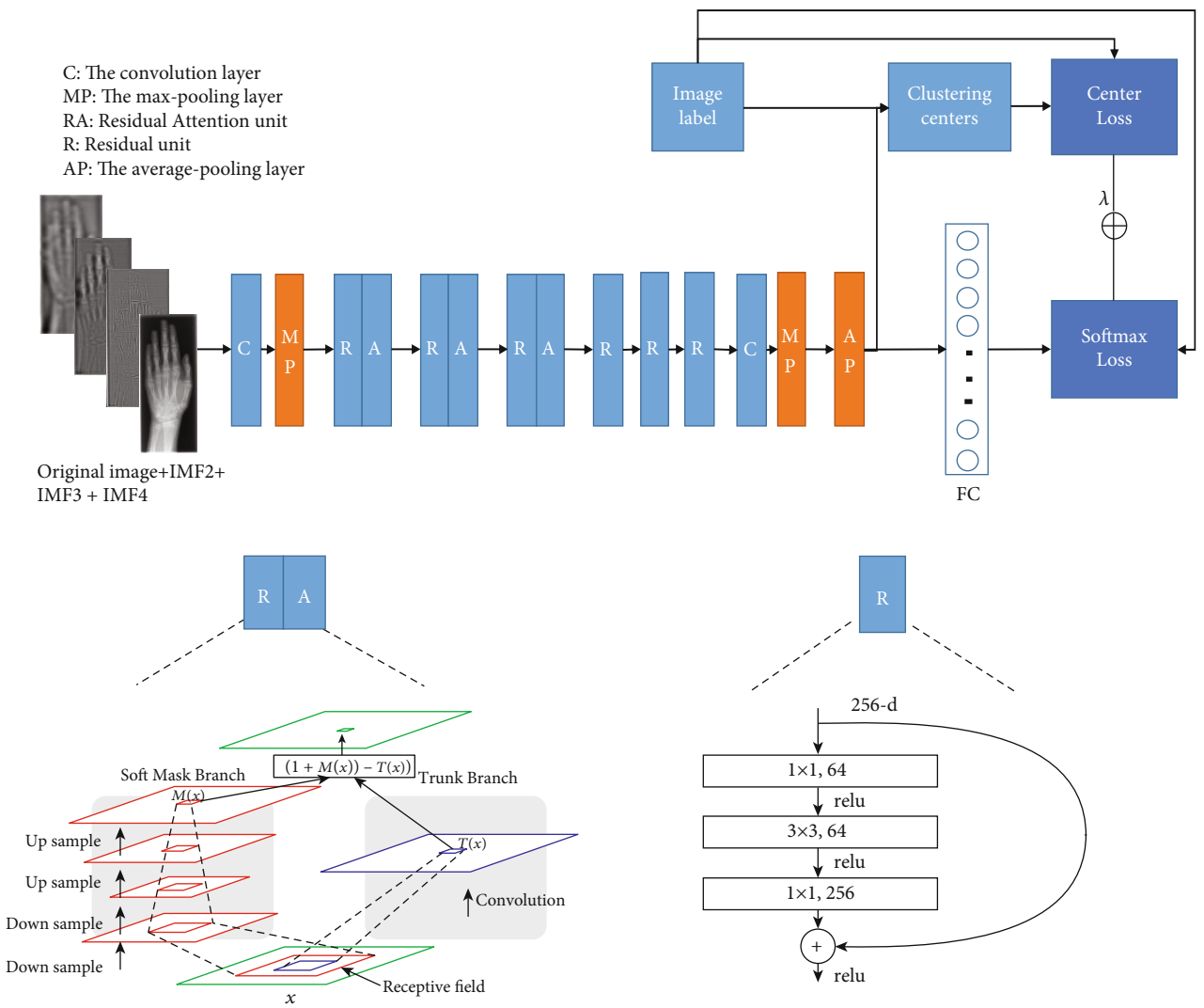


Figure 3: The proposed deep CNN framework for content-based medical image retrieval.

3.2.1. The Network Architecture. The proposed deep CNN architecture employs Residual Attention Network (RAN) [27] as the backbone network. In RAN, mixed attention activation function is used for both spatial and channel attention. The attention mechanism was implemented as multiple attention modules, and each module consisted of a mask branch and a trunk branch, in which the mask branch was used to select good properties of original features and

Table 1: The details of deep CNN used for medical image retrieval.

Layer name	Output size	Layer
Conv1	128×128	7×7, 64, stride 2
Max pooling	64×64	3×3 stride 2
Residual attention unit	64×64	$\begin{pmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1,256 \end{pmatrix} \times 1$ Attention mask block $\times 1$
Residual attention unit	32×32	$\begin{pmatrix} 1 \times 1, 128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{pmatrix} \times 1$ Attention mask block $\times 1$
Residual attention unit	16×16	$\begin{pmatrix} 1 \times 1, 256 \\ 3 \times 3,256 \\ 1 \times 1, 1024 \end{pmatrix} \times 1$ Attention mask block $\times 1$
Residual unit	8×8	$\begin{pmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{pmatrix} \times 3$
Conv2	8×8	3×3, 32, stride 1, padding: same
Max pooling	4×4	2×2 stride 1
Average pooling	1×1	4×4 stride 1
FC, softmax, Center loss	116	
Trunk depth	57	

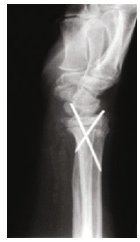
suppress noises from trunk features. Residual learning was introduced in the learning process of RAN; the mask branch was constructed as identical mapping. With the residual learning, Residual Attention Network can go very deep, and the training process was much efficient. For medical image retrieval task, the nearest-neighbor similarity search is the most common way used to rank retrieved images. If the length of vector used to compute the similarity between two compared medical images is too long, the retrieval process will be very time consuming and cannot be used in practice. Thus, a dimensionality reduction model is added to get concise while strong distinguishing features. Table 1 details the CNN structure used in this work.

3.2.2. Joint Loss Function. Wen et al. [52] firstly introduced center loss function in deep CNN for face recognition task. In their work, center loss function was linearly jointed with softmax loss function to form a mixture supervision signal to train deep CNN. These two loss functions that were used in conjunction with each other can achieve discriminative feature learning, that is, the deeply learned features contained intraclass compactness and interclass dispersion. Discriminative features are very suitable for medical image classification and retrieval task in which nearest-neighbor similarity search is most commonly used to accomplish the retrieve. Equation (3) formulates this joint loss function.

$$\begin{aligned} \text{Loss}_{\text{mixture}} = \text{Loss}_{\text{softmax}} + \lambda \text{Loss}_{\text{center}} &= -\sum_{i=1}^{m} \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_{y_i}^T x_i + b_{y_i}}} \\ &+ \frac{\lambda}{2} \sum_{i=1}^{m} \left\| x_i - c_{y_i} \right\|_2^2, \end{aligned} \tag{3}$$

where the left part is the original softmax loss and the right part is the center loss. The c_{yi} denotes the y_ith class center in the form of a feature vector. The parameter λ is empirically set as 0.002 in this paper's experiments.

3.2.3. Network Training Setting. As shown in Figure 3, the input of the network is the original medical image with its EMD components that contain IMF2, IMF3, and IMF4 got from EMD. The network training is developed and trained by using Keras on TensorFlow. The training processes are performed on a workstation with Ubuntu 18.04, having Intel(R) Xeon(R) Gold 6154 CPU with 256G RAM, and NVIDIA TITAN V graphic card with 12G RAM. Data



IRMA Code	1121-240-422-700
Technical code	X-ray, plain radiography, analog, overview image
Directional code	Sagittal, lateromedial
Anatomical code	Upper extremity (arm), radio carpal joint, right carpal joint
Biological code	Musculosceletal system

Figure 4: A sample image (arm) with the corresponding IRMA code.

argumentation and dropout are employed in the training process. The number of epochs is 500, the batch size is 16, the initial learning rate is 0.0001, and early stopping is on. When the network accuracy is not improved within 20 training iterations, the early stopping mechanism will be triggered. The 500 epoch setting is to make sure that in most cases, the network training is stopped by the early stopping mechanism.

4. Experimental Results

In this paper, the very challenging IRMA dataset is chosen to evaluate the proposed framework and compare with other methods reported in the literature. The proposed CNN model is evaluated in terms of classification performance and retrieval performance, respectively.

4.1. Database Description. IRMA (Image Retrieval in Medical Applications) database is a well-known medical image dataset for content-based medical image retrieval research, which was made by Aachen University of Technology (RWTH) [53]. This dataset was arbitrarily selected from a routine at the Department of Diagnostic Radiology, Aachen University of Technology. IRMA code is used to specify each image's class along four independent hierarchical axes: TTTT-DDD-AAA-BBB. In this code, T represents the technical code (imaging modality), D represents the directional code (body orientations), A represents the anatomical code (the body region examined), and B represents the biological code (the biological system examined). This dataset contains a total of 12,000 images divided into 116 classes, 11,000 image radiographs with known categories for training, and the rest 1000 radiographs as test. Figure 4 illustrates a sample image with the corresponding IRMA code.

4.2. Classification Performance

4.2.1. IRMA Error. ImageCLEF07 proposed the error evaluation procedure for IRMA Medical Image Annotation to calculate the retrieval error [54, 55]. The total IRMA error can be computed by the following formula:

$$\sum_{i=1}^{I} \frac{1}{b_i} \frac{1}{i} \delta\left(l_i, \hat{l}_i\right) \text{ with } \delta\left(l_i, \hat{l}_i\right) = \begin{cases} 0, & \text{if } l_j = \hat{l}_j, \forall j \le i, \\ 0.5, & \text{if } l_j = *, \exists j \le i, \\ 1, & \text{if } l_j \ne \hat{l}_j, \exists j \le i. \end{cases} \tag{4}$$

Here, $l_1^I = l_1, l_2, \cdots, l_i, \cdots, l_I$ is the *correct* code (for one axis) of an image, and $\hat{l}_1^I = \hat{l}_1, \hat{l}_2, \cdots, \hat{l}_i, \cdots, \hat{l}_I$ is the *classified* code (for one axis) of an image. I is the depth of the tree to which the classification is specified. If there is an incorrect classification at position $\hat{l}_I$, all succeeding decisions will be considered as wrong decisions.

4.2.2. Commonly Used Classification Performance Measure. To evaluate the performance of different methods for classification task, commonly used performance evaluation indicators include average precision (AP), average recall (AR), and F1 measure. These indicators are calculated as the following:

$$\begin{aligned} \text{AP} &= \frac{1}{M} \sum_{i=1}^{M} \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}, \\ \text{AR} &= \frac{1}{M} \sum_{i=1}^{M} \frac{\text{TP}_i}{\text{TP}_i + \text{TN}_i}, \\ \text{F1 measure} &= 2 \times \frac{\text{AP} \times \text{AR}}{\text{AP} + \text{AR}}, \end{aligned} \tag{5}$$

where TP is true positive, indicating the number of images correctly classified as class k; FP is false positive, indicating the number of images misclassified as class k; TN is true negative, indicating the number of images correctly classified as not class k; FN is false negative, indicating the number of images misclassified as not class k; and M means the total

Table 2: Comparison of our classification performance to other CNN reported in the literature for IRMA images. The IRMA error with ∗ was the best test value selected from the literature.

Methods	IRMA error
Proposed method	68.48
Parallel shrink CNN+Radon [9]	165.55*
Sequential shrink CNN+LBP [14]	168.05*
CNN+Radon [56]	210.35
CNNC+RBC [57]	224.13

Table 3: Comparison of classification performance of the proposed framework with other deep models and state-of-the-art on IRMA images.

Methods	IRMA error	AP	AR	F1 measure
VGG16 [18]	115.08	0.56	0.56	0.53
ResNet50 [19]	80.80	0.65	0.64	0.63
AttentionResNet56 [27]	76.83	0.65	0.66	0.64
AttentionResNet56+center loss [52]	73.85	0.68	0.66	0.66
Proposed method	68.48	0.67	0.67	0.66

number of classes that is 116 IRMA classes in this paper. As the F1 measure is more sensitive to data distribution, it is a suitable measure for classification problems on imbalanced datasets [10].

4.2.3. Classification Performance and Comparison. The performance of the proposed single-model framework for medical image classification is evaluated by the IRMA error and commonly used measures for image classification methods, which are detailed in Sections 4.2.1 and 4.2.2. Table 2 compares the IRMA error got by the proposed framework and several deep CNN-based methods reported in the literature [9, 14]. Table 2 shows that with the fast development of the deep CNN technique, much better classification accuracy (i.e., lower IRMA error score) can be gotten by employing a more powerful CNN model as a backbone network. In terms of the IRMA error, our proposed framework gets a much lower score than referenced deep CNN-based methods reported in the literature.

Considering the relative lag of the technology applied on IRMA dataset and the rapid development of the deep CNNs in computer vision area, Table 3 compares the classification accuracy measures on the IRMA dataset including IRMA error, AP, AR, and F1 measure of the proposed method with various state-of-the-art deep CNNs including VGG [18], ResNet [19], and AttentionResNet [27] that have achieved a very high recognition score on large image dataset challenges (such as ImageNet [58] and CoCo [59]). Table 3 shows that the proposed framework performs better in classifying IRMA images. The proposed framework and the compared deep CNNs are trained under the same condition, that is, using the same training dataset, same image argumentation strategy, same number of epochs, same learning rate, and so on.

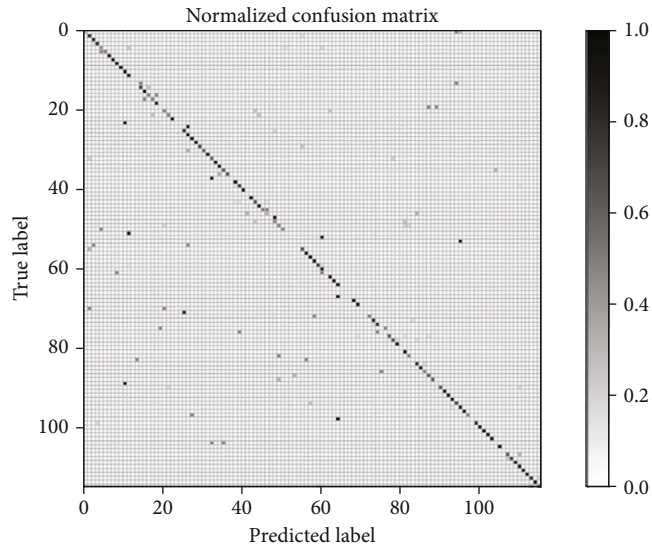


Figure 5: Confusion matrix of IRMA image classification with 116 classes using the proposed deep CNN.

Table 4: Comparison of our retrieval performance to other CNN reported in the literature for IRMA images. The proposed method used Cosine similarity measure to get the IRMA error in this table.

Methods	IRMA error
Proposed method	43.21
SVM+multiscale LBP [61]	146.55
Parallel shrink CNN+LBP, HOG, Radon [9]	165.55
Sequential shrink CNN+LBP [14]	168.05
TAUbiomed [60]	169.5
Diap [60]	178.93
CNN+Radon [56]	210.35
CNNC+RBC [57]	224.13
FEITIJS [60]	242.46
SuperPixel [57]	249.34
VPA [60]	261.16
SP-R [57]	311.8
MedGIFT [60]	317.53
SP-RBC [57]	356.57
IRMA [60]	359.29
MedGIFT [60]	420.91

For classification-based medical image retrieval, the retrieval performance depends entirely on the accuracy of classification, the higher classification accuracy means the better retrieval performance. As in Table 3, our proposed framework achieved the lowest IRMA error and the best F1 measure.

The confusion matrix is shown in Figure 5, where most classes can be classified rightly. There are 38.2% classes with accuracy better than 90%, 51.2% classes with accuracy better than 80%, and 59.1% classes with accuracy better than 70%.

Table 5: Comparison of retrieval performance of the proposed framework with other deep models and state-of-the-art on IRMA images. The IRMA error with ∗ was the best test value selected from the literature.

Methods	Vector dimension for similarity retrieval	IRMA error on Cosine similarity	mAP		
			Euclidean distance	Manhattan distance	Cosine similarity
Sequential shrink CNN+LBP [14]	8496	168.05*	—	—	—
Parallel shrink CNN+LBP, HOG, Radon [9]	8496 (LBP) 3528 (HOG) 1800 (Radon)	165.55*	—	—	—
SVM+multiscale LBP [61]	N subblocks × 4 × 4 × 1062 on SVM	146.55*	—	—	—
VGG16 [18]	512	65.53	0.53	0.53	0.53
AttentionResNet56 [27]	2048	49.66	0.45	0.45	0.62
AttentionResNet56+center loss [52]	2048	47.76	0.35	0.35	0.53
ResNet50 [19]	2048	47.53	0.72	0.72	0.73
Proposed method	32	43.21	0.84	0.85	0.86

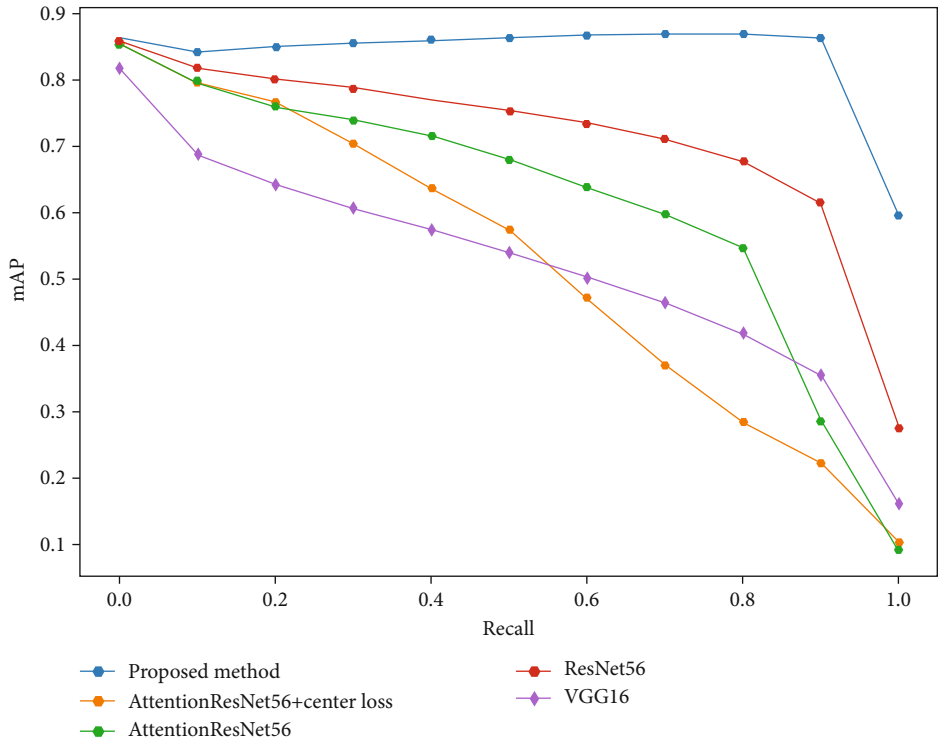


Figure 6: mAP vs. recall for medical retrieval on IRMA images.

4.3. Retrieval Performance

4.3.1. Retrieval Performance Measure. Precision and recall are two measures commonly used as retrieval performance evaluation measures [5].

$$P = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}},$$
$$R = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}}. \quad (6)$$

Besides precision and recall, mean average precision (MAP) is a very popular evaluation metric for algorithms that do search in medical image sets [5]. MAP combines precision and recall in one number. It is defined as the mean of average precision (AP) metric over all queries that can alleviate the bias during precision evaluation. The AP and mAP can be formulated as the following:

$$\text{AP}(q) = \frac{1}{N_R} \sum_{n=1}^{N_R} P_q(R_n), \quad (7)$$

where $P_q(R_n)$ is the precision value when the recall value is R_n and N_R indicates the top N_R-ranked relevant images for the query image q.

$$\text{mAP} = \frac{1}{|Q|} \sum_{q \in Q} \text{AP}(q), \quad (8)$$

where Q is the query image set and $|Q|$ is the number of the query image set.

4.3.2. Retrieval Performance and Comparison. In the proposed deep CNN framework, the feature vector for nearest-neighbor similarity searching of medical images is gotten from the last fully connected layer. For comparison, the proposed framework retrieval performance on the IRMA dataset is evaluated using both the IRMA error and the mean average precision (mAP). The calculation of the IRMA error in image retrieval follows the nearest-neighbor rule, that is, the query image's class label is determined by the most similar image returned in the retrieval process. Table 4 compares the retrieval performance achieved by the proposed framework with the other methods reported in the literature [9, 14, 56, 57, 60, 61] on the IRMA dataset with the IRMA error. The proposed deep CNN framework gets the lowest IRMA error in nearest-neighbor similarity retrieval. Table 5 compares the proposed framework with state-of-the-art deep CNNs on the IRMA error and mAP. For mAP, we test three usually used distance/similarity measures in image retrieval: Euclidean distance, Manhattan distance, and Cosine similarity, and the IRMA error is evaluated by using the best distance/-similarity measure: Cosine similarity. Table 5 shows that the proposed deep CNN framework gets the best mAP and the lowest IRMA error on these three distance/similarity measures and gets the highest score on Cosine similarity. In Table 5, we also list the vector length used for similarity retrieval. The feature vector for retrieval gotten from the proposed framework is just 32 dimensions that are much shorter than output vectors reported in literatures and state-of-the-

Table 6: Comparison of classification and retrieval performance of the proposed framework with other deep models and state-of-the-art on IRMA images with and without EMD components.

Methods	Classification				Retrieval (on Cosine similarity)			
	IRMA error		F1 measure		IRMA error		mAP	
	With EMD	Without EMD	With EMD	Without EMD	With EMD	Without EMD	With EMD	Without EMD
VGG16 [18]	115.08	98.29	0.53	0.52	65.53	57.18	0.53	0.54
ResNet50 [19]	80.8	92.62	0.63	0.55	47.53	62.2	0.73	0.70
AttentionResNet56 [27]	76.83	81.54	0.64	0.59	49.66	55.03	0.62	0.61
AttentionResNet56+center loss [52]	73.85	74.81	0.66	0.62	47.76	47.13	0.54	0.54
Proposed method	68.48	77.45	0.66	0.59	43.21	46.81	0.86	0.85

art deep CNNs, which illustrate the great potential of our method to implement large-scare medical retrieval. Suppose **a** and **b** are two feature vectors representing two medical images, the three distance/similarity measures are formulated as the following:

$$\begin{aligned}\text{Euclidean distance}(\mathbf{a},\mathbf{b}) &= \sqrt{\sum_{i=1}^{n}(a_i-b_i)^2},\\ \text{Manhattan distance}(\mathbf{a},\mathbf{b}) &= \sum_{i=1}^{n}|a_i-b_i|,\\ \text{Cosine similarity}(\mathbf{a},\mathbf{b}) &= \frac{\mathbf{a}\bullet\mathbf{b}}{\|\mathbf{a}\|\bullet\|\mathbf{b}\|}\end{aligned} \tag{9}$$

Figure 6 summarizes the retrieval performance of the proposed framework and state-of-the-art deep CNNs by the mAP-recall curve. And all these curves are calculated using the Cosine similarity measure.

4.4. Performance Comparison with and without EMD Components. To illustrate the effect of EMD components, Table 6 details the classification and retrieval measures between the proposed framework and the state-of-the-art deep CNNs with and without using EMD components. The results show that with EMD components, we can get higher performance in both classification and retrieval applications. With EMD components, deep CNNs can consistently achieve better classification and retrieval performance than without EMD components except for VGG16 on the IRMA error. This may be because the ResNet backbone is deeper than VGG16, so the CNNs based on the ResNet backbone can effectively handle more image information.

5. Conclusions

This paper has proposed a deep convolutional neural network for medical image retrieval task. By training deep CNN with input medical image and its multifrequency components (i.e., IMFs get from empirical mode decomposition (EMD)) in a supervised classification way, we have got a scheme that is very suitable for similarity-based medical image retrieval. Using an imbalanced IRMA medical image dataset, the proposed framework has surpassed existing algorithms with the highest classification accuracy and lowest retrieval error. The concise and distinguishable feature vector output from the proposed deep CNN has also shown great potential to handle large-scale medical image retrieval. We intend to further examine CBMIR on other medical datasets, different modalities, and 3D volumetric applications.

Data Availability

The IRMA2007 data used to support the findings of this study are available in a public repository at http://publications.rwth-aachen.de/search?ln=en&cc=Dataset&sc=1&p=IRMA&f=&action_search=Search.

Conflicts of Interest

The authors declared that they have no conflicts of interest in this work.

Acknowledgments

The authors would like to thank Bo Wen and Guihui Liu for their help for some experiments. This research is sponsored by the National Natural Science Foundation of China (Nos. 61561002 and 62062003), Natural Science Foundation of Ningxia (No. 2020AAC03213), Ningxia Medical Imaging Clinical Research Center Innovation Platform Construction Project (No. 2018DPG05006), "Image and Intelligent Information Processing Innovation Team" the State Ethnic Affairs Commission Innovation Team (Nos. PY1905 and PY1606), Ningxia Key Research and Development Project (special projects for talents) (2020BEB04022), North Minzu University Research Project of Talent Introduction under Grant 2020KYQD08, and General Research Project of North Minzu University (No. 2021XYZJK04).

References

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

[2] H. Zaidi, H. R. Vees, and M. Wissmeyer, "Molecular PET/CT imaging-guided radiation therapy treatment planning," *Academic Radiology*, vol. 16, no. 9, pp. 1108–1133, 2009.

[3] C. D. Marcus, V. Ladam-Marcus, C. Cucu, O. Bouché, L. Lucas, and C. Hoeffel, "Imaging techniques to evaluate the response to treatment in oncology: current standards and perspectives," *critical reviews in oncology hematology*, vol. 72, no. 3, pp. 217–238, 2009.

[4] C. E. Kahn, J. A. Carrino, M. J. Flynn, D. J. Peck, and S. C. Horii, "DICOM and radiology: past, present, and future," *Journal of the American College of Radiology*, vol. 4, no. 9, pp. 652–657, 2007.

[5] Z. Li, X. Zhang, H. Müller, and S. Zhang, "Large-scale retrieval for medical image analytics: a comprehensive review," *Medical Image Analysis*, vol. 43, pp. 66–84, 2018.

[6] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, no. 1, pp. 1–23, 2004.

[7] T. M. Lehmann, M. O. Güld, C. Thies et al., "Content-based image retrieval in medical applications," *Methods of Information in Medicine*, vol. 43, no. 4, pp. 354–361, 2004.

[8] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: current status and future directions," *Journal of digital imaging*, vol. 24, no. 2, pp. 208–222, 2011.

[9] A. Khatami, M. Babaie, A. Khosravi, H. R. Tizhoosh, and S. Nahavandi, "Parallel deep solutions for image retrieval from imbalanced medical imaging archives,," *Applied Soft Computing*, vol. 63, pp. 197–205, 2018.

[10] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 8–20, 2017.

[11] M. Srinivas, R. R. Naidu, C. S. Sastry, and C. K. Mohan, "Content based medical image retrieval using dictionary learning," *Neurocomputing*, vol. 168, pp. 880–895, 2015.

[12] L. Pan, Y. Qiang, J. Yuan, and L. Wu, "Rapid retrieval of lung nodule CT images based on hashing and pruning methods," *Biomed Research International*, vol. 2016, 10 pages, 2016.

[13] H.-C. Shin, K. Roberts, D. D.-F. Le Lu, J. Yao, and R. M. Summers, "Learning to read chest X-rays: recurrent neural cascade model for automated image annotation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2497–2506, Las Vegas, NV, USA, 2016.

[14] A. Khatami, M. Babaie, H. R. Tizhoosh, A. Khosravi, T. Nguyen, and S. Nahavandi, "A sequential search-space shrinking using CNN transfer learning and a Radon projection pool for medical image retrieval," *expert systems with applications*, vol. 100, pp. 224–233, 2018.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[16] A. Khatami, Y. Tai, A. Khosravi et al., "A deep learning-based model for tactile understanding on haptic data percutaneous needle treatment," in *International Conference on Neural Information Processing*, pp. 317–325, Springer, 2017.

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, The MIT Press, 2016.

[18] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, International Conference on Learning Representations, 2015.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[20] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[21] 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.

[22] A. A. A. Setio, F. Ciompi, G. Litjens et al., "Pulmonary nodule detection in CT images: false positive reduction using multiview convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[23] 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.

[24] H. C. 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.

[25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[26] 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.

[27] F. Wang, M. Jiang, C. Qian et al., *Residual attention network for image classification*6450–6458.

[28] H. K. Hoon, L. Haejun, and C. Duckjoo, "Medical image retrieval: past and present," *Healthcare Informatics Research*, vol. 18, no. 1, pp. 3–9, 2012.

[29] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1025–1039, 2013.

[30] M. Jiang, S. Zhang, H. Li, and D. N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 783–792, 2015.

[31] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," *Artificial Intelligence in Medicine in Europe*, pp. 126–135, 2009.

[32] S. Haas, R. Donner, A. Burner, M. Holzer, and G. Langs, "Superpixel-based interest points for effective bags of visual words medical image retrieval," in *MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support*, pp. 58–68, Springer, Berlin, Heidelberg, 2011.

[33] H. Müller, A. G. de Herrera, J. Kalpathy-Cramer, D. Demner-Fushman, S. K. Antani, and I. Eggel, *Overview of the ImageCLEF 2012 medical image retrieval and classiFIcation tasks*, Cross-Language Evaluation Forum, 2012.

[34] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial intelligence in medicine*, vol. 49, no. 2, pp. 117–125, 2010.

[35] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Directional binary wavelet patterns for biomedical image indexing

and retrieval," *Journal of Medical Systems*, vol. 36, no. 5, pp. 2865–2879, 2012.

[36] X. Xu, D. Lee, S. Antani, and L. R. Long, "A spine X-ray image retrieval system using partial shape matching," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 100–108, 2008.

[37] J. Liu, S. Zhang, W. Liu, X. Zhang, and D. N. Metaxas, "Scalable mammogram retrieval using anchor graph hashing," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 898–901, Beijing, China, 2014.

[38] M. Jiang, S. Zhang, J. Huang, L. Yang, and D. N. Metaxas, "Scalable histopathological image analysis via supervised hashing with multiple features," *Medical image analysis*, vol. 34, pp. 3–12, 2016.

[39] L. Song, X. Liu, L. Ma, C. Zhou, X. Zhao, and Y. Zhao, "Using HOG-LBP features and MMP learning to recognize imaging signs of lung lesions," in *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 1–4, Rome, Italy, 2012.

[40] K. Bunte, M. Biehl, M. F. Jonkman, and N. Petkov, "Learning effective color features for content based image retrieval in dermatology," *Pattern Recognition*, vol. 44, no. 9, pp. 1892–1902, 2011.

[41] M. M. Rahman, P. Bhattacharya, and B. C. Desai, "A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback," *International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 11, no. 1, pp. 58–69, 2007.

[42] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 294–297, New York, NY, USA, 2015.

[43] D. Semedo and J. Magalhães, *NovaSearch at ImageCLEFmed 2016 subfigure classification task*, pp. 386–398, 2016.

[44] J. Hofmanninger and G. Langs, "Mapping visual features to semantic profiles for retrieval in medical imaging," *Computer Vision and Pattern Recognition*, pp. 457–465, 2015.

[45] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the royal society a mathematical physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[46] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and P. Bunel, "Image analysis by bidimensional empirical mode decomposition," *Image and vision computing*, vol. 21, no. 12, pp. 1019–1026, 2003.

[47] K. Oberleithner, M. Sieber, C. N. Nayeri et al., "Three-dimensional coherent structures in a swirling jet undergoing vortex breakdown: stability analysis and empirical mode construction," *Journal of fluid mechanics*, vol. 679, pp. 383–414, 2011.

[48] S. Al-Baddai, P. Marti-Puig, E. Gallego-Jutglà et al., "A recognition–verification system for noisy faces based on an empirical mode decomposition with Green's functions," vol. 24, no. 5, pp. 3809–3827, 2020.

[49] H.-S. Oh and D. Kim, "Image decomposition by bidimensional ensemble patch transform," vol. 135, pp. 173–179, 2020.

[50] Z. Wu, N. E. Huang, and X. Chen, "The multi-dimensional ensemble empirical mode decomposition method," *Advances in Adaptive Data Analysis*, vol. 1, no. 3, pp. 339–372, 2011.

[51] Y. H. Wang, C. H. Yeh, H. W. Young, K. Hu, and M. T. Lo, "On the computational complexity of the empirical mode decomposition algorithm," *Physica a statistical mechanics and its applications*, vol. 400, pp. 159–167, 2014.

[52] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European Conference on Computer Vision*, pp. 499–515, Cham, 2016.

[53] T. M. Lehmann, M. O. Güld, T. Deselaers et al., "Automatic categorization of medical images for content-based retrieval and data mining," *Computerized Medical Imaging and Graphics*, vol. 29, no. 2-3, pp. 143–155, 2005.

[54] T. M. Lehmann, M. O. Güld, C. Thies et al., "IRMA–content-based image retrieval in medical applications," *Studies in health technology and informatics*, vol. 107, pp. 842–846, 2004.

[55] T. M. Lehmann, H. Schubert, D. Keysers, M. Kohnen, and B. B. Wein, "The IRMA code for unique classification of medical images," *Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation*, vol. 5033, pp. 440–451, 2003.

[56] A. Khatami, M. Babaie, A. Khosravi, H. R. Tizhoosh, S. M. Salaken, and S. Nahavandi, "A deep-structural medical image classification for a Radon-based image retrieval," in *Canadian Conference on Electrical and Computer Engineering*, pp. 1–4, Windsor, ON, Canada, 2017.

[57] X. Liu, H. R. Tizhoosh, and J. Kofman, "Generating binary tags for fast medical image retrieval based on convolutional nets and Radon transform," in *International Joint Conference on Neural Network*, pp. 2872–2878, Vancouver, BC, Canada, 2016.

[58] 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.

[59] T. Y. Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," in *European Conference on Computer Vision*, pp. 740–755, Cham, 2014.

[60] T. Tommasi, B. Caputo, P. Welter, M. O. Güld, and T. M. Deserno, "Overview of the CLEF 2009 medical image annotation track," in *Workshop of the Cross-Language Evaluation Forum for European Languages*, pp. 85–93, Springer, Berlin, Heidelberg, 2009.

[61] Z. Çamlica, H. R. Tizhoosh, and F. Khalvati, "Medical image classification via SVM using LBP features from saliency-based folded data," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pp. 128–132, Miami, FL, USA, 2015.

Hindawi
BioMed Research International
Volume 2020, Article ID 6619076, 10 pages
https://doi.org/10.1155/2020/6619076

Research Article

Spiculation Sign Recognition in a Pulmonary Nodule Based on Spiking Neural P Systems

Shi Qiu,[1] **Jingtao Sun**,[2] **Tao Zhou**,[3,4] **Guilong Gao**,[5] **Zhenan He**,[6] **and Ting Liang**[2,7]

[1]*Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China*
[2]*Department of Radiology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China*
[3]*School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China*
[4]*School of Science, Ningxia Medical University, Yinchuan 750004, China*
[5]*Key Laboratory of Ultra-Fast Photoelectric, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences Xi'an, 710119, China*
[6]*Shaanxi Institute of Medical Device Quality Supervision and Inspection, Xi'an 712046, China*
[7]*Department of Biomedical Engineering, the Key Laboratory of Biomedical Information Engineering of the Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710061, China*

Correspondence should be addressed to Tao Zhou; zhoutaonxmu@126.com and Ting Liang; liangting31500@126.com

Received 10 October 2020; Revised 4 December 2020; Accepted 11 December 2020; Published 24 December 2020

Academic Editor: Changming Sun

The spiculation sign is one of the main signs to distinguish benign and malignant pulmonary nodules. In order to effectively extract the image feature of a pulmonary nodule for the spiculation sign distinguishment, a new spiculation sign recognition model is proposed based on the doctors' diagnosis process of pulmonary nodules. A maximum density projection model is established to fuse the local three-dimensional information into the two-dimensional image. The complete boundary of a pulmonary nodule is extracted by the improved Snake model, which can take full advantage of the parallel calculation of the Spike Neural P Systems to build a new neural network structure. In this paper, our experiments show that the proposed algorithm can accurately extract the boundary of a pulmonary nodule and effectively improve the recognition rate of the spiculation sign.

1. Introduction

A pulmonary nodule is an early pattern of lung cancer. Malignant lesions might occur in some pulmonary nodules and even threaten patients' lives seriously [1]. The spiculation sign is the feature of a pulmonary nodule. It is a radial and unbranched strip shadow extending from the boundary of a pulmonary nodule to the surrounding pulmonary parenchyma [2]. Its detection may cost more time and energy of doctors.

The diagnosis of benign and malignant pulmonary nodules can be divided into imaging detection and "biopsy." The most accurate detection method is "biopsy," but it cannot predict the development trend of pulmonary nodules. Imaging analysis is still the mainstream detection method [3, 4]. It is also a main method to predict the development trend of benign and malignant pulmonary nodules from the perspective of imaging [5, 6]. "Biopsy" needs to sample the suspected lung lesions for detection. In the detection process, the instrument needs to be deep into the lung, which is easy to cause discomfort to patients. The suspected area for "biopsy" should be determined in advance. It needs to be analyzed by modeling from the perspective of imaging, so it is very important to start prepositioning from the perspective of imaging. "Biopsy" is the gold standard for judging benign and malignant pulmonary nodules. But the probability of malignant pulmonary nodules is far less than that of malignant. For this reason, not all pulmonary nodules must be biopsied. Main imaging features of pulmonary nodules include lobulation sign, spiculation sign, and cavity sign. It

is necessary to identify the signs in biopsy of high probability pulmonary nodules. According to the sign features of pulmonary nodules, a single model cannot realize recognition accurately and effectively. Therefore, we need to analyze the signs and establish the model one by one. "Biopsy" can only detect the current benign and malignant pulmonary nodules, but cannot predict the development of pulmonary nodules. But the imaging is different, it can compare the change rate of the same lesion point in different time periods, predict the development area of pulmonary nodules in the future, and further guide the diagnosis. Therefore, our research is significant.

The main signs of pulmonary nodules are lobulation sign, spiculation sign, cavity sign, and calcification. The spiculation sign has the highest deterioration rate, and it is difficult to distinguish the lobulated sign. Therefore, our research is based on the spiculation sign in this paper. Pulmonary nodules present a limited number of pixels in the image, and pulmonary nodules are volume data with three-dimensional structure. As a result, CT cannot accurately locate the signs of pulmonary nodules and make accurate judgment. Aiming at this problem, a density projection algorithm is proposed to integrate local 3D information into two-dimensional images for accurate diagnosis.

With the development of computer imaging technology, computer-aided diagnosis becomes possible for doctors and also has been successfully applied into the detection of pulmonary nodules: Qiu et al. [2] establishes a model to detect solitary pulmonary nodules. Gavrielides et al. [7] built a three-dimensional model to analyze the volume of pulmonary nodules. El-Baz et al. [8] judges the malignant degree of pulmonary nodules through analyzing morphological characteristics of pulmonary nodules. Brandman and Ko [9] establish a complete process including the detection of pulmonary nodules and the distinguishment and management of signs. Chen et al. [10] establish a neural network and a regression model to distinguish pulmonary nodules. Huang et al. [11] introduce the practical application of membrane calculation and achieves good results. Fan et al. [12] analyze the sign of pulmonary nodules from a mathematical and statistical perspective. Vinay et al. [13] construct an optimal classifier to distinguish the spiculation sign from a three-dimensional perspective. Dhara et al. [14] quantify the speculation sign on the basis of a three-dimensional model. Han et al. [15] focus on boundary characteristics to analyze the benign and malignant pulmonary nodules. Wang et al. [16] establish an image enhancement model to highlight pulmonary nodules. Choi and Choi [17] use a fixed threshold to segment pulmonary nodules. Rubin [18] sets seed points for local growth of pulmonary nodules. Shen et al. [19] establish a bidirectional coding system to improve the efficiency of the proposed algorithm. Qiang et al. [20] apply the active contour model for the segmentation of pulmonary nodules. Messay et al. [21] realize the segmentation of pulmonary nodules through analyzing the characteristics of CT pixel distribution from the linear regression perspective. Zhang et al. [22] analyze the spiking neural P systems based on the principle and puts forward a fast solution algorithm. Kumar et al. [23] classify pulmonary nodules by depth features. Bartholmai et al. [24] analyze the characteristics of pulmonary nodules with a computer. Firmino et al. [25] analyze the malignant degree of pulmonary nodules from the sign perspective. Dhara et al. [26] establish a gradient model to extract pulmonary nodules. Gonçalves et al. [27] establish the Hessian matrix to segment pulmonary nodules. Wang et al. [28] establish a data-driven model to focus on the pulmonary nodule area. Soliman et al. [29] establish the Adaptive Appearance-Guided Shape Model to simulate the distribution of pulmonary nodules. Froz et al. [30] classify pulmonary nodules with the support vector machine. Hoogi et al. [31] improve the level set algorithm for the pulmonary nodule segmentation. Wang et al. [32] apply the spiking neural P systems to realize the target tracking and path planning. Shakir et al. [33] establish a three-dimensional level set algorithm based on the two-dimensional segmentation. Qiu et al. [34] classify pulmonary nodules based on the geometric theory. Xie et al. [35] fuse multiple features to distinguish pulmonary nodules. Wang et al. [36] propose a set of complete data training algorithm to classify pulmonary nodules. Pang et al. [37] Automatic lung segmentation based on texture and deep features of hrct images with interstitial lung disease. Rong et al. [38] improve the spike neural P systems and improve the diagnosis accuracy. Cao et al. [39] used two-stage convolutional neural networks for nodule detection. Xu et al. [40] used multiresolution CT screening images to detect nodules.

Currently, the main problems of the computer-aided diagnosis of pulmonary nodules can be summarized as follows: (1) the two-dimensional and three-dimensional features of pulmonary nodules are difficult to be balanced during the modeling process. (2) The accurate segmentation of pulmonary nodules cannot be realized with gray values and without boundary features. (3) An effective distinguishing mechanism cannot be established after obtaining features of pulmonary nodules.

Therefore, in this paper, a spiculation sign recognition algorithm is proposed after studying the doctors' diagnosis process of pulmonary nodules. (1) A maximum intensity projection model is established to fuse the three-dimensional information into the two-dimensional image to reduce the missed rate of spiculation signs. (2) The accurate extraction of pulmonary nodules can be realized by the improved Snake model to strengthen the boundary effect. (3) A neural network framework based on the Spike Neural P Systems is constructed through focusing on boundary features of pulmonary nodules.

2. Algorithm

The spiculation sign recognition process of pulmonary nodules is simulated by the computer, as shown in Figure 1. (1) The maximum intensity projection algorithm is constructed to fully display the features of pulmonary nodules. (2) The boundary of pulmonary nodules is focused by the improved Snake algorithm. (3) The Spiking Neural P Systems is optimized to realize the sign recognition.

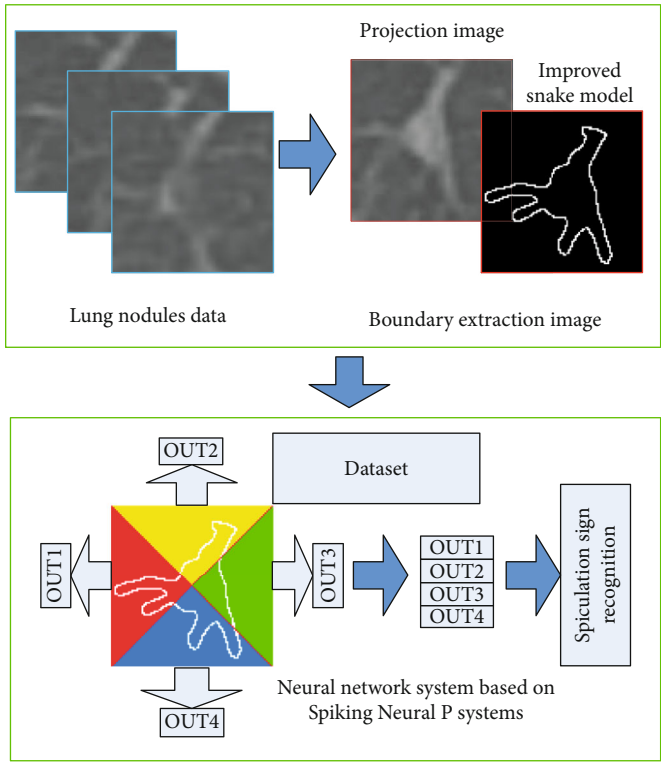


Figure 1: Flow chart of the spiculation sign recognition algorithm.

2.1. Projection Algorithm. The spiculation sign is the main feature to distinguish benign pulmonary nodules from malignant ones. It is defined as a radial and unbranched stripe shadow extending from the boundary of a pulmonary nodule to the surrounding pulmonary parenchyma. According to the local highlight of a pulmonary nodule, its section structure is extracted layer by layer to construct a model from the perspective of local three-dimensional information.

The maximum gray value along the ray direction of continuous multiframes is used by MIP as the gray value of the corresponding point on the projection image [41],

$$\mathrm{MIP}(x, y) = \max\left(I_0(x, y) \cdots I_N(x, y)\right), \tag{1}$$

where $\mathrm{MIP}(x, y)$ is the gray value at the point (x, y) on the MIP image. N is the number of projection layers. $I_k(x, y)$ is the gray value at the point (x, y) on the k-th image in the original CT sequence images. MIP images contain local three-dimensional features, which can restore the local three-dimensional information of pulmonary nodules, as shown in Figure 2.

2.2. The Segmentation Algorithm of Pulmonary Nodule. A pulmonary nodule is displayed in the highlighted area and occupies a limited number of pixels in CT images.

A benign pulmonary nodule has features of small area, high luminance, and smooth boundary; however, a malignant pulmonary nodule has features of large area, high luminance, and blurred boundary. Complete segmentation is the premise of the pulmonary nodule distinguishment.

2.2.1. The Snake Model. The Snake model algorithm can perform the target segmentation from the perspective of internal energy and external energy [42]. It has the following advantages: image data, initial estimation, target contour, and knowledge-based constraints are unified in one process. It can automatically converge to the state of minimum energy after proper initialization. Minimizing the energy from coarse to fine in scale space can greatly expand the capture area and reduce the complexity. Meanwhile, the Snake model algorithm also has its disadvantages: It is sensitive to the initial position, and Snake needs to be placed near the image features depending on other mechanisms. It may converge to the local extremum or even diverge because of the nonconvexity of the Snake model. Dong et al. [43] introduce the deep learning theory to constrain the Snake algorithm to segment targets. Rajinikanth et al. [44] achieve the three-dimensional target segmentation based on the Snake algorithm and the Otsu algorithm. Ma et al. [45] fuse the local phase position, and the Snake algorithm alleviates the problem of convergence to the local extremum.

When the Snake model achieves the balance of internal energy and external energy, the optimal segmentation effect is obtained. The energy functional is defined as:

$$\begin{cases} E = \int_0^1 \{E_{\mathrm{in}}[C(s)] + E_{\mathrm{out}}[C(s)]\}\, ds, \\ C(s) = (x(s), y(s)) s \in [0, 1], \end{cases} \tag{2}$$

where $C(s)$ is a contour curve and $E_{\mathrm{in}}[C(s)]$ is an internal energy function. $E_{\mathrm{in}}[C(s)]$ is only related to the curve itself,

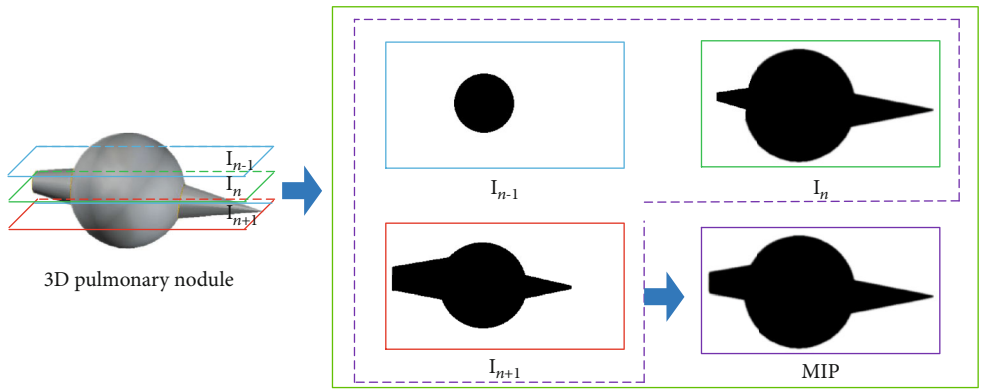


Figure 2: MIP effect image.

so that the curve keeps continuity and smoothness during deformation. $E_{\text{out}}[C(s)]$ is an external energy function, and $E_{\text{out}}[C(s)]$ is only related to the image itself, which can drive the curve to move towards the target boundary continuously.

$$E_{\text{in}}[C(s)] = \frac{1}{2}\left[\alpha\left|C'(s)\right|^2 + \beta\left|C''(s)\right|^2\right], \tag{3}$$

where α is the elastic energy weight coefficient and β is the rigid energy weight. The minimization of variational principle $C(s)$ should satisfy the Euler equation:

$$\alpha C'' - \beta C'''' - \nabla E_{\text{out}} = 0. \tag{4}$$

The GVF model [46] introduces the gradient vector flow $V(x, y) = (u(x, y), v(x, y))$ to replace the external force of the Snake model, then the energy functional of the external force field is

$$\varepsilon_{\text{GVF}} = \min\left\{\iint\left\{w\left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right) + |\nabla f|^2|V - \nabla f|^2\right\}dxdy\right\}, \tag{5}$$

where w is the weight coefficient to control the smoothness of the external force field. $f(x, y)$ is an image boundary mapping function. When the curve is far from the target contour, the first term plays a major role. On the contrary, the second term plays a major role in expanding the search scope. By solving

$$\begin{cases} w\nabla^2 u - (u - f_x)\left(f_x^2 + f_y^2\right) = 0, \\ w\nabla^2 v - (v - f_x)\left(f_x^2 + f_y^2\right) = 0, \end{cases} \tag{6}$$

where the GVF field is obtained, where∇^2 is a Laplacian operator. The Laplace operator produces an isotropic smoothing effect on the external force field and cannot protect the boundary.

2.2.2. The Improved Model. As the traditional Snake algorithm is easy to converge to the local extreme and cannot protect boundary, we have analyzed the Laplace operator: The Laplace operator can be decomposed into normal and tangent components, and the normal direction component can promote the contour line to converge to the deep concave part. Thus, $w|J_v P|$ term is added to make the curve converge to the small deep concave boundary. The improved function is as follows:

$$\varepsilon = \min\left\{\iint\left\{m(x, y)|\nabla V|^2 + h(x, y)\left(w|J_v P|^2 + |V - \nabla f|^2\right)\right\}dxdy\right\}, \tag{7}$$

where w, $g(x, y)$, and $h(x, y)$ are weighting functions and J_v is the Jacobian matrix of external force field. In order to enhance the corresponding boundary, we construct

$$P = \begin{bmatrix} -\dfrac{I_{xy}}{\sqrt{I_{xx}^2 + I_{xy}^2}}, \dfrac{I_{xx}}{\sqrt{I_{xx}^2 + I_{xy}^2}} \\ -\dfrac{I_{yy}}{\sqrt{I_{yx}^2 + I_{yy}^2}}, \dfrac{I_{yx}}{\sqrt{I_{yx}^2 + I_{yy}^2}} \end{bmatrix}, \tag{8}$$

to increase the accuracy of corner positioning.

In Eq. (7), $|\nabla V|^2$ has a strong smoothing effect on the boundary. To reduce boundary weakening, $|\nabla V|^2$ is replaced by

$$G = \left(1 + |\nabla V|^2\right)^{q(|\nabla f|)/2} q(|\nabla f|) = 1 + \frac{1}{1 + |\nabla f|}. \tag{9}$$

In the smoothing area, $|\nabla f| \to 0$, $q(|\nabla f|) \to 2$, the external force field has an isotropic diffusion effect. At the boundary, $|\nabla f| \to \infty$, $q(|\nabla f|) \to 1$, $G \to |\nabla V|$, the external force field only diffuses along the boundary direction to prevent boundary leakage and improve the antinoise performance. The energy function is

$$\varepsilon = \min\left\{\iint\left\{m(x, y)G + h(x, y)\left(w|J_v P|^2 + |V - \nabla f|^2\right)\right\}dxdy\right\}, \tag{10}$$

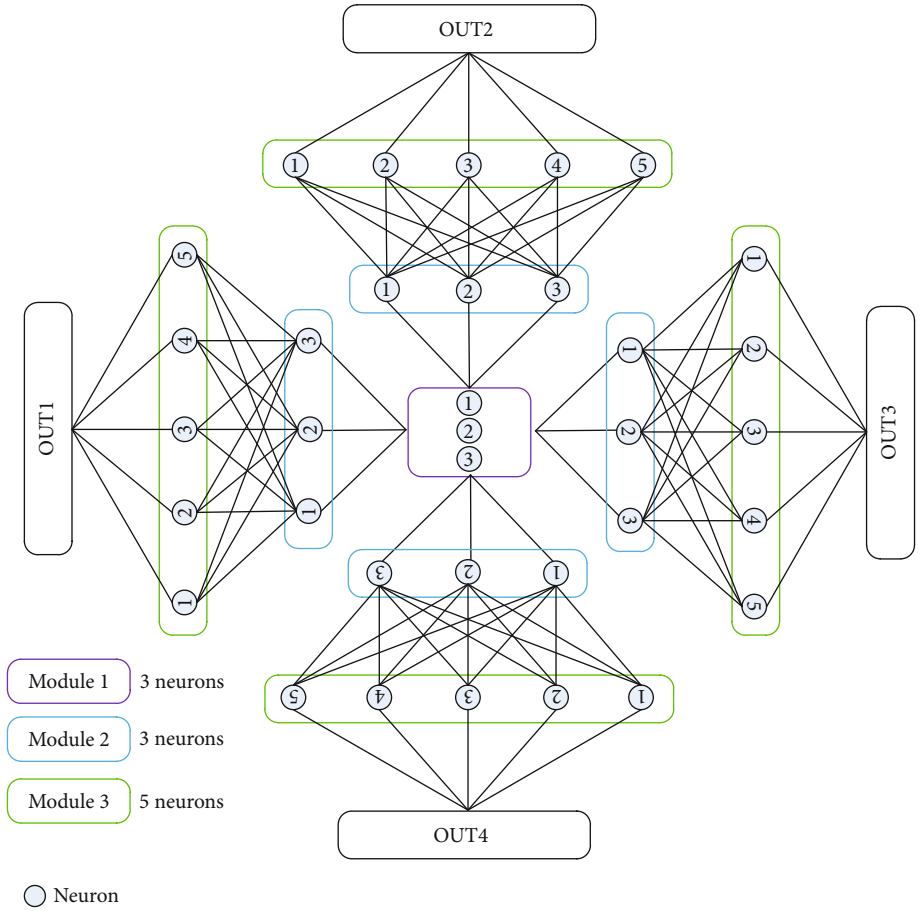


Figure 3: Network framework.

which corresponds to the Euler equations $Z(u) = 0$ and $Z(v) = 0$.

The iterative formula of numerical solution $l(u_{i,j}^{n+1})$, $l(v_{i,j}^{n+1})$ is

$$\begin{aligned} l\left(a_{i,j}^{n+1}\right) = &\left(1 - \Delta t \cdot h(|\nabla f|)a_{i,j}^{n} + \Delta t \cdot g(|\nabla f|)\right. \\ &\cdot \left[\left(1 + |\nabla a|^2\right)^{q-2/2} a_{tt} + \left((q-2)|\nabla a|^2\left(1 + |\nabla a|^2\right)^{q-4/2}\right.\right. \\ &\left.\left. + \left(1 + |\nabla a|^2\right)^{q-2/2} + 1\right) u_{nn}\right] + \Delta t \\ &\cdot \left[w^2 h(|\nabla f|)\left(P_{11}^2 a_{xx} + P_{12}^2 a_{yy} + 2P_{11}P_{12}\right) + c_a\right], \end{aligned} \tag{11}$$

where $c_u = h(|\nabla f|)f_x$, $c_v = h(|\nabla f|)f_y$. Δt is the iteration step. $(u_{i,j}^{n}, v_{i,j}^{n})$ represents the field forces at coordinates (i, j) with n iterations. Pulmonary nodules are extracted layer by layer to obtain complete pulmonary nodules.

2.3. Neural Network System Based on Spiking Neural P Systems. The SN P systems are a parallel computing model derived from organisms [47]. Wang et al. [48] introduce the fuzzy set theory on the basis of SN P systems, which solves the problem of fault diagnosis to a certain extent. The topological structure of SN P systems is composed of a directed graph. Each neuron in the system is represented by a node, and the synapse between two adjacent neurons in the system is represented by edge. It is similar to the topological structure of the artificial neural network. There are abundant theoretical and applied researches in ANN, so the learning rules in ANN can be introduced into SN P systems.

The neural network based on SN P systems can be defined as

$$\begin{aligned} &\mathrm{B} = (O, \sigma_1, \cdots, \sigma_m, \mathrm{syn}, y, \mathrm{in}, \mathrm{out}), \\ &\sigma_i = (n_i, R_i), \end{aligned} \tag{12}$$

where O represents a set of pulses; σ_m represents the m-th neuron in System B; $n_i \geq 0$ represents the number of original pulses; R_i represents a set of all rules in neuron σ_i; The form of excitation rule is $E/a^c \to a$, $c \geq 1$; The rule of oblivion is $a^s \to \lambda$, $s \geq 1$; y represents the learning function of system; in and out represent the input and output neurons of the system, respectively. Define the rule as $E/a^c \to \alpha k(i, Q_j)$, $k \geq 1$, $c \geq 1$, $1 \leq j \leq |R_i|$. When the rule is called, all neurons in σ_i and Q_j establish the connection state.

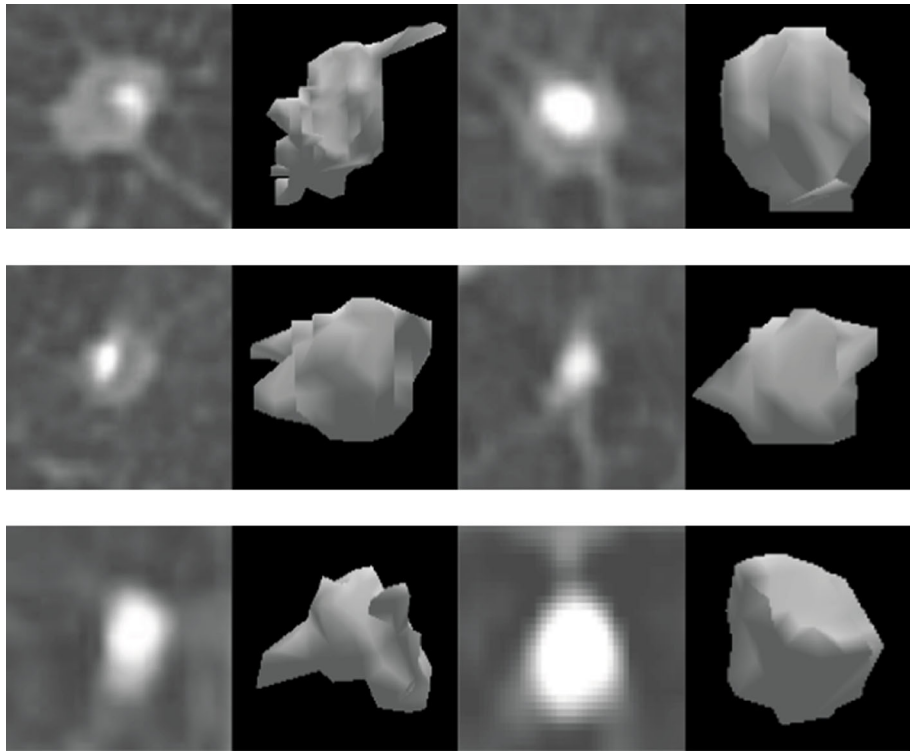


Figure 4: Experimental data.

Syn represents a synapse between σ_i and σ_j; $w_{ij}(t)$ represents the weight of synapse (i, j). $T = \{w_{ij}(t) \mid t = 1,2,3 \cdots\}$ represents a set of weights on synapses(i, j) at different times.

According to the state of time t and $w_{ij}(t)$, the synaptic weight set $w_{ij}(t+1)$ at time $t+1$ can be obtained by y; pr (σ_i) and po(σ_i) represent the label set of presynaptic neurons and postsynaptic neurons of σ_i, respectively.

If σ_i contains b pulses and $a^b \in L(E)$, $E/a^c \to \alpha k(i, Q)$ is used. If the rules in the system are excited, c pulses will be consumed. Then, the next step will be performed according to the value of α:

(i) For $\alpha = +$, if $1 \le |Q - \mathrm{pr}(\sigma_i)| \le k$, σ_i selects all neuron tags in $Q - \mathrm{pr}(\sigma_i)$ to create synapses. If $|Q - \mathrm{pr}(\sigma_i)| > k$, σ_i randomly selects k neuron tags in $Q - \mathrm{pr}(\sigma_i)$ to create synapses. If $Q - \mathrm{pr}(\sigma_i) = \varnothing$ or $\mathrm{pr}(\sigma_i) = \varnothing$, C pulses are consumed but synapses are not established. In this case, the principle of synaptic creation rules is similar to that of standard rules of oblivion

(ii) For $\alpha = -$, if $|\mathrm{pr}(\sigma_i)| \le k$, all synapses are deleted in $\mathrm{pr}(\sigma_i)$. If $|\mathrm{pr}(\sigma_i)| > k$, k neurons are selected in $\mathrm{pr}(\sigma_i)$ and the synaptic connection with each selected neuron is deleted

(iii) For $\alpha = \mp$, synapses are created at the time t and deleted at the time $t+1$. Conversely, for $\alpha = \pm$, synapses are deleted at the time t and created at the time $t+1$. In this case, the use of rules is similar to that of $\alpha = +$ and $\alpha = -$. From time t to time $t+1$, σ_i is always in an open state, but σ_i uses other rules at time $t+2$

If σ_i has k pulses and $a^k \in L(E)$, $k \ge c$, the excited rule $E/a^c \to a^p$; d is used. When this rule is used, σ_i will delete c pulses. At the same time, p pulses are sent to all neurons connected to σ_i after d time intervals. When the excited rule is used to the d-th time intervals, σ_i is in a closed state. Rules and processing pulses can only be used by σ_i when the execution conditions are met. If σ_i uses the excitation rule $E/a^c \to a^p$ at t-th step, σ_i at t-th, $t+1-$th, $\cdots t+d-1-$th step is not executed. After $t+d$ steps, σ_i is in the excited state.

If a neuron has s pulses, the rule of oblivion $E'/a^s \to \lambda$, $s \ge 1$ is used. When this rule is used, σ_i will consume s pulses. No new pulse will be produced.

The state of System P at a certain time is expressed as $C_r = <k_1/t_1, \cdots, k_m/t_m>, 1 \le i \le m$, where k_i represents the number of pulses stored in neuron σ_i in this state; t_i represents the time taken for σ_i to be reactivated. At the beginning of System P calculation, all neurons meet the excitation rule conditions. By rules, the state of the system is transferred. $C_1 \Rightarrow C_2$ means that the system is transferred from state C_1 to state C_2. When all neurons in the system have been activated, the termination state means that there are no rules in the neurons that can be activated again. If a system is able to calculate till the termination state, then the calculation is regarded as the one that can be terminated.

According to the state of time t and $w_{ij}(t)$, the synaptic weight set $w_{ij}(t+1)$ is obtained at time $t+1$. Theoretically, if there is a transfer of M_t^{t+1} from time t to time $t+1$ in the system, and the set of weights on the synapse is $w_{ij}(t)$.

Then, under the transfer of ${M_t}^{t+1}$, the set function of synaptic weights at time $t+1$ is $w_{ij}(t+1) = y({M_t}^{t+1}, w_{ij}(t))$.

2.4. Network Connection. Based on the above analysis, the boundary extraction image is combined with the neural network system of SN P systems. The parallelism of SN P systems and the flexibility of neural networks are taken full advantage.

(1) Mark the boundary of a pulmonary nodule as 1, which is regarded as a pulse signal. Nonboundary areas are marked as 0

(2) Normalize the boundary image size of a pulmonary nodule to 5×7

(3) The neurons are divided into three parts, as shown in Figure 3. The flow direction of a pulse signal is from Module 1 to Module 2 and then to Module 3. Three neurons of Module 1 establish the neural connection of Module 2 through the defined SN P systems rules, and the weights of all synapses are 1. There is only one excitation rule for the neurons of Module 2 and Module 3, that is, if the neuron contains pulses, the neuron is excited until the number of pulses in the neuron changes to 0, and the calculation is terminated. Module 2 has four layers, and each layer contains three neurons. Module 3 has four layers, and each layer contains five neurons. The neurons of Module 2 and Module 3 are connected by synapses

3. Experiment and Result Analysis

All the experimental data are from the database of the International Early Lung Cancer Action Project and the American Association of Lung Imaging Databases, as shown in Figure 4. 514 pulmonary nodules with spiculation signs and 501 pulmonary nodules without spiculation signs are labeled by two professional doctors as the detection basis. The ratio of training data and test data is 1 : 1.

3.1. Image Segmentation. The area overlap measure (AOM) is used to evaluate the segmentation effect.

$$\mathrm{AOM}(A, B) = \frac{S(A \cap B)}{S(A \cup B)} \times 100\%. \quad (13)$$

AOM is the overlap degree of area. A is the standard image. B is the segmentation result image. $S(.)$ represents the pixel number of the corresponding area. The larger the AOM value, the better the segmentation effect.

Different algorithms are used to segment common pulmonary nodules and pulmonary nodules with speculation sign, as shown in Table 1. It illustrates that the segmentation effect for common pulmonary nodules is better than that for pulmonary nodules with spiculation sign. That is because common pulmonary nodules have high gray value and high density, and pulmonary nodules with spiculation sign have high gray values including small protrusions. The fixed threshold [17] algorithm achieves segmentation of pulmonary nodules by selecting threshold artificially, and the result is good. But the threshold setting is manual. The gradient model [26] algorithm focuses on the boundary to extract pulmonary nodules. AAGSM [29] used an initial shape of pulmonary nodules to constrain segmentation of pulmonary nodules. LS [31] algorithm establishes the iterative model to achieve segmentation of pulmonary nodules. The Snake [38] algorithm establishes internal force and external force balance mechanism to extract pulmonary nodules. The Esnake [40] algorithm introduces the Otsu algorithm to improve Snake and achieves good results. On the basis of the Snake algorithm, our algorithm protects boundary information and suppresses falling into local minimum. It has a strong segmentation effect for common pulmonary nodules and pulmonary nodules with speculation sign.

Table 1: The effect comparison of algorithms.

Algorithm	AOM %	
	Common	Spiculation
Fixed threshold [17]	94	92
Gradient model [26]	85	76
AAGSM [29]	86	79
LS [31]	89	83
Snake [38]	91	84
Esnake [40]	93	87
Ours	94	90

3.2. The Speculation Discrimination Effect. The ROC curve is introduced to measure the effect of all algorithms. The recognition results of the original pulmonary nodule image by different algorithms are shown in Figure 5(a), and the recognition results of different algorithms in MIP pulmonary nodule images are shown in Figure 5(b). It can be seen that the MIP algorithm can better reflect the boundary features of pulmonary nodules and improve the distinguishing effect of spiculation sign. The fractal model (FM) [34] uses the fractal operator to calculate the fractal degree of pulmonary nodules to judge the signs of pulmonary nodules. The nerve network model (NNM) [10] algorithm introduces a learning mechanism to realize feature learning, which requires a large number of samples to train parameters. 3DM [13] establishes a three-dimensional pulmonary nodule model and analyzes the pulmonary nodule signs from a spatial perspective, which can realize the identification of pulmonary nodule signs, but the algorithm has high complexity. The feature fusion model (FFM) [35] extracts the gray value and boundary information of pulmonary nodules to realize the identification of pulmonary nodules. Our algorithm fuses the pulmonary nodule information from three locations and proposes a time series analysis algorithm, which achieves good results. The proposed algorithm in this paper focuses on the boundary of the pulmonary nodule spiculation sign and integrates the SN P systems into the neural network. It gives full play to the advantages of the SN P systems and has a better effect.

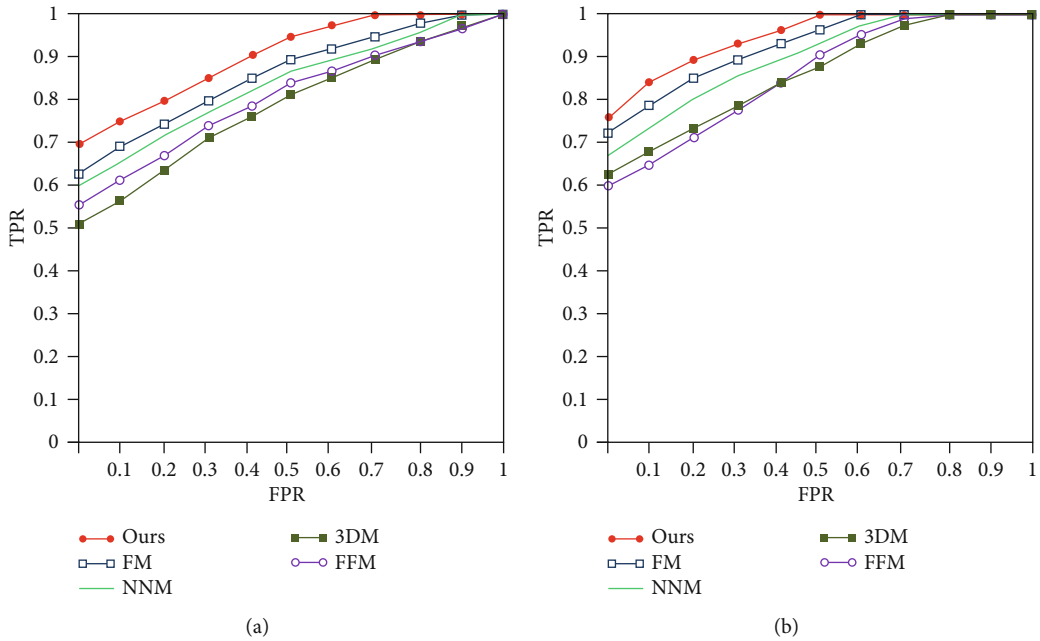


Figure 5: ROC curve. (a) The recognition results of original pulmonary nodule image by different algorithms. (b) The recognition results of original pulmonary nodule MIP image by different algorithms.

4. Conclusion

In view of the recognition of pulmonary nodules with computer, a complete recognition system of speculation sign of pulmonary nodules is proposed from the doctors' perspective. The MIP algorithm is proposed to restore the three-dimensional local structure of pulmonary nodules. The improved Snake algorithm can extract the boundary information of pulmonary nodules completely. The neural network system based on SNP systems can help doctors to make accurate diagnosis with computer-aided. On the basis of existing datasets, we will expand the amount of data. By labeling the dataset, it is of great significance to integrate the imaging features and pathological features of different time periods into the model and carry out the research on the prediction of benign and malignant development trend of pulmonary nodules.

Data Availability

All used data is within the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by Postdoctoral Science Foundation of China under Grant No. 2020M682144 and 62062003, the Natural Science Foundation of China under Grant 61561040, North Minzu University Research Project of Talent Introduction under Grant 2020KYQD08, Shanxi National Science Foundation under Grant 2020JQ-518, and the Open Project Program of the State Key Lab of CAD&CG under Grant No. A2026, Ningxia Key Research and Development Project under Grant 2020BEB04022.

References

[1] M. Winkels and T. S. Cohen, "Pulmonary nodule detection in CT scans with equivariant CNNs," *Medical Image Analysis*, vol. 55, pp. 15–26, 2019.

[2] S. Qiu, Q. Guo, D. Zhou, Y. Jin, T. Zhou, and Z. He, "Isolated pulmonary nodules characteristics detection based on CT images," *IEEE Access*, vol. 7, pp. 165597–165606, 2019.

[3] B. Qiu, J. Furst, A. Rasin, R. Tchoua, and D. Raicu, "Learning latent spiculated features for lung nodule characterization," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1254–1257, Montreal, QC, Canada, 2020.

[4] L. Liu, Q. Dou, H. Chen, J. Qin, and A. Heng, "Multi-task deep model with margin ranking loss for lung nodule analysis," *IEEE Transactions on Medical Imaging*, vol. 39, no. 3, pp. 718–728, 2020.

[5] W. Wang and G. Chakraborty, "Deep learning for automatic identification of nodule morphology features and prediction of lung cancer," in *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)*, Morioka, Japan, 2019.

[6] G. Zhang, Z. Yang, L. Gong et al., "An appraisal of nodule diagnosis for lung cancer in CT images," *Journal of Medical Systems*, vol. 43, no. 7, p. 181, 2019.

[7] M. A. Gavrielides, L. M. Kinnard, K. J. Myers et al., "A resource for the assessment of lung nodule size estimation methods: database of thoracic CT scans of an anthropomorphic phantom," *Optics Express*, vol. 18, no. 14, pp. 15244–15255, 2010.

[8] A. El-Baz, M. Nitzken, F. Khalifa et al., "3D shape analysis for early diagnosis of malignant lung nodules," in *Biennial International Conference on Information Processing in Medical Imaging*, pp. 772–783, Springer, Berlin, Heidelberg, 2011.

[9] S. Brandman and J. P. Ko, "Pulmonary nodule detection, characterization, and management with multidetector computed tomography," *Journal of Thoracic Imaging*, vol. 26, no. 2, pp. 90–105, 2011.

[10] H. Chen, J. Zhang, Y. Xu, B. Chen, and K. Zhang, "Performance comparison of artificial neural network and logistic regression model for differentiating lung nodules on CT scans," *Expert Systems with Applications*, vol. 39, no. 13, pp. 11503–11509, 2012.

[11] X. Huang, J. Shan, and V. Vaidya, "Lung nodule detection in CT using 3D convolutional neural networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 379–383, Melbourne, VIC, Australia, 2017.

[12] L. Fan, S. Y. Liu, Q. C. Li, H. Yu, and X. S. Xiao, "Multidetector CT features of pulmonary focal ground-glass opacity: differences between benign and malignant," *The British Journal of Radiology*, vol. 85, no. 1015, pp. 897–904, 2012.

[13] K. Vinay, A. Rao, and G. Hemantha Kumar, "Prediction of lung nodule characteristic rating using best classifier model," *International Journal of Computer Applications*, vol. 56, no. 18, 2012.

[14] A. K. Dhara, S. Mukhopadhyay, N. Alam, and N. Khandelwal, "Measurement of spiculation index in 3D for solitary pulmonary nodules in volumetric lung CT images," in *Medical Imaging 2013: Computer-Aided Diagnosis, International Society for Optics and Photonics*, Lake Buena Vista (Orlando Area), FL, United States, 2013.

[15] F. Han, G. Zhang, H. Wang et al., "A texture feature analysis for diagnosis of pulmonary nodules using LIDC-IDRI database," in *2013 IEEE International Conference on Medical Imaging Physics and Engineering*, pp. 14–18, Shenyang, China, 2013.

[16] Y. X. J. Wang, J. S. Gong, K. Suzuki, and S. K. Morcos, "Evidence based imaging strategies for solitary pulmonary nodule," *Journal of Thoracic Disease*, vol. 6, no. 7, p. 872, 2014.

[17] W. J. Choi and T. S. Choi, "Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 37–54, 2014.

[18] G. D. Rubin, "Lung nodule and cancer detection in CT screening," *Journal of Thoracic Imaging*, vol. 30, no. 2, p. 130, 2015.

[19] S. Shen, A. A. Bui, J. Cong, and W. Hsu, "An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy," *Computers in Biology and Medicine*, vol. 57, pp. 139–149, 2015.

[20] Y. Qiang, X. Zhang, G. Ji, and J. Zhao, "Automated lung nodule segmentation using an active contour model based on PET/CT images," *Journal of Computational and Theoretical Nanoscience*, vol. 12, no. 8, pp. 1972–1976, 2015.

[21] T. Messay, R. C. Hardie, and T. R. Tuinstra, "Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset," *Medical Image Analysis*, vol. 22, no. 1, pp. 48–62, 2015.

[22] G. Zhang, H. Rong, F. Neri, and M. J. Pérez-Jiménez, "An optimization spiking neural P system for approximately solving combinatorial optimization problems," *International Journal of Neural Systems*, vol. 24, no. 5, article 1440006, 2014.

[23] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in *2015 12th Conference on Computer and Robot Vision*, pp. 133–138, Halifax, NS, Canada, 2015.

[24] B. J. Bartholmai, C. W. Koo, G. B. Johnson et al., "Pulmonary nodule characterization, including computer analysis and quantitative features," *Journal of Thoracic Imaging*, vol. 30, no. 2, pp. 139–156, 2015.

[25] M. Firmino, G. Angelo, H. Morais, M. R. Dantas, and R. Valentim, "Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy," *Biomedical Engineering*, vol. 15, no. 1, pp. 1–17, 2016.

[26] A. K. Dhara, S. Mukhopadhyay, P. Saha, M. Garg, and N. Khandelwal, "Differential geometry-based techniques for characterization of boundary roughness of pulmonary nodules in CT images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 3, pp. 337–349, 2016.

[27] L. Gonçalves, J. Novo, and A. Campilho, "Hessian based approaches for 3D lung nodule segmentation," *Expert Systems with Applications*, vol. 61, pp. 1–15, 2016.

[28] S. Wang, M. Zhou, Z. Liu et al., "Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation," *Medical Image Analysis*, vol. 40, pp. 172–183, 2017.

[29] A. Soliman, F. Khalifa, A. Elnakib et al., "Accurate lungs segmentation on CT chest images by adaptive appearance-guided shape modeling," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 263–276, 2017.

[30] B. R. Froz, A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, "Lung nodule classification using artificial crawlers, directional texture and support vector machine," *Expert Systems with Applications*, vol. 69, pp. 176–188, 2017.

[31] A. Hoogi, C. F. Beaulieu, G. M. Cunha et al., "Adaptive local window for level set segmentation of CT and MRI liver lesions," *Medical Image Analysis*, vol. 37, pp. 46–55, 2017.

[32] X. Wang, G. Zhang, F. Neri et al., "Design and implementation of membrane controllers for trajectory tracking of nonholonomic wheeled mobile robots," *Integrated Computer-Aided Engineering*, vol. 23, no. 1, pp. 15–30, 2016.

[33] H. Shakir, T. M. R. Khan, and H. Rasheed, "3-D segmentation of lung nodules using hybrid level sets," *Computers in Biology and Medicine*, vol. 96, pp. 214–226, 2018.

[34] S. Qiu, J. Li, M. Cong, C. Wu, Y. Qin, and T. Liang, "Detection of solitary pulmonary nodules based on brain-computer interface," *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 4930972, 10 pages, 2020.

[35] Y. Xie, J. Zhang, Y. Xia, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT," *Information Fusion*, vol. 42, pp. 102–110, 2018.

[36] Z. Wang, J. Xin, P. Sun, Z. Lin, Y. Yao, and X. Gao, "Improved lung nodule diagnosis accuracy using lung CT images with uncertain class," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 197–209, 2018.

[37] T. Pang, S. Guo, X. Zhang, and L. Zhao, "Automatic lung segmentation based on texture and deep features of hrct images with interstitial lung disease," *BioMed Research International*, vol. 2019, Article ID 2045432, 8 pages, 2019.

[38] H. Rong, K. Yi, G. Zhang, J. Dong, P. Paul, and Z. Huang, "Automatic implementation of fuzzy reasoning spiking neural P systems for diagnosing faults in complex power systems," *Complexity*, vol. 2019, Article ID 2635714, 16 pages, 2019.

[39] H. Cao, H. Liu, E. Song et al., "A two-stage convolutional neural networks for lung nodule detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 1–2015, 2020.

[40] X. Xu, C. Wang, J. Guo et al., "DeepLN: a framework for automatic lung nodule detection using multi-resolution CT screening images," *Knowledge-Based Systems*, vol. 189, article 105128, 2020.

[41] S. Hassan and J. Dobson, "4D imaging experience for treatment of lung SBRT," *Clinical Oncology*, vol. 30, no. 6, p. e59, 2018.

[42] T. Mitchel, Q. Xuan, J. S. Kim, G. Chirikjian, and C. Li, *A novel 3-D full body model of snake locomotion in complex 3-D terrain*, APS, 2018, X06-010.

[43] S. Dong, G. Luo, G. Sun, K. Wang, and H. Zhang, "A left ventricular segmentation method on 3D echocardiography using deep learning and snake," in *2016 Computing in Cardiology Conference (CinC)*, pp. 473–476, Vancouver, BC, Canada, 2016.

[44] V. Rajinikanth, N. S. Madhavaraja, S. C. Satapathy, and S. L. Fernandes, "Otsu's multi-thresholding and active contour snake model to segment dermoscopy images," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 8, pp. 1837–1840, 2017.

[45] L. Ma, H. Kiyomatsu, K. Nakagawa, J. Wang, E. Kobayashi, and I. Sakuma, "Accurate vessel segmentation in ultrasound images using a local-phase-based snake," *Biomedical Signal Processing and Control*, vol. 43, pp. 236–243, 2018.

[46] F. Zhang, X. Zhang, K. Cao, and R. Li, "Contour extraction of gait recognition based on improved GVF Snake model," *Computers & Electrical Engineering*, vol. 38, no. 4, pp. 882–890, 2012.

[47] T. Song, L. Pan, T. Wu, P. Zheng, M. L. D. Wong, and A. Rodriguez-Paton, "Spiking neural P systems with learning functions," *IEEE Transactions on Nanobioscience*, vol. 18, no. 2, pp. 176–190, 2019.

[48] T. Wang, G. Zhang, J. Zhao, Z. He, J. Wang, and M. J. Perez-Jimenez, "Fault diagnosis of electric power systems based on fuzzy reasoning spiking neural P systems," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1182–1194, 2014.

Hindawi
BioMed Research International
Volume 2020, Article ID 6636321, 9 pages
https://doi.org/10.1155/2020/6636321

Research Article

NSCR-Based DenseNet for Lung Tumor Recognition Using Chest CT Image

Zhou Tao,[1] **Huo Bingqiang**,[1] **Lu Huiling**,[2] **Yang Zaoli**,[3] **and Shi Hongbin**[4]

[1]*School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China*
[2]*School of Science, Ningxia Medical University, Yinchuan 750004, China*
[3]*College of Economics and Management, Beijing University of Technology, Beijing 100124, China*
[4]*Urinary Surgery, General Hospital of Ningxia Medical University, Yinchuan 750004, China*

Correspondence should be addressed to Lu Huiling; lu_huiling@163.com

Received 24 October 2020; Revised 15 November 2020; Accepted 4 December 2020; Published 16 December 2020

Academic Editor: Lin Gu

Nonnegative sparse representation has become a popular methodology in medical analysis and diagnosis in recent years. In order to resolve network degradation, higher dimensionality in feature extraction, data redundancy, and other issues faced when medical images parameters are trained using convolutional neural networks. Lung tumors in chest CT image based on nonnegative, sparse, and collaborative representation classification of DenseNet (DenseNet-NSCR) are proposed by this paper: firstly, initialization parameters of pretrained DenseNet model using transfer learning; secondly, training DenseNet using CT images to extract feature vectors for the full connectivity layer; thirdly, a nonnegative, sparse, and collaborative representation (NSCR) is used to represent the feature vector and solve the coding coefficient matrix; fourthly, the residual similarity is used for classification. The experimental results show that the DenseNet-NSCR classification is better than the other models, and the various evaluation indexes such as specificity and sensitivity are also high, and the method has better robustness and generalization ability through comparison experiment using AlexNet, GoogleNet, and DenseNet-201 models.

1. Introduction

Chest CT images offer the advantages of easy access, cost-effectiveness, and low radiation dosage needed, making it the most common screening procedure in daily clinical practice. Diagnostic testing of multiple diseases of the chest from CT images by radiologists can provide useful references for the diagnosis and treatment of lung diseases. Lung cancer [1] is one of the malignant tumors with a high rate of morbidity and mortality, posing a serious threat to human health. Early diagnosis and early detection are crucial to the treatment of lung cancer. Computer-assisted diagnostic technology (CAD) [2] has been widely used in the diagnosis and treatment of various diseases, especially lung cancer detection, which is one of the most common applications of CAD technology. The introduction of computer-aided diagnosis technology has an important and positive effect on the early detection and diagnosis of lung cancer, so it has great prospects for development in the field of assisting doctors in diagnosing and treating lung cancer.

In recent years, deep learning [3] had achieved great success in the field of image processing due to its excellent learning capabilities. Deep learning, exemplified by DenseNet [4], has been increasingly applied in the field of medical imaging; good results have been achieved in clinically assisted classification, identification, detection, and segmentation for benign and malignant tumors, brain functions, cardiovascular diseases, and other major diseases. Residual neural networks (ResNet) [5, 6] reduce feature redundancy and reuse existing features by sharing parameter shortcut connections and preserving intermediate features. Khened et al. [7] proposed a fully convolutional multiscale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. In Alzheimer's disease diagnosis, hippocampus analysis by combination of 3-D DenseNet and shapes are putted forward by Cui and Liu [8]. Tong et al. [9] proposed a channel-attention-based DenseNet network for remote

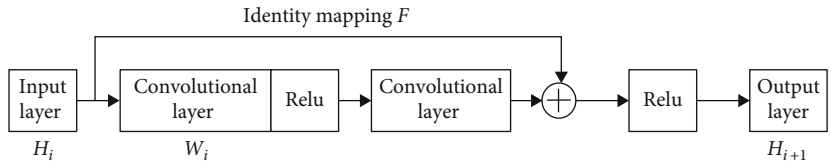


Figure 1: Residual block.

sensing image scene classification. However, the trade-off is that it is difficult to rediscover the underlying features using high-level information; DenseNet effectively leverages high-level information to rediscover new features at the bottom layer, enhancing feature transmission across the network and enabling enhanced feature reuse, effectively reducing the number of parameters. Chen et al. [10] proposed a new DenseNet and ResNet-based dual asymmetric feature learning network, DualCheXNet, which uses two homogeneous DCNNs to learn each other supplemented with more accurate features for multilabeled thoracic disease classification, which is relatively robust; Dai et al. [11] proposed the improved lung nodule classification identification algorithm based on DenseNet; the model is based on DenseNet and uses intermediate dense projection method to obtain three-dimensional information about pulmonary nodules and train the network using Focal Loss to enable the network to focus on learning the difficult resolved lung nodules, with good experimental results; Zhu and Qin [12] proposed an improved U-Net convolutional neural network lung nodule detection algorithm using convolutional and pooling operation to retrieve high-level features, enable high-speed flow of feature information between input and output layers through DenseNet, and generate in combination with expansion convolution. Multiscale features improve the utilization of low-level features of pulmonary nodules. Li et al. [13] used a DenseNet for computer-aided diagnosis of lung cancer, which uses a patch-based, multiresolution DenseNet to extract features and classify them using four different integration methods.

Sparse representation (SR) and collaborative representation (CR) have become a popular methodology in pattern classification and computer vision for computer-aided diagnosis (CAD) and tumor recognition in recent years [14]. These methods first encode the query sample as a linear combination of the given training samples and then assign the query sample to the corresponding class with the minimal distance or approximation error. One seminal work in this category is the sparse representation- (SR-) based classifier (SRC). Sparse representation models often contain two stages: sparse coding and dictionary learning. Li et al. [15] propose a nonnegative dictionary-based sparse representation and classification scheme for ear recognition. The nonnegative dictionary includes the Gabor feature dictionary extracted from the ear images and nonnegative occlusion dictionary learned from the identity occlusion dictionary. A test sample with occlusion can be sparsely represented over the Gabor feature dictionary and the occlusion dictionary. The sparse coding coefficients are noted with nonnegativity and much more sparsity, and the nonnegative dictionary has shown increasing discrimination ability. Mi et al. [16] propose a robust supervised sparse representation (RSSR) model, which uses a two-phase robust representation to compute a sparse coding vector. Huber loss is employed as the fidelity term in the linear representation, which improves the competitiveness of correct class in the first phase. Then, training samples with weak competitiveness are removed by supervised way. In the second phase, the competitiveness of correct class is further boosted by Huber loss. Zhang et al. [17] propose a nonlinear nonnegative sparse representation model: NNK-KSVD. In the sparse coding stage, a nonlinear update rule is proposed to obtain the sparse matrix. In the dictionary learning stage, the proposed model extends the kernel KSVD by embedding the nonnegative sparse coding. The proposed nonnegative kernel sparse representation model was evaluated on several public image datasets for the task of classification. Fuzzy discriminative sparse representation (FDSR) is proposed by Ghasemi et al. [18]; the proposed fuzzy terms increase the interclass representation difference and the intraclass representation similarity. Also, an adaptive fuzzy dictionary learning approach is used to learn dictionary atoms. A robust sparse representation for medical image classification is proposed based on the adaptive type-2 fuzzy learning (T2-FDL) system by Ghasemi et al. [19]. In the proposed method, sparse coding and dictionary learning processes are executed iteratively until a near-optimal dictionary is obtained. Moradi and Mahdavi-Amiri [20] propose a sparse representation-based method for segmentation and classification of lesion images. The main idea of our framework is based on a kernel sparse representation, which produces discriminative sparse codes to represent features in a high-dimensional feature space. Our novel formulation for discriminative kernel sparse coding jointly learns a kernel-based dictionary and a linear classifier. We also present an adaptive K-SVD algorithm for kernel dictionary and classifier learning. In order to solve the semantic gap problem between low-level features and high-level image semantic, which will largely degrade the classification performance, Zhang et al. [21] propose a multiscale nonnegative sparse coding-based medical image classification algorithm.

This paper presents methods for classification of for benign and malignant lung tumors based on non-negative, sparse, and collaborative representation classification of DenseNet (DenseNet-NSCR). First, CT modal medical images were collected and preprocessed. The dataset is then trained in a DenseNet to construct a DenseNet model to extract the full connection layer feature vector. It was

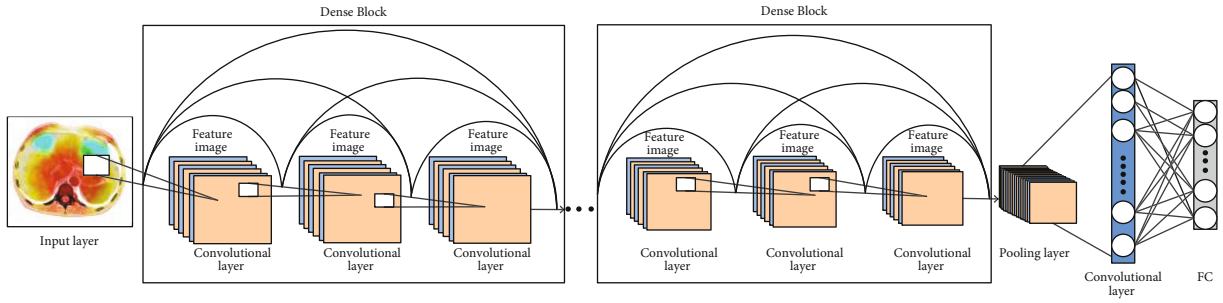


Figure 2: Residual neural network, resNet-101 diagram structure.

concluded with the results of lung tumor classification in the NSCR classifier, compared by a total of nine models, AlexNet +SVM, AlexNet+SRC, AlexNet+NSCR, GoogleNet+SVM, GoogleNet+SRC, GoogleNet+NSCR, DenseNet+SVM, DenseNet+SRC, and DenseNet+NSCR. The DenseNet+NSCR model outperforms the other models with better robustness and generalization capabilities.

2. Basic Principle

2.1. The Basic Structure of DenseNet. DenseNet is typically composed of multiple Dense Blocks and transition layer structures overlap to form a multilayer neural network. Its internal Dense Block structure uses the residual neural network's shortcut connection [5] structure. The deep residual neural network is usually composed of multiple residual block structures overlapping each other. Neighboring convolutional layers are connected by a shortcut to form a residual block. The residual block structure is shown in Figure 1, where H_i is input, H_{i+1} is output, W_i is weight, and F denotes the identity mapping. The residual block mapping is represented in Figure 1 as

$$H_{i+1} = \mathrm{Re}\, lu(H_i + F(H_i, W_i)). \quad (1)$$

The DenseNet structure uses dense connections in model building as shown in Figure 1, where the current network layer is connected to each subsequent layer. The feature map within each Dense Block is of the same size, and the features learned by the DenseNet are reused within the network. The dense connections between the DenseNet layers facilitate the flow of information throughout the network. Its nonlinear function is shown in Eq. (2) where x_i denotes the output of layer l. $[x_0\ x_1\ x_2 \cdots\ x_{l-1}]$ indicates the collocation of feature maps from the input layer to the l-1 layer. H_i denotes the nonlinear function which is a combined operation containing the batch normalization (BN) layer, the Relu layer, and the convolutional layer. As a result, the training of the deep network becomes more efficient and the performance of the model is improved as shown in Figure 2.

DenseNet has fewer parameters for network training compared to ResNet networks. Also, the use of dense connections alleviates the overfitting problem for models with small datasets. For the transition layer, it mainly connects two Dense Blocks, which contain a 1×1 convolution and 2×2 average pooling to reduce the feature map size. If the Dense Block of the previous layer outputs m feature maps, the transition layer can generate θ feature maps, where $0 \le \theta \le 1$ is called the compression factor; when $\theta = 1$, the feature map remains unchanged; when $\theta < 1$, the transition layer can further compress the model. In this paper, DenseNet's $k = 32$ and $\theta = 0.5$ are used.

$$x_l = H_i([x_0 x_1 x_2 \cdots x_{l-1}]). \quad (2)$$

DenseNet has the following features: firstly, DenseNet effectively alleviates the gradient vanishing problem caused by an overly deep network. DenseNet effectively strengthens feature forward transmission by acquiring the loss function of all preceding layers for each layer, so that deeper networks can be trained; secondly, compared to ResNet, which uses summation to transmission features, DenseNet uses inception's concatenation channel merge, which merges all previous layer outputs together as the current input, thus significantly improving feature transmission efficiency; thirdly, residual neural networks reduce feature redundancy and reuse existing features by sharing parameters across layers and preserving intermediate features, with the disadvantage that it is difficult to rediscover the underlying features using high-level information; DenseNet effectively leverages high-level information to rediscover new features at the bottom layer, enhancing feature transmission across the network and enabling and enhancing feature reuse; fourthly, DenseNet effectively reduces the number of parameters compared to ResNet which has a larger number of parameters.

2.2. NSCR Algorithm. There are many redundant or irrelevant features in high-dimensional data, thus facing the curse of dimensionality. On one hand, high computational time and space are required; on the other hand, problems such as overfitting occur in classification tasks. Therefore, data dimension reduction is a challenging task in machine learning. The sparse representation of high-dimensional feature data is one of the recent research hotspots in the field of machine learning, and SRC/CRC/NRC's [16, 22]core idea is that test samples that are represented approximately by linear combinations of training samples from all classes, and then, the test samples are assigned to the corresponding class with minimum distance or approximate error. However, the coding coefficients in

Table 1: The NSCR-based classifier.

The NSCR based classifier	
1	Input: training sample matrix $X = [X_1, \cdots, X_k]$ and query sample y
2	Normalize each column of matrix X and query sample y to the unit L_2 norm
3	The encoding vector of y on X is solved by the NSCR model
4	Calculate the coefficient matrix: $\hat{c} = \arg \min_c \left\|y' = Xc\right\| + \alpha\|c\|_2^2 + \beta c \, s.t. c \geq 0$
5	Calculate residual similarity: $r_k = \|y - X_k\hat{c}_k\|_2$
6	Output label category: $\text{label}(y) = \arg \min \{r_k\}$

the sparse representation classifier SRC/CRC will be negative, which in practice makes the problem of the corresponding weights of positive and negative coding coefficients offset, which affects the sample classification accuracy to some extent. Nonnegative representation of classification NRC coding coefficients for classification ideas are restricted to nonnegative, and non-negative representation enhances the representation of homogeneous samples' capabilities while limiting the representation of heterogeneous samples. Despite the success of the three classifiers, SRC/CRC/NRC, in the image recognition task, they have their corresponding localization. When using the entire training image to reconstruct the test image y, on the one hand, both SRC and CRC are generated in the coding coefficient vector deviation. The reason is that from a generative point of view, it is not physically feasible to reconstruct real-world images from training images with complex negative (minus) and positive (plus) coefficients. NRC constrains the coding coefficients to be nonnegative, but due to the lack of proper regularization, NRC classification is not flexible enough to deal with real-world problems. NSCR [23] combines the advantages of sparse, collaborative, and nonnegative representations to be physically more robust and generalizable than previous sparse, collaborative, and nonnegative representations.

The NSCR classifier can be reconstructed as a bivariate problem bounded by a linear equation and can be solved under the alternate direction [24] method (ADMM) of the multiplicative subframe. Each subproblem can be solved efficiently in closed form and can converge to a global optimum. Extensive experiments of NSCR on various visual classification datasets have verified the effectiveness of NSCR classifier, and NSCR classification is better than advanced classification algorithms such as SVM and SRC. Based on the above discussion, the NSCR algorithm for a given test sample and training sample matrix X, X consists of several classes of samples, where $X = [X_1, \cdots, X'_k] \in R^{D \times N}$; its algorithmic idea is shown in Table 1:

2.3. Evaluation Metrics. In this paper, the evaluation metrics [25] include accuracy, sensitivity, specificity, F-score value, and Matthews correlation coefficient (MCC), which are described as follows:

Accuracy, sensitivity, and specificity were calculated by true positive (TP), false positive (FP), true negative (TN), and false negative (FN). TP indicates a benign tumor was predicted correctly, FP indicates a malignant tumor was predicted incorrectly, TN indicates a malignant image was predicted correctly, and FN indicates that benign tumors were predicted incorrectly. They are calculated by the following formulae. The calculation formula is as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{3}$$

$$\text{Specificity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{5}$$

The F value is a summed average of the percentages of completeness and accuracy. It is used as a trade-off between accuracy and recall. The calculation formula is as follows:

$$F = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}. \tag{6}$$

MCC is a more comprehensive evaluation metric that reflects the reliability of the algorithm. When the number of categories is different, the value of the measure considered to be balanced ranges from -1 to +1. The MCC takes the value of 1 when the prediction error is 0 for both FP and FN, which means that the classification is completely correct; when the prediction error is 0 for both TP and TN, the MCC takes the value of -1, which means that the classification is completely wrong. It is calculated as follows:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \tag{7}$$

3. NSRC-Based DenseNet Model

Target the network degradation problem when training CT modal medical images using convolutional neural networks, high dimensionality, and data redundancy during feature extraction and other problems. This paper combines the DenseNet-based feature extraction method and the classification recognition method based on nonnegative, sparse, and collaborative representation, in the proposal of a DenseNet-based nonnegative, sparse, and collaborative representation (DenseNet-NSCR) classification of benign and malignant lung tumors. The steps of the calculation as a whole are divided into image

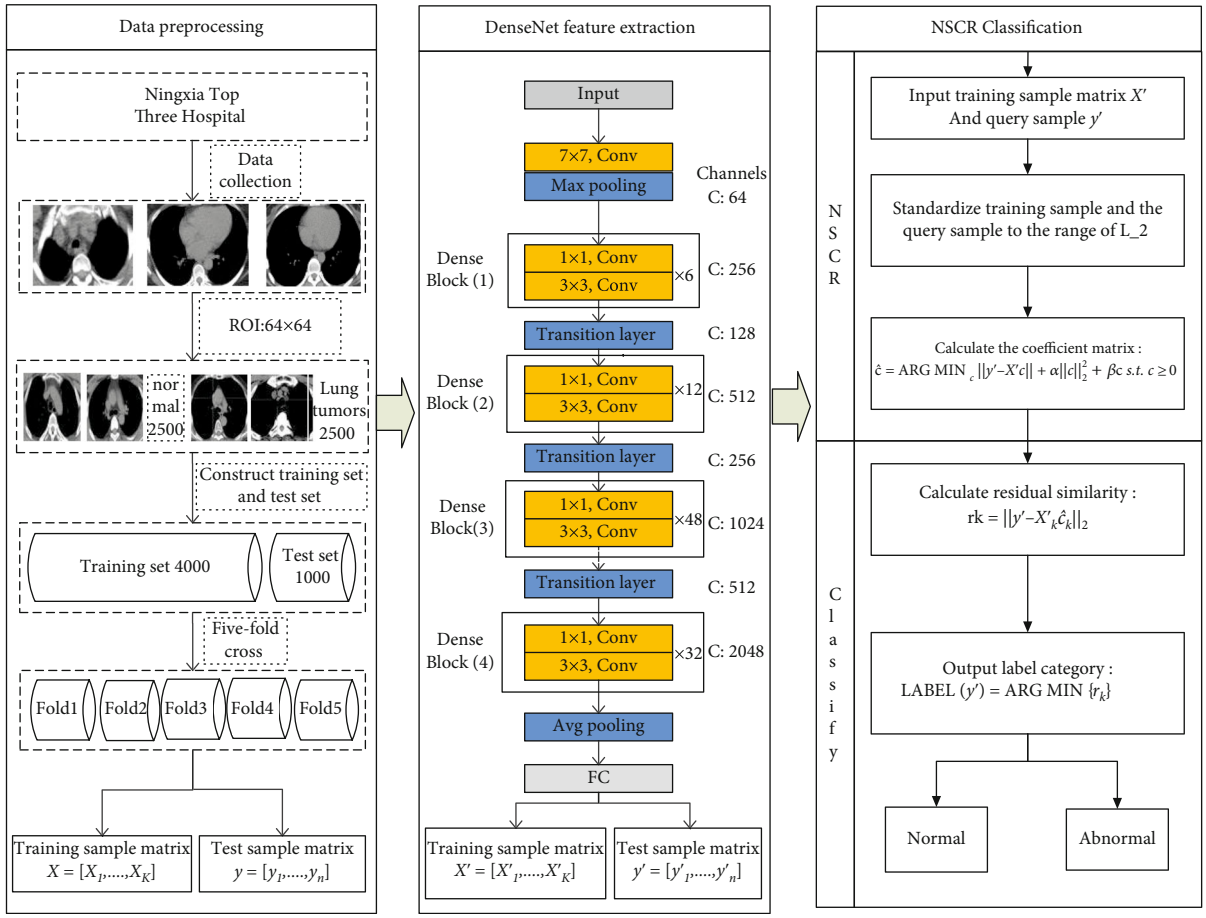


Figure 3: Flowchart of NRC DenseNet-based algorithm.

Table 2: Accuracy under the regularization parameters.

		α				
		0.01	0.05	0.1	0.5	1
β	0.01	99.37	99.33	99.32	99.15	99.01
	0.05	99.47	99.36	99.33	99.17	99.01
	0.1	99.48	99.40	99.31	99.20	99.00
	0.5	99.03	99.15	99.24	99.12	98.93
	1	98.33	97.06	97.05	99.01	98.79

preprocessing, DenseNet feature extraction, and NSCR classification.

3.1. Image Preprocessing. (1) *Data collection*: 5000 raw images of lung CT models were collected from a hospital in Ningxia of China between 2014 and 2016. The number of both benign and malignant lung tumors was 2500 cases [26].

(2) *Data preprocessing*: the original images of the lung CT models have numbered accordingly and recolored into grayscale images. Based on the clinical markers, the focal areas were intercepted from the full-grayscale images and normalized to the same size as the ROI images, e.g., 64 px × 64 px, to obtain CT modal samples, which were divided into benign samples and lung malignancy samples. The benign sample and the lung malignancy sample were each 2500 samples. The two types of targets were divided into a test set and a training set of 4000 and 1000 cases, respectively, according to a certain ratio, and constructed with its corresponding binary labels, where the benign label is 1 and the lung malignancy label is 2.

3.2. Dense Neural Network-DenseNet. (1) *Transfer learning*: the dense neural network, DenseNet-201 model is first pretrained on a large natural image dataset, ImageNet, with the parameters from the pretrained network as the initialization parameters in the network where the growth rate of the DenseNet is $k = 32$ while the compression rate of the transition layer is $\theta = 0.5$.

(2) *DenseNet partial feature extraction*: the datasets and labels are input into the pretrained dense neural network, DenseNet-201, respectively, and a single-module network based on the DenseNet model, which is CT-DenseNet, is constructed; DenseNet is trained to extract the feature vectors of training samples and test samples at the full-joint layer.

3.3. NSCR Classification Identification. Extract the feature vectors of training sample matrices and test sample matrices at the full connection layer of a DenseNet, input the feature matrix as an NSCR classifier, standardize all

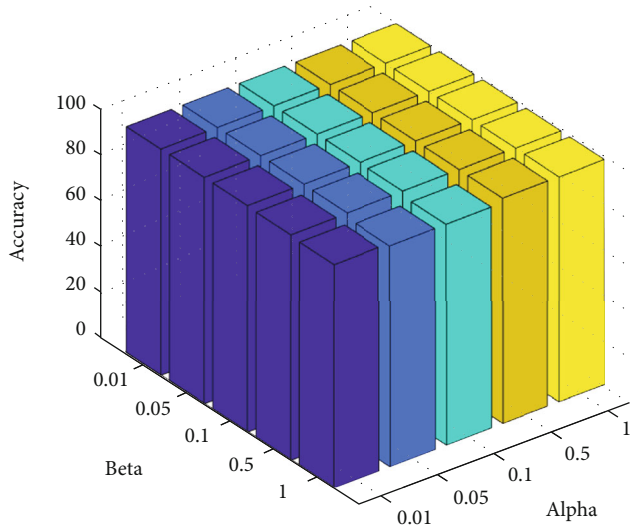


Figure 4: Regularization parameters α and β.

Table 3: Comparison of accuracy and training time results for different models.

Dataset / Model	CT Accuracy (/%)	CT Training time (/s)
AlexNet+SVM	97.50	224.58
AlexNet+SRC	96.32	604.20
AlexNet+NSCR	98.52	334.39
GoogleNet+SVM	97.90	662.44
GoogleNet+SRC	98.02	1081.21
GoogleNet+NSCR	98.82	773.97
DenseNet-201+SVM	98.26	3182.66
DenseNet-201+SRC	98.32	3962.32
DenseNet-201+NSCR	99.10	3234.63

training sample matrices, and test sample matrices to the L_2 paradigm and solve the coefficient matrix, which in turn is used to find the reconstruction error for each category. Finally, the final classification identification is completed based on the similarity of the reconstruction residuals as follows:

(1) For the training sample $X' = [X'_1, \cdots, X'_K]$, $X_i \in \mathrm{CT}$, and for the testing sample $y' = [y'_1 \cdots, y'_n]$, $y_i \in \mathrm{CT}$. After dense neural network, DenseNet-201 feature extraction, a training sample matrix $X' = [X'_1, \cdots, X'_K]$, and a test sample matrix of the feature space $y' = [y'_1 \cdots, y'_n]$ are obtained

(2) Standardize each column X' of the matrix and the query sample y' to the range of L_2

(3) The nonnegative sparse and collaborative representation processing of y' with the training sample X' in feature space is done to obtain the matrix of representation coefficient $\hat{c}$:

$$\hat{c} = \arg \min_c \left\| y' = X'c \right\| + \alpha \|c\|_2^2 + \beta c \; s.t. c \geq 0 \tag{8}$$

(4) Classify the residual similarity of nonnegative, sparse, and collaborative representation of test samples by training samples:

$$r_k = \left\| y' - X'_k \hat{c}_k \right\|_2 \tag{9}$$

(5) Output the label categories corresponding to the residual results:

$$\mathrm{Label}\left(y'\right) = \arg \min \{r_k\} \tag{10}$$

The NSCR-based DenseNet model DenseNet-NSCR is shown in Figure 3.

4. Algorithm Simulation Experiments

4.1. Experimental Environment. Software environment: Windows10 operating system, MatlabR2019a;

Hardware environment: Intel(R)Core(TM)i5-7200U CPU @2.50GHz 2.70GHz, 4.0GB memory, 500GB hard disk.

4.2. Results and Analysis of Experiments. To ensure the reliability of the data, the five-fold crossover method was used in this experiment. All samples were divided into five equal parts. Each copy contains equal proportions of the number of samples in different categories; 4 sets of data were used as training samples at a time, while the remaining 1 sample was used as a test sample, and each result was averaged to get the final result. That is, the number of training samples each session is 4000, the number of test samples is 1000, and the average of five experiments is taken. Experiments are conducted on three different network models, AlexNet, GoogleNet, and DenseNet, and three classification algorithms: the SVM, the SRC, and the NSCR. The results of the experimental comparison of the two combined models are as follows:

4.2.1. Experiment 1: NSCR Regularization Parameter Optimization. The regularization parameters α and β affect the performance of the NSCR classifier to achieve the optimal performance of the NSCR classifier. In this experiment, the regularization parameters α and β were selected as 0.01, 0.05, 0.1, 0.5, and 1, respectively, with CT medical images as the dataset, the dataset was randomly divided 7 : 3, and a five-fold crossover experiment was performed. The optimal regularization parameters α and β were found with classification accuracy as the index.

As shown in Table 2, the selection of different regularization parameters α and β affects the performance of the NSCR classifier. When $\alpha = 0.01$ and $\beta = 0.1$, the NSCR classification accuracy is 99.48% and the performance of the NSCR

Table 4: Comparison of CT results for different network models and classification algorithms.

Network model	Classification algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-score (%)	MCC (%)
AlexNet	SVM	97.50	97.40	97.60	97.50	95.00
	SRC	96.32	96.36	96.28	96.32	92.64
	NSCR	98.52	98.44	98.60	98.52	97.04
GoogleNet	SVM	97.90	97.76	98.04	97.90	95.80
	SRC	98.02	99.20	96.84	98.04	96.07
	NSCR	98.82	99.20	98.44	98.82	97.64
DenseNet-201	SVM	98.26	98.32	98.20	98.26	96.52
	SRC	98.32	99.60	97.04	98.34	96.67
	NSCR	99.10	99.60	98.60	99.10	98.20

classifier is optimal. To better indicate the effect of the selection of the regularization parameters α and β on the classification results, the three indicators were plotted in a three-dimensional histogram, as shown in Figure 4.

4.2.2. Experiment 2: Comparison of Accuracy of Different Models and Time. This experiment focuses on the recognition accuracy and training time of nine algorithms (AlexNet+SVM, AlexNet+SRC, AlexNet+NSCR, GoogleNet +SVM, GoogleNet+SRC, GoogleNet+NSCR, DenseNet +SVM, DenseNet+SRC, and DenseNet+NSCR) for training and recognition on CT sampling space and probes the effects of different network models, different classification algorithms, and different sampling spaces on the recognition rate and training time of dense neural networks, as shown in Table 3.

In the first case, different network models are used with the same classification algorithm. In experiment 1, there are three sets of comparison experiments, namely, (AlexNet +SVM, GoogleNet+SVM, and AlexNet+SVM), (AlexNet +SRC, GoogleNet+SRC, and DenseNet-201+SRC), and (AlexNet+NSCR, GoogleNet+NSCR, and DenseNet-201 +NSCR). To illustrate with the third group, in the CT sampling space, the accuracy of the DenseNet-201+NSCR model proposed in this paper is 0.28% and 0.58% higher, and the training time is 2460.66 s and 2900.24 s higher than the AlexNet+NSCR and GoogleNet+NSCR models, respectively. Not surprisingly, the DenseNet-201 has deep network layers, rich extracted image features, and high classification accuracy compared to other models. However, the cost is a significant increase in training time. The other two sets of results are similar and will not be recounted here.

In the second case, the same network and different classification algorithms are used. In experiment 1, there are three groups of comparison experiments, which are (AlexNet +SVM, AlexNet+SRC, and AlexNet+NSCR), (GoogleNet +SVM, GoogleNet+SRC, and GoogleNet+NSCR), and (DenseNet-201+SVM, DenseNet-201+SRC, and DenseNet-201 +NSCR). To illustrate the third set, in the CT sample space, the classification accuracy of the DenseNet-201+NSCR model proposed in this paper is better than that of the DenseNet-201+SVM is 0.84% higher and 0.78% higher than DenseNet-201+SRC. In terms of training time, it is 51.97 s more than the DenseNet-201+SVM model and 727.69 s lower than the DenseNet-201+SRC model. Compared to the first two cases, the overall training time is significantly improved. However, after the network model is determined, the increase in training time complexity compared to the SVM classifier is relatively reduced. Moreover, the time complexity is significantly reduced compared to the SRC classification algorithm. Not surprisingly, under the same network model, the nonnegative, sparse, and collaborative representation classification algorithm NSCR has better classification accuracy, which better solves the optimization problem of high-dimensional data and with a much lower time cost compared to SVM and SRC.

4.2.3. Experiment 3: Comparison of Different Combinations of Networks and Classifier Algorithms. The experiment focuses on nine algorithms (AlexNet+SVM, AlexNet+SRC, AlexNet +NSCR, GoogleNet+SVM, GoogleNet+SRC, GoogleNet +NSCR, DenseNet-201+SVM, DenseNet-201+SRC, and DenseNet-201+NSCR) trained on CT sampling space, in terms of accuracy, sensitivity, specificity, F value, and MCC for a total of five metrics to evaluate the merits of the algorithm. The results are shown in Table 4.

As shown in Table 4, the DenseNet-201+NSCR algorithm performance are all better than The DenseNet-201 +NSCR algorithm has better metrics than other algorithms in terms of accuracy, sensitivity, specificity, F value, and MCC on the CT dataset improved by 2.78%, 3.24%, 2.32%, 2.78, and 5.56%, respectively. To point out the differences between the different algorithms on each indicator more clearly, the mean values of these five indicators are plotted on a line graph with the three network models in horizontal coordinates and the five evaluation indicators in vertical coordinates, respectively, as shown in Figure 5.

Through these two experiments and the correlation analysis, one can easily see that, with the same network model, this paper compares three classification algorithms, SVM, SRC, and NSCR, and the result of the experiments show that NSCR classification outperforms SVM and SRC classification algorithms for DenseNet in medical image extraction. The NSRC algorithm has better robustness for the problems of DenseNet in which features extracted from medical images appear to have high dimensionality and data redundancy. With the same classification algorithm, this paper compares three network models, AlexNet, GoogleNet, and DenseNet-

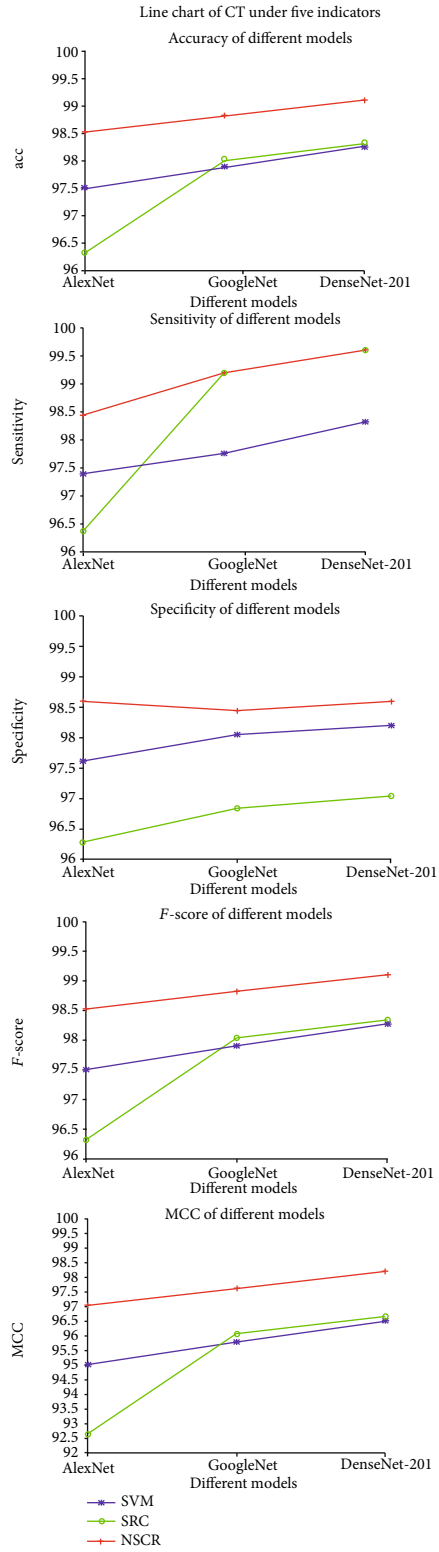


Figure 5: Line graphs of different indicators under CT and PET/CT datasets.

201. The results show that DenseNet outperforms AlexNet and GoogleNet models, and DenseNet effectively uses high-level information to rediscover new features at the bottom layer, enhances the propagation of features across networks, and implements and strengthens feature reuse. The result shows that DenseNet outperforms AlexNet and GoogleNet, especially the DenseNet-201+NSCR model with deep network depth, strong network generalization capability, high classification accuracy, and better accuracy, sensitivity, specificity, F value, and MCC than the other models.

5. Conclusion

In this paper, a DenseNet based on nonnegative, sparse, and collaborative representation classification for benign and malignant classification of lung tumors (DenseNet-NSCR) is proposed. First, CT medical images were collected and preprocessed. The dataset is then trained in a DenseNet to construct a DenseNet model to extract the full connection layer feature vector. Finally, the lung tumor classification results were obtained in the NSCR classifier and compared by AlexNet+SVM, AlexNet+SRC, AlexNet+NSCR, GoogleNet +SVM, GoogleNet+SRC, GoogleNet+NSCR, DenseNet-201 +SVM, DenseNet-201+SRC, and DenseNet-201+NSCR for a total of nine models. The DenseNet+NSCR model outperforms the other models with better robustness and generalization capabilities.

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 they have no conflicts of interest to report regarding the present study.

Acknowledgments

Thanks to a hospital in Ningxia of China for offering the PET, CT, and PET/CT medical image of lung tumor. This work was supported in part by the Natural Science Foundation of China under Grant 62062003, Key Research and Development Project of Ningxia (Special projects for talents) under Grant 2020BEB04022, and North Minzu University Research Project of Talent Introduction (2020KYQD08).

References

[1] K. Gopi, J. Selvakumar, and H. G. Amir, "Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier," *Expert Systems with Applications*, vol. 134, pp. 112–119, 2019.

[2] C. Kaushal, S. Bhat, and D. Koundal, "Recent trends in computer assisted diagnosis (cad) system for breast cancer diagnosis using histopathological images," *Innovation & Research in Biomedical En*, vol. 40, no. 4, pp. 211–227, 2019.

[3] M. Andreas, S. Christopher, and L. Tobias, "A gentle introduction to deep learning in medical image processing," *Zeitschrift für Medizinische Physik*, vol. 29, pp. 86–101, 2019.

[4] H. Gao, L. Zhuang, and V. D. M. Laurens, "Densely connected convolutional networks," in *IEEE Conference on Computer*

Vision and Pattern Recognition(CVPR), pp. 2261–2269, Hawaii, USA, 2017..

[5] K. M. He, X. Y. Zhang, and S. Q. Ren, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[6] K. M. He, X. Zhang, and S. Ren, "Identity mappings in deep residual networks," in *Computer Vision-ECCV 14th European Conference*, pp. 630–645, Amsterdam, The Netherlands, 2016.

[7] M. Khened, V. A. Kollerathu, and G. Krishnamurthi, "Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers," *Medical Image Analysis*, vol. 51, pp. 21–45, 2019.

[8] R. X. Cui and M. H. Liu, "Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 5, 2019.

[9] W. Tong, W. T. Chen, W. Han, X. J. Li, and L. Z. Wang, "Channel-attention-based DenseNet network for remote sensing image scene classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, 2020.

[10] B. Z. Chen, J. X. Li, and X. B. Guo, "DualCheXNet: dual asymmetric feature learning for thoracic disease classification in chest x-rays," *Biomedical Signal Processing and Control*, vol. 53, p. 101554, 2019.

[11] Y. J. DAI, C. L. SONG, and S. J. YAN, "Benign or malignant lung nodules classification model based on modified DenseNet," *Chinese Journal of Medical Imaging Technology*, vol. 34, pp. 1104–1109, 2018.

[12] H. ZHU and P. L. QIN, "U-Net pulmonary nodule detection algorithm based on multi-scale feature structure," *Computer Engineering*, vol. 45, no. 4, pp. 254–261, 2018.

[13] X. C. Li, L. L. Shen, and X. P. Xie, "Multi-resolution convolutional networks for chest x-ray radiograph based lung nodule detection," *Artificial Intelligence in Medicine*, vol. 103, article 101744, 2020.

[14] X. Jun, W. P. An, Z. Lei, and Z. David, "Sparse, collaborative, or nonnegative representation: which helps pattern classification?," *Pattern Recognition*, vol. 88, pp. 679–688, 2019.

[15] Y. Li, L. Wei, and L. Yang, "Non-negative dictionary based sparse representation classification for ear recognition with occlusion," *Neurocomputing*, vol. 171, pp. 540–550, 2016.

[16] J. X. Mi, Y. R. Sun, and J. Lu, "Robust supervised sparse representation for face recognition," *Cognitive Systems Research*, vol. 62, pp. 10–22, 2020.

[17] Y. G. Zhang, T. W. Xu, and J. M. Ma, "Image categorization using non-negative kernel sparse representation," *Neurocomputing*, vol. 269, pp. 21–28, 2017.

[18] G. Majid, K. Manoochehr, E. Farshad, and S. Arash, "FDSR: a new fuzzy discriminative sparse representation method for medical image classification," *Artificial Intelligence in Medicine*, vol. 106, p. 101876, 2020.

[19] G. Majid, K. Manoochehr, E. Farshad, and S. Arash, "T2-FDL: a robust sparse representation method using adaptive type-2 fuzzy dictionary learning for medical image classification," *Expert Systems with Applications*, vol. 158, p. 113500, 2020.

[20] M. Nooshin and M. A. Nezam, "Kernel sparse representation based model for skin lesions segmentation and classification," *Computer Methods and Programs in Biomedicine*, vol. 182, article 105038, 2019.

[21] R. J. Zhang, J. Shen, F. S. Wei, X. Li, and K. S. Arun, "Medical image classification based on multi-scale non-negative sparse coding," *Artificial Intelligence in Medicine*, vol. 83, pp. 44–51, 2017.

[22] J. P. Gou, H. W. Wu, and H. P. Song, "Double competitive constraints-based collaborative representation for pattern classification," *Computers & Electrical Engineering*, vol. 84, article 106632, 2020.

[23] X. Jun, X. Zhou, and P. Wang, *Jointly non-negative, sparse and collaborative representation for image recognition*, 2019.

[24] S. Boyd, N. Parikh, and E. Chu, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations & Trends in Machine Learnin*, vol. 3, pp. 1–122, 2010.

[25] T. Zhou, H. L. Lu, F. Y. Hu, S. Qiu, and C. Y. Wu, "A model of high-dimensional feature reduction based on variable precision rough set and genetic algorithm in medical image," *Mathematical Problems in Engineering*, vol. 2020, 18 pages, 2020.

[26] T. Zhou, H. L. Lu, W. W. Wang, and Y. Xia, "GA-SVM based feature selection and parameter optimization in hospitalization expense," *Applied soft computing*, vol. 75, pp. 323–332, 2019.