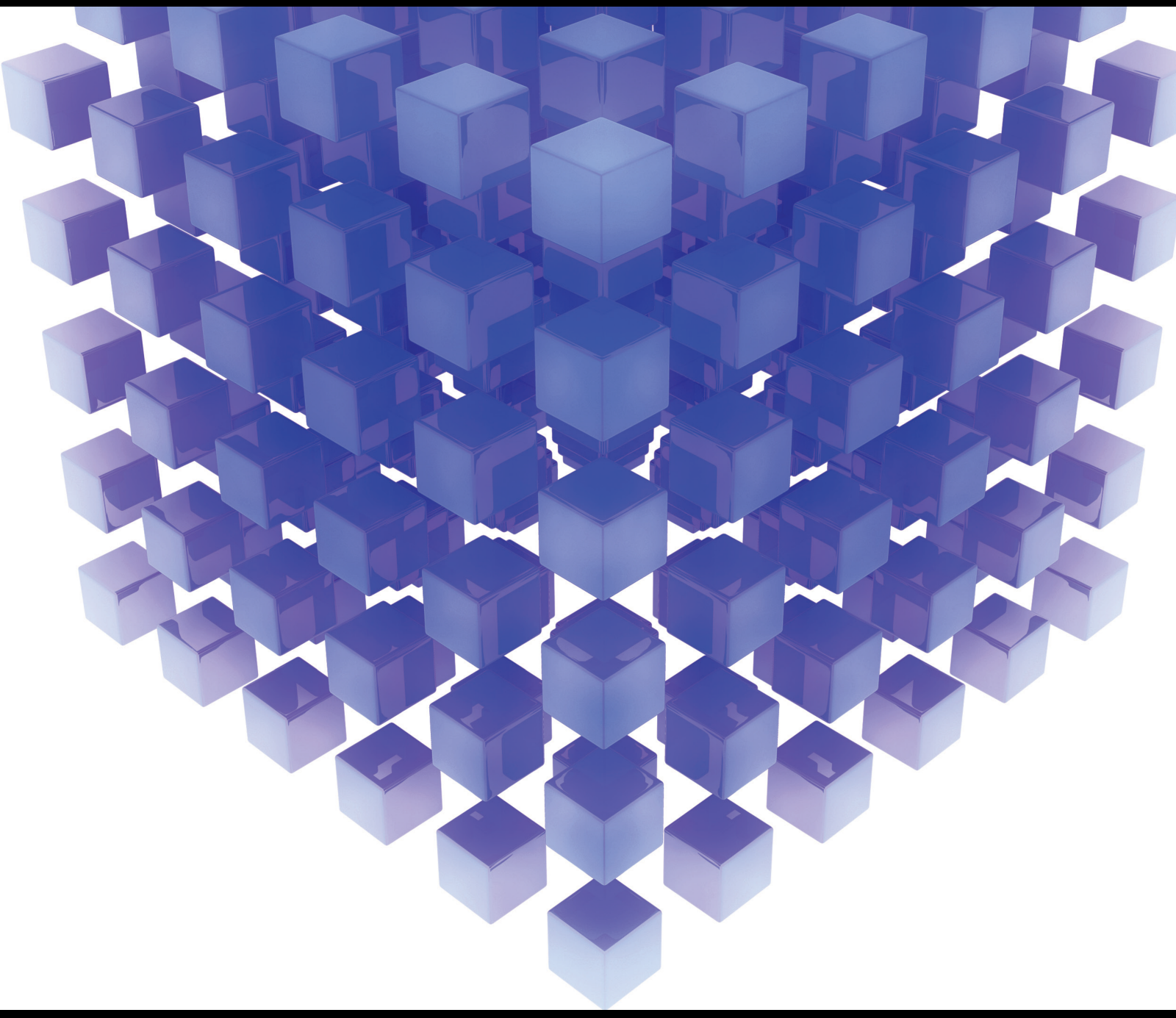
Mathematical Problems in Engineering

Deep Learning in Biological and Medical Data Analysis

Lead Guest Editor: Xiaofeng Li
Guest Editors: Jiliang Tang, Xu-Qing Tang, and J.H. Pang



Deep Learning in Biological and Medical Data Analysis

Mathematical Problems in Engineering

Deep Learning in Biological and Medical Data Analysis

Lead Guest Editor: Xiaofeng Li
Guest Editors: Jiliang Tang, Xu-Qing Tang, and J.H. Pang

Chief Editor

Academic Editors

Contents

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 2362851, 10 pages
https://doi.org/10.1155/2022/2362851

Research Article

X-Ray Breast Images Denoising Method Based on the Convolutional Autoencoder

Chensheng Yang, Junbo Ye, Yanwei Wang, and Chengling Song

School of Mechanical Engineer, Heilongjiang University of Science & Technology, Harbin, Heilongjiang, China

Correspondence should be addressed to Yanwei Wang; wangyanwei@usth.edu.cn

Received 2 June 2022; Revised 23 August 2022; Accepted 3 October 2022; Published 22 November 2022

Academic Editor: Xiaofeng Li

Considering the potential risk of X-ray to patients, denoising of low-dose X-ray medical images is imperative. Inspired by deep learning, a convolutional autoencoder method for X-ray breast image denoising is proposed in this paper. First, image symmetry and flip are used to increase the number of images in the public dataset; second, the number of samples is increased further by image cropping segmentation, adding simulated noise, and producing the dataset. Finally, a convolutional autoencoder neural network model is constructed, and clean and noisy images are fed into it to complete the training. The results show that this method effectively removes noise while retaining image details in X-ray breast images, yielding higher peak signal-to-noise ratio and structural similarity index values than classical and novel denoising methods.

1. Introduction

Medical images are frequently utilized in modern clinical diagnosis and therapy to aid in disease diagnosis and treatment evaluation, among other things. They provide a crucial foundation for clinical diagnosis and treatment. Unlike natural photos, medical images generate a lot of signal-related noise during the creation process; therefore, the contrast is lower and the noise is more visible [1]. As a type of medical imaging, X-ray is of great value in early breast cancer detection. Patients should receive mammography with the lowest possible radiation dose [2]. The most common way to reduce the radiation dose is to reduce the X-ray flux by decreasing the operating current and shortening the exposure time of the X-ray tube. However, the weaker the X-ray flux, the more-noisy the reconstructed image will be. Noise can blur the image, obscure important information in the image, and make disease analysis and diagnosis more difficult. Therefore, it is important to study how to remove the noise from X-ray images.

In the past few decades, many scholars have been proposing new image denoising algorithms as image noise has been intensively studied. Traditional image denoising models can be classified into four categories based on spatial domain, transform domain, sparse representation, and natural statistics. Among them, the representative methods are the median filtering method based on the spatial domain [3], which ignores the characteristics of each pixel, and the image will be more seriously blurred after denoising; the BLS-GSM [4] based on the transform domain lose some useful information while denoising; the NLSC [5] based on the sparse representation has a long computation time and low denoising efficiency. The natural statistics-based BM3D [6] can only filter a specific noise. Although these algorithms are effective in removing noise, they often inevitably result in loss of texture information in the medical image and excessive smoothening of edges, which adversely affects diagnosis. It is still a challenging problem to remove the noise while retaining the detailed information in the medical image.

With the improvement of hardware computing power, the powerful learning and fitting capabilities of neural networks have shown great potential in image processing [7]. For example, convolutional neural networks (CNNs) have been applied to image classification, target detection, image segmentation, image denoising, etc. Currently, many scholars

have used CNN to denoise medical images such as low-dose CT images [8, 9], OCT images [10], and MRI images [11, 12], ultrasonography images [13], and so on. Kim et al. improved the BM3D method by proposing a method to assign different weights to each block according to the degree of denoising [14]. Although the detailed information of the image can be well recovered, the Gibbs effect will be produced after denoising, and the artifacts will be produced which cannot be eliminated. Guo et al. proposed a median filtering method based on adaptive two-level threshold to solve the problems of low contrast and blurred boundary of traditional weighted median filters [15]. This method has a very good denoising effect on CT images of COVID-19, but it is not suitable for denoising mixed noise, which is easy to cause image blur and discontinuity. Jia et al. proposed a pyramid dilated CNN [16], which uses dilated convolution to expand the network's receptive field and obtain more image details. This method has good denoising effect on both gray image and color image. However, the dilation rate needs be adjusted according to the size of the object in the input image to avoid image discontinuity. Huang et al. proposed a denoising GAN based on the U-Net discriminator [17], which can not only give feedback to each pixel in the image but also use U-Net to focus on the global structure at the semantic level. This method has a very good denoising effect on low-dose CT images, but the model is complicated and difficult to train. However, there are few studies on the denoising of X-ray breast images. This inspired us to use a CNN-based denoising model to remove noise from X-ray breast images and improve the quality of X-ray breast images. This paper proposes an X-ray breast image denoising method based on a convolutional autoencoder. First is by expanding the public breast dataset of Mammographic Image Analysis Society MiniMammographic Database (MIAS) and then centrally cropping the key parts to intercept the data containing key medical information while further expanding the number of samples; second is by adding Gaussian noise and salt and pepper noise to the sample data to generate noisy images and then combining the clean and noisy images into a dataset; and the final one is building an autoencoder denoising model based on CNN and completing the training. The results of the experiments show that the method can effectively remove various levels of blending noise in medical images while retaining image detail texture. The following are the main contributions of this paper: (1) We achieved better results than the current denoising methods, with higher peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM); (2) we proposed a mathematical model to simulate blending noise which was implemented by code to add different levels of blending noise in medical images; (3) in the MIAS dataset, the performance of this method is tested from the perspective of multiple index analysis, which fully verifies the efficiency and superiority of this method; and (4) we completed end-to-end modeling for medical image denoising.

2. Proposed Method

2.1. Simulated Medical Image Noise. In imaging, complex noise sources include Gaussian, impulse, pretzel, and scatter noise [18]. Salt and pepper noise, Gaussian noise, and other types of noise are common in medical images. Because noise-free images are not readily available in clinics, and the network model of proposed method requires high-resolution noise-free images to train the network. As a result, the open dataset is chosen, and the open dataset is a clear clean image free of noise obtained by adding simulated noise to the image to obtain paired training data to train the noise removal network model. In this paper, noise is added to an image by adding noise components to the values of the image's corresponding pixels. The noise model can be defined as follows:

$$N_{(h,w,c)} = O_{(h,w,c)} + Y_{(h,w,c)} + G_{(h,w,c)}, \tag{1}$$

where (h, w, c) represents the pixel point in the image, h represents the image height, w represents the image width, c represents the number of channels, and when $c = 1$ represents the single-channel gray image, $O_{(h,w,c)}$ represents original images, $Y_{(h,w,c)}$ represents salt and pepper nose data, and $G_{(h,w,c)}$ represents Gaussian noise data.

2.1.1. Generation of Gaussian Noise. Gaussian noise is a kind of noise that obeys Gaussian distribution, where μ represents the expectation of Gaussian noise data and σ represents the variance of Gaussian noise data. In this paper, the method of adding Gaussian noise to medical images is as follows:

$$\begin{aligned} Z &\sim N(\mu, \sigma), \\ G_{(h,w,c)} &= kZ_{(h,w,c)}, \end{aligned} \tag{2}$$

where k represents noise intensity, $Z_{(h,w,c)}$ represents denoising images, considering the pixel of images which may over 255 before adding noise, and the pixel of the noisy image is limited to avoid data overflow in the computer, and the restriction is as follows:

$$N_{(h,w,c)} \begin{cases} 0, N_{(h,w,c)} < 0, \\ 255, 255 < N_{(h,w,c)}, \\ N_{(h,w,c)}, 0 \le N_{(h,w,c)} \le 255. \end{cases} \tag{3}$$

2.1.2. Salt and Pepper Noise Generation. Salt and pepper noise generation is also a common noise in the image, as the name suggests, and pepper represents black, and salt represents white. Salt and pepper noise represents white or black points. The generation way is as follows:

$$N_{(h,w,c)} \begin{cases} 255, 0 \le W \le 1 - \text{SNR and } Z = 1, \\ 0, 0 \le W \le 1 - \text{SNR and } Z = 0, \\ O_{(h,w,c)}, \text{other}, \end{cases} \tag{4}$$

$$\begin{aligned} W &\sim U(0, 1), \\ Z &\sim B(1, 0.5), \end{aligned}$$

where SNR represents signal-to-noise ratio, which is expressed in the range of [0, 1], W obeys the uniform distribution of parameter [0, 1], Z obeys the 0.1 distribution when $P(1) = 0.5$ and $W \in [0, (1 - \text{SNR})]$, and the value of

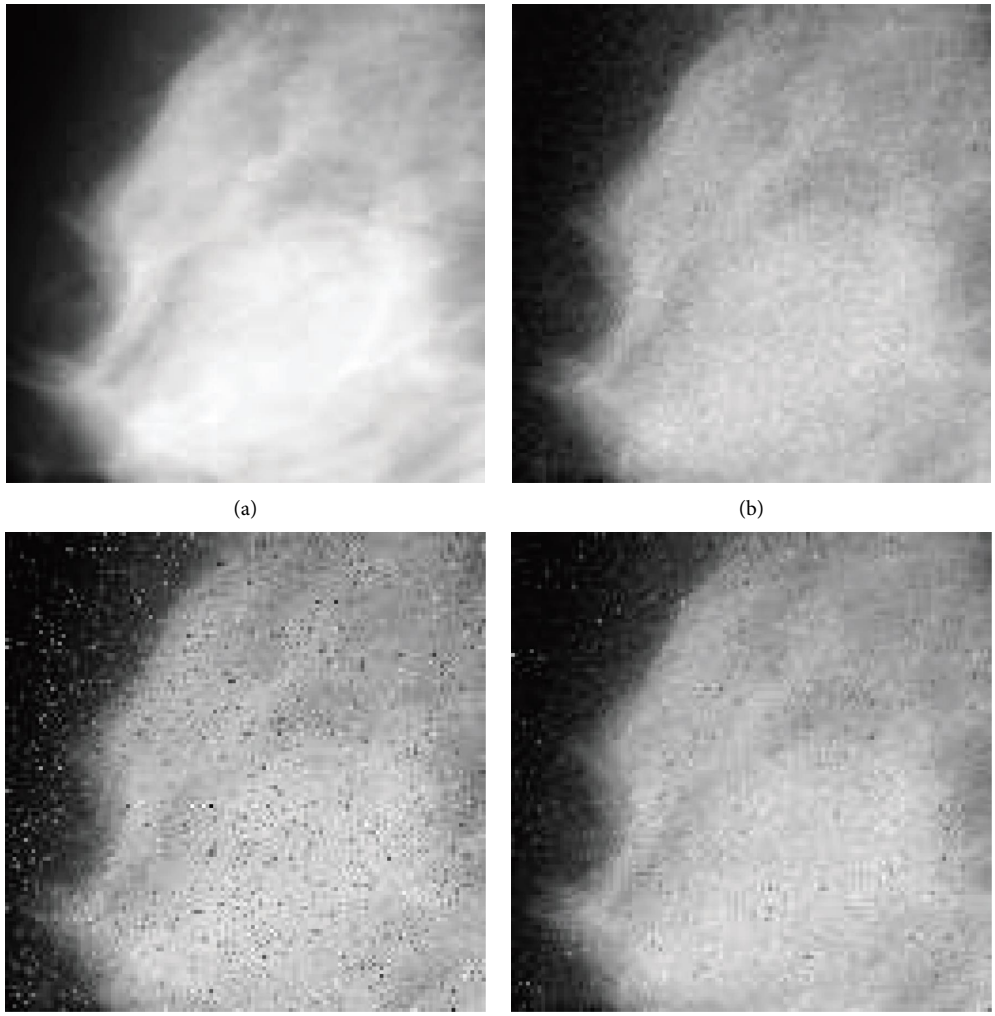


Figure 1: Noise images. (a) Origin image. (b) Gaussian noise. (c) Salt and pepper noise. (d) Mixed noise.

pixels in the original image in each channel has a 50% chance of being 255 or 0. When the pixel value is 255, it shows black salt and pepper noise points, and when it is 0, it shows white salt and pepper noise points.

In Figure 1, (a) represents the original image, (b) represents the image after adding Gaussian noise, (c) represents the image after adding salt and pepper noise, and (d) represents the mixed image after adding Gaussian noise and salt and pepper noise.

2.2. Convolutional Autoencoder Model. The concept of autoencoder was first proposed by Rumelhart et al. [19], which was originally applied to dimension reduction of complex data. The autoencoder includes encoding and decoding and trains the network through reverse propagation, so that the output is equal to the input [20]. The encoder can reduce the dimension of data compression, reduce the amount of data, and retain the most critical feature information. The function of the decoder is opposite to the encoder. The decoder restores the compressed data and restores the input data through decoding. The full connection layer of the traditional autoencoder will stretch the data into one dimension, thus losing the spatial information of the two-dimensional image data. The CNN has strong performance in extracting spatial feature information of images, which can compensate for the loss of spatial information when an autoencoder extracts features. Based on the traditional autoencoder, convolutional autoencoder combines the advantages of CNN and realizes the deep neural network through the superposition of convolution layer, activation layer, and pooling layer to complete the extraction of image detail features [21]. This paper uses the convolution layer, pooling layer, activation layer, and deconvolution layer in the CNN to construct an encoder and decoder. Encoders and decoders can be defined as

$$\begin{aligned} M &= f\left(W_i * X + b_i\right), \\ N &= f\left(W_i * X + b_i\right), \end{aligned} \tag{5}$$

where W_i and b_i represent the weight matrix and bias of each convolution layer, respectively; i represents the i convolution layers; f and h represent the encoder and decoder, respectively; X represents the input medical image data; M

represents the feature information of input image data after feature extraction by the encoder; and N represents the medical image data generated after the feature information is decoded.

2.3. Loss Function and Optimization Algorithm. The mean squared difference loss function is used in this paper to compute the difference between the denoised noisy image and the clear image. The mean squared difference of image reconstruction in this paper is as follows:

$$\begin{aligned} \mathrm{MSE} &= \frac{1}{M}\sum_{i=1}^{M}\left(f_p(y_i,\theta)-f_i\right)^2, \\ \theta &= \{w,b\}, \end{aligned} \tag{6}$$

where M represents the number of training samples; θ represents the parameters in the network model; w represents weight; and b represents a bias. $f_p(y_i,\theta)$ represents the output image of the denoising model, and f_i denotes the noise-free image corresponding to the output image of the model. When the mean square difference is smaller, the reconstruction effect is better. To make the model converge faster and better [22], the learning process of the model is optimized using the Adam algorithm, which is based on the gradient descent method but differs from the traditional stochastic gradient descent algorithm that can make the model converge faster and better. To update all the weights, stochastic gradient descent uses a single learning rate that does not change during the training process. In contrast, the Adam algorithm calculates the first-order moment estimates and second-order moment estimates of the gradient to design independent adaptive learning rates for different parameters, which has significant advantages for model optimization of large-scale datasets. The Adam algorithm's optimization procedure is as follows:

$$\begin{cases} g_t = \nabla_\theta F_t(\theta_{t-1}), \\ m_t = \beta_1 m_{t-1} + (1-\beta_1)g_1, \\ v_t = \beta_2 v_{t-1} + (1-\beta_2)g_t^2, \\ \hat{m_t} = \dfrac{m_t}{\left(1-\beta_1^t\right)}, \\ \hat{v_t} = \dfrac{v_t}{\left(1-\beta_2^t\right)}, \end{cases} \tag{7}$$

$$\theta_i^t = \theta_{i-1}^i - \alpha\frac{\hat{m_i^t}}{\sqrt{\hat{v_t^i}}+\varepsilon},$$

where t is the current time step; F is the optimization function of the model; g_t is the gradient of the optimization objective function F at the time step t; θ is the model parameter vector; and θ_i^t is the value of the first element θ_i in the parameter vector θ when t time step. α is the step length; m is the first-order moment estimation of gradient; β_1 is the exponential decay rate of m; v is the second-order moment estimation of gradient; β_2 is the exponential decay rate of v; and β_1, $\beta_2 \in [0,1)$; $\varepsilon = 10^{-8}$ equation (6) represents each update of the parameter vector θ. However, as the neural network is trained, the distribution of the parameters of the next layer will change from the parameters of the previous layer. This causes a constant change in the distribution of parameters, which slows down training. This issue can be addressed by incorporating a normalization layer into the network model and performing normalization for each small batch of samples [23]. Suppose the input of a layer in the model is $x = (x^1, x^2, \ldots, x^n)$ and the set of samples is $B = (x_1, x_2, \ldots, x_m)$. The batch normalization method is as follows:

$$\begin{aligned} \mu_B &= \frac{1}{m}\sum_{i=1}^{m} x_i, \\ \sigma_B^2 &= \frac{1}{m}\sum_{i=1}^{m}\left(x_i-\mu_B\right)^2, \\ \hat{x^{(n)}} &= \frac{x^{(n)}-\mu_B}{\sigma_B}, \\ y^{(n)} &= \gamma^{(n)} * \hat{x^{(n)}} + \beta^{(n)}, \end{aligned} \tag{8}$$

where $x^{(n)}$ is the nth dimension of the input x; μ_B is the sample set B's expectation; σ_B^2 is the variance of the sample set B; $x^{(k)}$ is the input's regularization result; $y^{(n)}$ is the batch regularization result of $x^{(n)}$; and $\gamma^{(n)}$ and $\beta^{(n)}$ is the parameters to be learned. The network model species' parameters are then continuously updated using back-propagation and optimization algorithms to produce the best denoising model.

3. Design of the Denoising Method

3.1. Denoising Process. Figure 2 depicts the overall flow of the denoising method. To begin with, a public dataset is chosen for data preprocessing. The MIAS dataset with 322 high-definition $1024 * 1024$ mammographic images are used in this paper. The MIAS dataset is then cropped, rotated, and flipped to increase the volume of training data to 3000 images with $336 * 336$ resolution. Second, simulated noise is added to the cropped and expanded clear image data, and the noise images and clear images are combined into a dataset that can be passed into the model, with the samples in the dataset divided into an 8 : 1 : 1 training, test, and validation set. The convolutional autoencoder neural network model is then built, and the training set is fed into the model to complete the model's training. The validation set is then fed into the trained model, which checks the model's denoising effect and outputs the model parameters. Finally, the model is loaded and the test set is fed into it to test the model's denoising effect.

Figure 2 depicts the process of the denoising method used in this paper: first, the public dataset is converted into a

Figure 2: Flow chart of the denoising method.

dataset, and then the dataset is divided into a training set, a validation set, and a test set. The model training is completed with the help of the training set. The validation set is used to determine whether the trained model is overfitting. The test set is used to validate the model's denoising effect. Finally, the model is then saved.

3.2. Model Structure. The model's goal is to create an end-to-end mapping of noisy image noise to a clear image. The process of converting a noisy image to a clear image is known as image denoising. This paper's convolutional autoencoder denoising network is divided into two parts: encoder and decoder. The encoder compresses and downscales the input noisy image through feature extraction via convolution before transforming it into an abstract mathematical description. By deconvolution, the decoder converts this abstract description into an image. As the cost function is reduced, the decoder generates images that are increasingly similar to the target image. The structural model of convolutional autoencoder proposed in this paper is shown in Figure 3, where the input is the noisy image and the output is the image with completed noise reduction processing. The model contains a total of 12 convolutional layers, the first 6 layers belong to the encoder structure, and the last 6 layers belong to the decoder structure. The first four layers of the encoder contain 32 convolutional kernels, the fifth layer contains 64 convolutional kernels, and the sixth layer contains 128 convolutional kernels. The first two layers of the decoder contain 64 convolutional kernels, layers 3 and 4 contain 32 convolutional kernels, layer 5 contains 16 convolutional kernels, and the last layer contains 1 convolutional kernel. The size of all convolutional kernels in the model is $3 * 3$. A $3 * 3$ convolution kernel has the same perceptual field after three convolutions as a $7 * 7$ convolution kernel after 1 convolution; that is, a small-sized convolution kernel can obtain the same perceptual field as a large-sized convolution kernel through multiple convolution operations, and the multiple convolution operations of a small-sized convolution kernel can increase the network depth and improve the network's nonlinear fitting ability [24]. Furthermore, the small convolution kernel reduces the number of parameters and improves the model's convergence speed. As a result, small-sized convolutional kernels of $3 * 3$ are used in the models developed in this paper. A ReLU layer is added as the activation function to ensure the gradient descent speed and model convergence. After the encoder's third layer, the model adds a maximum pooling layer. The pooling layer can reduce redundant information while also expanding the receptive field.

Figure 3 depicts the end-to-end medical image denoising model developed in this paper, with the noisy image as the input, and the regenerated image after noise reduction as the output.

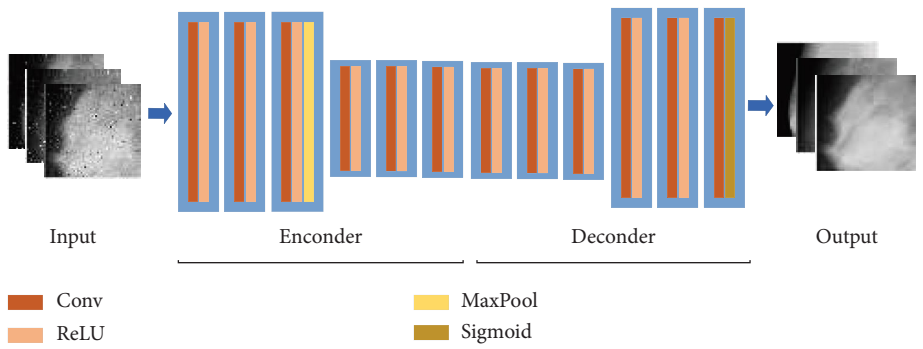


Figure 3: Structure model of the convolutional autoencoder.

4. Experimental Analysis and Results

4.1. Experimental Set-Up

(1) *Data for Training and Testing.* To train the proposed model's denoising ability under different levels of noise, we used the MIAS dataset to generate the training dataset. In the dataset used for model training, four different intensities of Gaussian noise and four different intensities of salt and pepper noise were set. The standard deviations of Gaussian noise are set to 40, 50, 60, and 70, respectively; the signal-to-noise ratios of salt and pepper noise were set to 0.99, 0.97, 0.95, and 0.93, respectively. The change of mixed noise intensity in the proposed method is completed by setting parameters in the program. An image training set consists of 2400 $336 * 336$ MIAS breast X-ray images. The test set and validation set were 300 MIAS breast X-ray images of $336 * 336$.

(2) *Super Parameter Setting and Model Training.* Based on a large number of experimental results, the depth of the model is determined to be 12 layers by weighing the receptive field and the model depth. When the model is trained, the number of mini-batch is 4, the initial learning rate is 0.01, and after three rounds of iteration, the learning rate is 0.0001, so that the model continues to learn. If the loss function of the model does not change after five consecutive iterations, the model stops iterative training.

(3) *Experimental Environment.* To train the model with a better denoising effect, a large number of experiments are performed in this paper. The experimental environment mainly includes hardware and software. The hardware configuration is mainly Intel Core i7 4810M CPU and NVIDIA GeForce960M GPU, in which the graphics card memory is 2G, and the running memory is 12G. The software environment includes operating system Windows 10 64-bit, programming language Python 3.8, deep learning framework PyTorch 1.10, and Python data integration package Anaconda.

(4) *Evaluation Indicators.* Because it is a medical image, this paper employs both objective and subjective evaluation methods. The subjective evaluation is the human's subjective reaction to the denoised image. PSNR is the most commonly used image quality evaluation criterion among image researchers, but the visual effect of an image is not identical to the PSNR. Because the PSNR index is not always the highest for the best visual effect in a set of images, it cannot be used as an absolute criterion to assess image quality [25]. SSIM structural similarity, on the other hand, comprehensively takes into account nonstructural distortions such as brightness and contrast as well as structural distortions such as junction noise intensity and blurring degree [26]. This means that in the context of visual perception, SSIM can better reflect the image quality. Therefore, in this experiment, we comprehensively considered PSNR and SSIM to evaluate the denoising effect of the model. The calculation formula is as follows:

$$\mathrm{PSNR} = 10 * \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right),$$

$$\begin{cases} l(x,y) = \dfrac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}, \\ c(x,y) = \dfrac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + c_2}, \\ s(x,y) = \dfrac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}, \\ \mathrm{SSIM}(x,y) = \left[l(x,y)^{\alpha} * c(x,y)^{\beta} * s(x,y)^{\gamma}\right], \end{cases} \tag{9}$$

where MAX is the maximum value of pixels in the image, and all the images involved in this paper are 8-bit color depth, so MAX = 255; MSE is the mean square value; x and y are input images; μ is the mean value of the input image; σ is the variance of the input image; and α, β, γ are power exponents used to adjust the importance of input images. Since the input image is equally important, here we take the power

Table 1: The PSNR (dB) and SSIM value of different methods at the MIAS dataset (SNR = 0.99, $k = 0.1$).

Different methods	$\sigma = 40$		$\sigma = 50$		$\sigma = 60$		$\sigma = 70$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Noise image	23.37	0.33	22.76	0.27	21.97	0.21	21.21	0.17
Literature [14]	37.39	0.90	36.19	0.87	34.83	0.82	34.22	0.80
Literature [15]	37.51	0.90	37.24	0.89	35.26	0.83	34.68	0.81
Literature [16]	38.37	0.91	38.03	0.90	37.37	0.88	35.29	0.84
Literature [17]	38.91	0.92	38.25	0.91	37.28	0.88	35.87	0.86
Proposed	**38.95**	**0.93**	**38.31**	**0.92**	**37.48**	**0.91**	**36.36**	**0.90**

Table 2: The PSNR (dB) and SSIM of different methods at the MIAS dataset ($k = 0.1$, $\sigma = 60$).

Different methods	SNR = 0.99		SNR = 0.97		SNR = 0.95		SNR = 0.93	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Noise image	21.97	0.21	19.22	0.14	17.55	0.10	16.32	0.07
Literature [14]	37.15	0.90	36.15	0.88	34.72	0.85	32.41	0.82
Literature [15]	36.82	0.89	35.64	0.87	33.48	0.84	31.52	0.81
Literature [16]	37.24	0.91	36.23	0.88	36.04	0.88	34.91	0.85
Literature [17]	37.29	0.91	36.59	0.89	36.21	0.88	35.73	0.87
Proposed	**37.48**	**0.91**	**37.07**	**0.90**	**36.47**	**0.89**	**35.93**	**0.88**

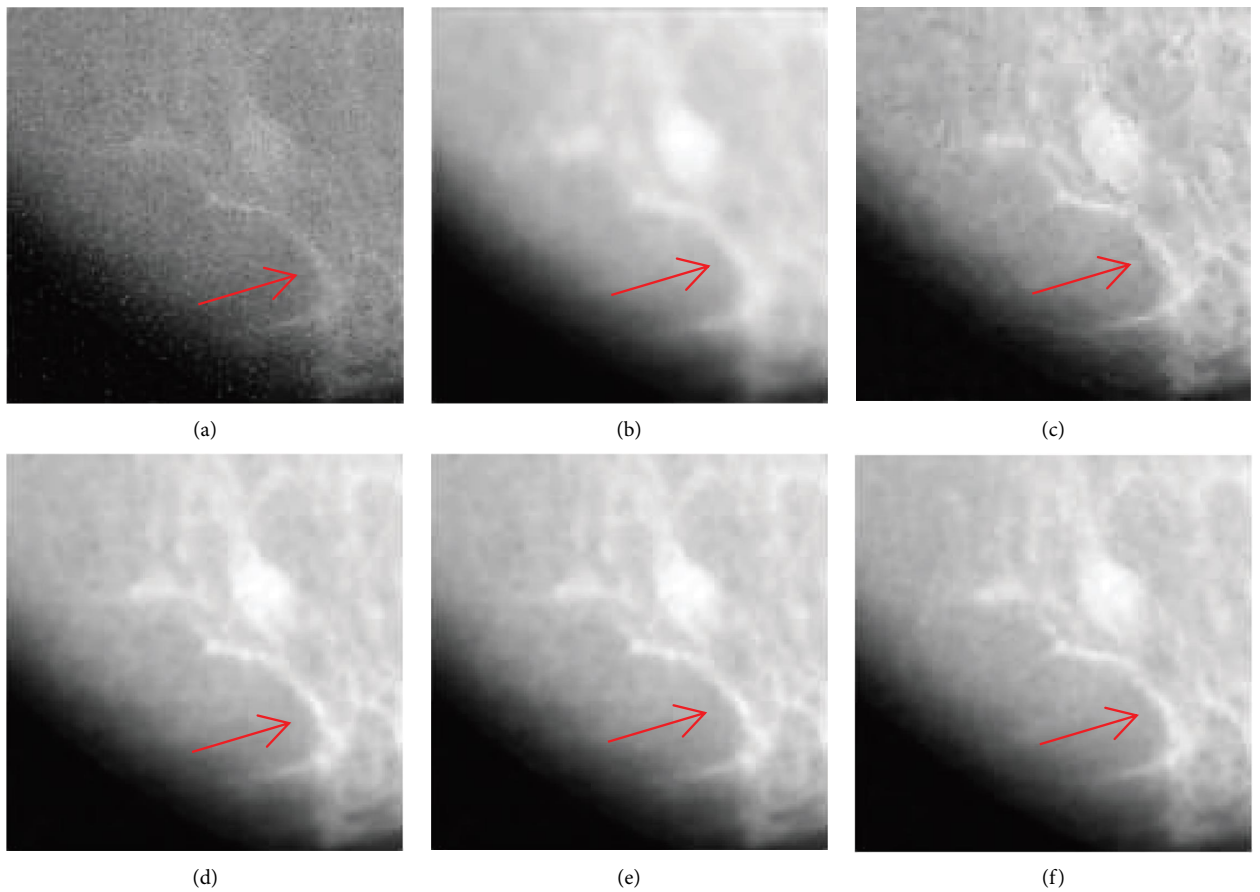


Figure 4: Denoising results of different methods. (a) Noisy image (PSNR = 21.53 dB) (SSIM = 0.26). (b) Literature [14] (PSNR = 33.47 dB) (SSIM = 0.87). (c) Literature [15] (PSNR = 33.58) (SSIM = 0.83). (d) Literature [16] (PSNR = 36.18 dB) (SSIM = 0.88). (e) Literature [17] (PSNR = 37.41 dB) (SSIM = 0.90). (f) Proposed method (PSNR = 37.83 dB) (SSIM = 0.91).

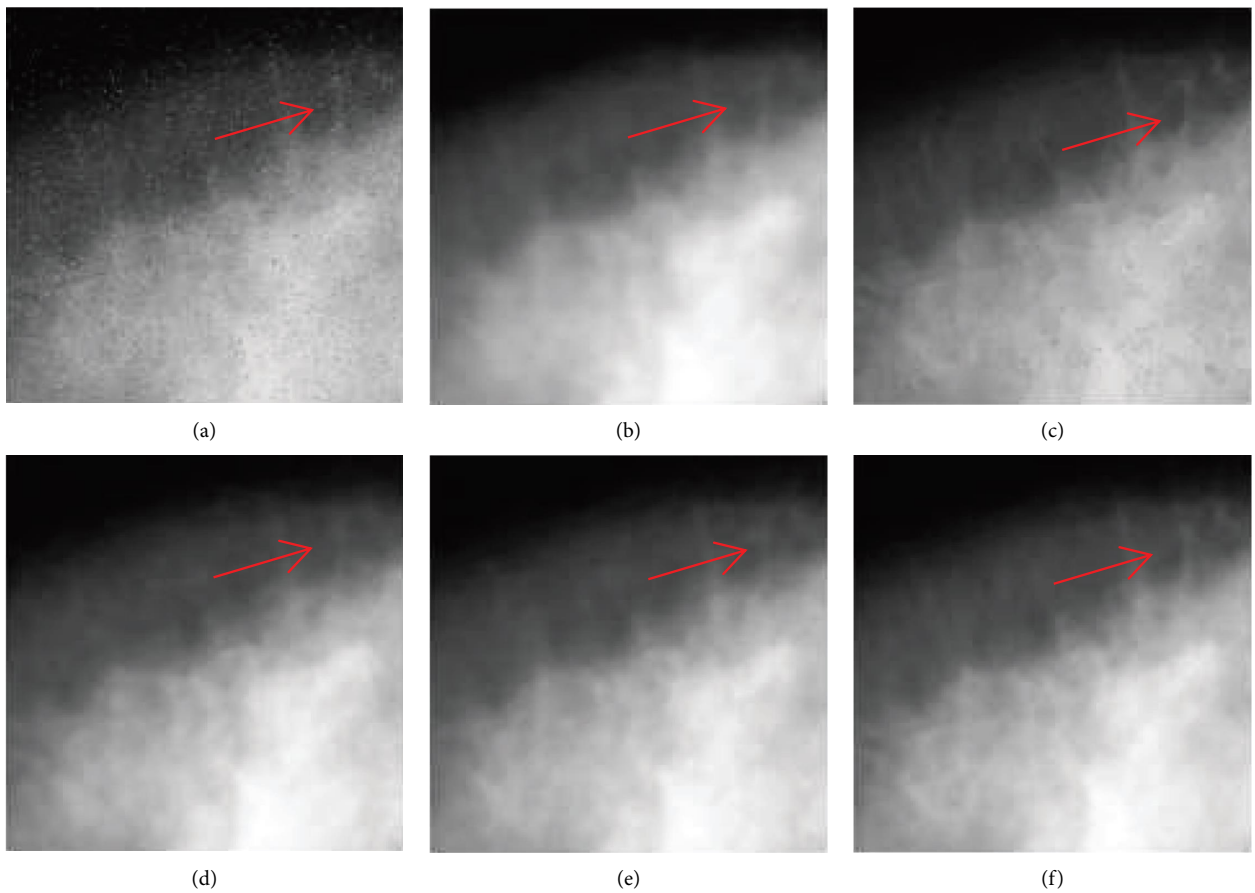


Figure 5: Denoising results of different methods. (a) Noisy image (PSNR = 22.58 dB) (SSIM = 0.27). (b) Literature [14] (PSNR = 32.17 dB) (SSIM = 0.51). (c) Literature [15] (PSNR = 33.78) (SSIM = 0.75). (d) Literature [16] (PSNR = 33.49 dB) (SSIM = 0.82). (e) Literature [17] (PSNR = 34.91 dB) (SSIM = 0.86). (f) Proposed method (PSNR = 35.17 dB) (SSIM = 0.87).

index as 1; l represents brightness, c represents contrast, and s represents structure.

4.2. Results and Discussion. To show the proposed method's effectiveness in removing noise from X-ray breast pictures and its benefits in maintaining texture. In this study, we conduct an experimental comparison of the proposed method's denoising performance with that of novel denoising methods. These methods include those mentioned in literatures [14], [15], [16], and [17]. In our experiments, we denoise the dataset constructed based on MIAS, and the experimental results are shown in Tables 1 and 2. All methods present visually well denoised results to some degree. When $\sigma = 40$, the PSNR of literature [16] and literature [15] are 0.98 dB and 0.12 dB greater than literature [14], respectively. The PSNR of literature [17] is 0.54 dB higher than literature [16], while the PSNR of the proposed approach in this study is 0.04 dB higher than literature [17]. The average SSIM values of literature [16] are improved by 0.01 compared to literature [14] and literature [15], and the SSIM of literature [17] is improved by 0.01 compared to literature [16]. The SSIM of the proposed method in this paper is improved by 0.01 compared to literature [17]. In addition, with the increase in noise intensity, $\sigma = 50$, $\sigma = 60$, and $\sigma = 70$, and it can be seen from the experimental data in Table 1 that the proposed method in this paper performs well on PSNR and SSIM. When SNR = 0.99, the PSNR of literature [16] and literature [14] are 0.42 dB and 0.33 dB greater than literature [15], respectively. Literature [17] is 0.05 dB greater than literature [16], while the PSNR of the proposed approach in this study is 0.19 dB higher than literature [17]. The average SSIM values of literature [16], literature [17] and literature [14] are improved by 0.02, 0.02, and 0.01 compared to literature [15], respectively, and the SSIM of the proposed method has the same value with literature [17] and literature [16]. In addition, with the decrease in signal-to-noise ratio, SNR = 0.97, SNR = 0.95, and SNR = 0.93, and it can be seen from the experimental data in Table 2 that the proposed method performs well on PSNR and SSIM.

We chose two images from the test set to display in Figures 4 and 5 to further confirm the visualization effect of the technique for X-ray image denoising described in this research. In Figure 4, the noise intensity factor is 0.1, the variance of the noisy image is = 0.6, and the SNR is 0.99. In Figure 5, the noise intensity factor is 0.1, the variance of the noisy image is 0.6, and the SNR is 0.93. It is observed that the literature [14] produces a significant blurring effect while removing noise and fails to recover the detailed texture information of the image. A weakness can be seen in

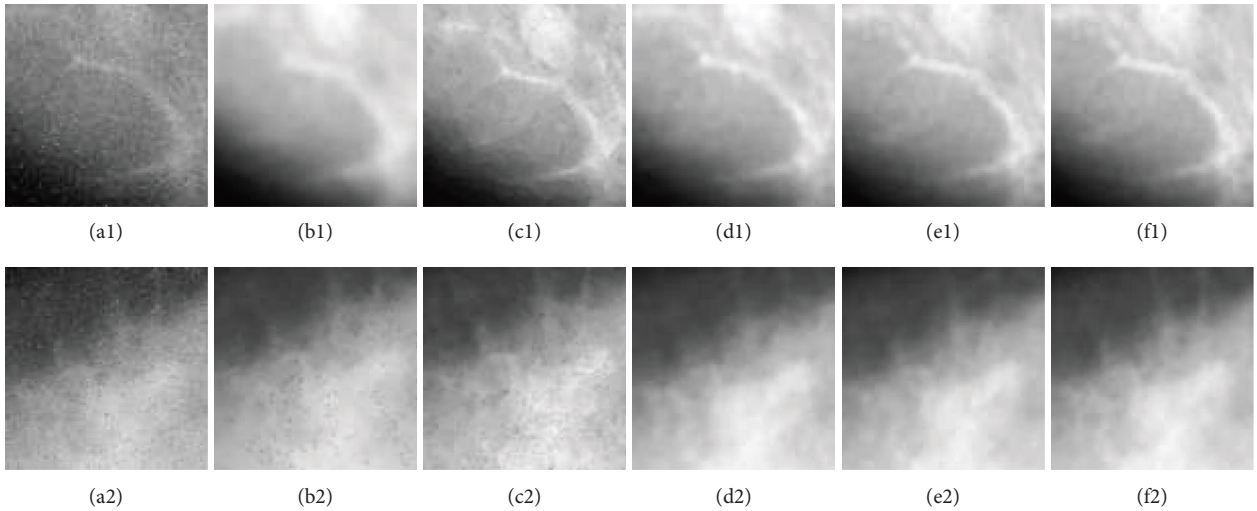


Figure 6: Local detail enlarged image.

the harmonious interplay between noise removal and preservation of edge features and texture information. Literature [15] is a highly competitive technique that has shown higher performance in both noise removal and texture information preservation. The Gibbs effect, which cannot be removed, will be produced by BM3D since it executes the filtering operation in the transform domain, which is comparable to windowing the image signal. After denoising, the Gibbs effect will produce pseudo-textures that resemble scratches. The clinical diagnosis might be significantly impacted by pseudo-texture. Literature [16] is a relatively novel method at present, and the pyramid convolution method expands the perceptual field without increasing the number of parameters, and thus more image details can be obtained. Figure d shows the denoising results of the method in the literature [16] for mixed noise, which not only removes the noise in the image but also retains the image detail information. The proposed method obtains a better denoising effect than literature [16], and the image detail texture is clearer. Literature [17] is a U-net based denoising method that has a simple structure and uses a completely different feature fusion approach to stitch the features in channels to form thicker features. Figure e shows the denoising results of the method in the literature [17] for mixed noise, and the noise in the image is removed more cleanly. The proposed denoising method in this paper obtains comparable denoising results to those of the literature [17].

To better show the detailed information about image denoising, we zoom in on the areas marked by red arrows in Figures 4 and 5, as shown in Figure 6. Figures 6(a1)–6(f1) correspond to the part pointed by the red arrow in Figure 4, respectively.

Figures 6(a2)–6(f2) correspond to the part pointed by the red arrow in Figure 6, respectively. From Figures 6(b1), 6(b2), 6(c1), and 6(c2), it can be found that the literature [14, 15] cannot produce a good denoising effect on the mixed noise in X-ray breast images, and the denoised images are blurred, and there are artifacts. Figures 6(d1), 6(d2), 6(e1), and 6(e2) show the enlarged details after denoising the noisy images in literature [16, 17]. Besides, these two methods can remove the noise in X-ray breast images and retain the details better. Figures 6(f1) and 6(f2) show the detailed information of the X-ray breast image denoised by the proposed method. It can be seen that the denoising effect of the proposed method on X-ray breast images is slightly better than that of the literature [16, 17]. By showing the enlarged details, it can be seen that the proposed method has a better denoising effect on X-ray breast images, compared with the current novel denoising methods.

5. Conclusions

In this paper, the image denoising method is described. In view of the shortcomings of the current traditional denoising methods, a convolution autoencoder neural network denoising model for medical images is proposed by using the powerful extraction ability of the convolution neural network for image information and the learning ability of the autoencoder network structure. A solution to mixed noise in medical images is presented. The proposed method has the following advantages: (1) Denoising effect is obvious, in the removal of mixed noise in medical images, while retaining the details of the image texture information; (2) model denoising is fast and can complete a large number of medical image denoising tasks in a short time, which has a clinical application value; (3) end-to-end modeling is realized, and the denoised image can be obtained by inputting the noise image without manually adjusting the parameters. In addition, the experiment in this paper is limited by the memory of graphic processing unit. In the future, we can consider increasing the memory of the graphic processing unit and further optimize the model from the total amount of sample data and mini-batch.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest with any financial organizations regarding the material reported in this manuscript.

Acknowledgments

This work was supported by Natural Science Foundation of Heilongjiang Province of China under grant number LH2021F039.

References

[1] S. V. Mohd Sagheer and S. N. George, "A review on medical image denoising algorithms," *Biomedical Signal Processing and Control*, vol. 61, p. 102036, 2020.

[2] K. A. S. H. Kulathilake, N. A. Abdullah, A. Q. M. Sabri, and K. W. Lai, "A review on Deep Learning approaches for low-dose Computed Tomography restoration," *Complex & Intelligent Systems*, pp. 1–33, 2021.

[3] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, Mar 2007.

[4] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, Nov 2003.

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision (ICCV)*, Kyoto, Japan, September 2010.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.

[7] G. Wang, "A perspective on deep imaging," *IEEE Access*, vol. 4, pp. 8914–8924, 2016.

[8] H. M. Shan, Y. Zhang, Q. Yang et al., "Correction for "3D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2D trained network" [jun 18 1522-1534]," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, p. 2750, Dec 2018.

[9] Y. Lei, J. Zhang, and H. Shan, "Strided self-supervised low-dose ct denoising for lung nodule classification," *Phenomics*, vol. 1, 2021.

[10] F. Shi, N. Cai, Y. Gu et al., "DeSpecNet: a CNN-based method for speckle reduction in retinal optical coherence tomography images," *Physics in Medicine and Biology*, vol. 64, no. 17, p. 175010, 2019.

[11] B. Srinivas and G. Sasibhushana Rao, "A novel DeepCNN model for denoising analysis of MRI brain tumour images," *International Journal of Intelligent Information and Database Systems*, vol. 13, no. 2/3/4, pp. 393–410, 2020.

[12] P. C. Tripathi and S. Bag, "CNN-DMRI: a convolutional neural network for denoising of magnetic resonance images," *Pattern Recognition Letters*, vol. 135, pp. 57–63, Jul 2020.

[13] L. Bo, K. Xu, D. Feng, H. Mi, and Z. Jian, "Denoising convolutional auto-encoder based B-mode ultrasound tongue image feature extraction," in *Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

[14] G. I. Ri, S. J. Kim, and M. S. Kim, "Improved BM3D method with modified block-matching and multi-scaled images," *Multimedia Tools and Applications*, vol. 81, no. 9, pp. 12661–12679, Apr 2022.

[15] S. L. Guo, G. W. Wang, L. N. Han, X. W. Song, and W. T. Yang, "COVID-19 CT image denoising algorithm based on adaptive threshold and optimized weighted median filter," *Biomedical Signal Processing and Control*, vol. 75, p. 103552, 2022.

[16] X. L. Jia, Y. L. Peng, J. Li, Y. H. Xin, B. Ge, and S. G. Liu, "Pyramid dilated convolutional neural network for image denoising," *Journal of Electronic Imaging*, vol. 31, no. 02, Article ID 023024, 2022.

[17] Z. Z. Huang, J. P. Zhang, Y. Zhang, and H. M. Shan, "DU-GAN: generative adversarial networks with dual-domain U-Net-Based discriminators for low-dose CT denoising," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[18] A. E. Ilesanmi and T. O. Ilesanmi, "Methods for image denoising using convolutional neural network: a review," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2179–2198, 2021.

[19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct 1986.

[20] 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, Aug 2013.

[21] J. Masci, U. Meier, D. Ciresan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International Conference on Artificial Neural Networks*, pp. 52–59, Springer, Berlin, Heidelberg, 2011.

[22] D. Kingma and J. J. C. S. Ba, *Adam: A Method for Stochastic Optimization*, Computer Science, San Diego, CA, USA, 2014.

[23] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning JMLR.org*, pmlr, New York, NY, USA, June 2015.

[24] K. Simonyan and A. J. C. S. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[25] W. Zhou and A. C. J. I. S. P. M. Bovik, "Mean squared error: love it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[26] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proceedings of the 20th International Conference on Pattern Recognition, ICPR 2010*, pp. 23–26, Istanbul, Turkey, August 2010.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 1752390, 15 pages
https://doi.org/10.1155/2022/1752390

Research Article

Breast Ultrasound Image Segmentation Algorithm Using Adaptive Region Growing and Variation Level Sets

Yanwei Wang,[1] Junbo Ye,[1] Tianxiang Wang,[1] Jingyu Liu ,[2] Hao Dong ,[2] and Xin Qiao[3]

[1]*School of Mechanical Engineer, Heilongjiang University of Science & Technology, Harbin, Heilongjiang, China*
[2]*School of School of Computer Science and Information Engineering, Harbin Normal University, Harbin, Heilongjiang, China*
[3]*Functional Inspection Department, Harbin Maternal and Child Health Care and Family Planning Service Center, Harbin, Heilongjiang, China*

Correspondence should be addressed to Jingyu Liu; rgzn2020@hrbnu.edu.cn

Received 5 March 2022; Revised 23 August 2022; Accepted 2 September 2022; Published 3 October 2022

Academic Editor: Xiaofeng Li

To address the features of strong noise, blurred boundaries, and poor imaging quality in breast ultrasound images, we propose a method for segmenting breast ultrasound images using adaptive region growing and variation level sets. First, this method builds a template layer from the difference between the marked image and the original image. Second, the Otsu algorithm is used to measure the target and background using the maximum class variance method to set the threshold. Finally, through the level set of the pixel neighborhood, the boundary points of the adaptive region growth are specified by the level set of the pixel neighborhood, and it is therefore possible to accurately determine the contour perimeter and area of the lesion region. The results demonstrate that the value of Jaccard and Dice for benign tumors is greater than 0.99. Therefore, the segmentation effect of breast images can be achieved by utilizing a breast ultrasound image segmentation approach that uses adaptive region growth and variation level sets.

1. Introduction

Medical image segmentation is a key technology in medical image processing, and its task is to extract interesting objects from medical images to serve the clinician's computer-aided diagnosis. In recent years, breast ultrasound image segmentation has become a challenging task because of the simultaneous consideration of speckle noise, ultrasound image quality, and the location and size of breast lesions [1]. Several auxiliary systems assist doctors in effectively detecting some of the conditions in breast ultrasound images, via image processing and pattern recognition algorithms [2]. Segmentation of breast ultrasound images is critical to systematically localize the specific locations of breast lesions; however, speckle noise, image quality, and location and size of breast lesions are the limitations of this method.

There are several automatic, semi-automatic, and manual breast ultrasound image segmentation methods available in modern society [3–5]. Histogram threshold is a common method in breast ultrasound image segmentation. In this method, a histogram threshold is set in breast ultrasound image segmentation and is widely used in grayscale image segmentation. Another method is the active contour model, which is a framework for delineating object contours from background images. It is an edge-based segmentation method, in which active contour models are widely used in breast ultrasound images [6], and it also applies in enhancement [7]. A hybrid scheme combining region-based and boundary-based techniques for breast ultrasound image segmentation was developed using these methods, in which seed points are automatically generated by an empirical rule-based formula, and boundary points are thereafter determined by region growing (RG) and directional gradient operation, which combines the clustering algorithm used in computed tomography detection [8]. Compared to RG, adaptive region growing (ARG) is more effective in segmentation [9], improved genetic algorithms [10], enhanced seeded region growing [11], automated inspection [12], magnetic resonance imaging [12], and other types of

imaging [13], and it also combines clustering [14], statistical characters, and point cloud images [15]. In addition, Markov random field models have been used for breast ultrasound image segmentation. It provides efficient use of pixel correlations, and the segmentation results are further improved by applying a maximum posterior segmentation prediction scheme. However, it has a complex iterative process and is time consuming. A neural network is also used in breast ultrasound image segmentation to simplify processing. This transforms the segmentation problem into a classification decision based on a set of features. The neural-network-based image segmentation method is based on an adaptive reference point classification algorithm, and the extraction of tumor images is performed based on a cost function, which is defined according to the boundary and region information of the tumor in the frequency and spatial domains. Neural network-based methods are typically used when the lesions are compact and round, and the algorithms do not perform well in other cases [16–18]. They also combine other methods to improve accuracy [19]. Subsequently, the RG method was applied to mammary image segmentation to improve the degree of automation.

Medical image segmentation is a challenging task because of the low contrast between the region of interest and other textures. Vakanski et al.[20] proposed an algorithm based on convolutional neural networks (CNN) for medical image segmentation, which included U-Net and full convolutional neural networks. To further improve the accuracy and precision of detection, an image segmentation method combining level set systems was used as a piece of picture division [21]. In this method, the region growth combined level set was applied in mammary image segmentation, and it improved the segmentation accuracy. To improve the robustness and accuracy, Regonda et al. [22] performed a machine learning approach for medical image segmentation by employing a CNN. At this time, an adaptive segmentation algorithm combined with machine learning was developed. To improve the accuracy, optimized methods are used in image segmentation and image classification. Ahmad et al. [23] proposed a probability-based optimal deep learning feature fusion applied in ultrasound image segmentation, which can detect tumors in ultrasound images automatically. Subsequently, Shwetha et al. [24] proposed a design and analysis algorithm with high precision for breast tumor segmentation in mammograms and ultrasound images and compared the features of breast ultrasound. Praveen and Rajendra Prasad [25] used multi-ROI segmentation for effective texture features of mammogram images, which is effective for texture features, and Shrivastava and Bharti [26] suggested density applied in breast tumor detection for further research. To further improve the intelligence level of segmentation, in 2020, Lei et al. [27] proposed an automatic breast anatomy segmentation method for automated whole breast ultrasound images, but experiments have proved that it is currently difficult to effectively address the breast anatomy segmentation problem using this method. To address this problem, an ARG algorithm based on optimization is applied in medical image segmentation.

In conclusion, it is known from the above literature that in RG, the edges are optimized using methods such as deep learning, CNN, and level sets. Considering the boundary of the ending condition, the level set method (LSM) is used to track various interfaces and shapes. The strength of the level set method (LMD) is that it automatically handles topological changes in the tracked shape because it uses an implicit representation to represent active contours. Active contour model (ACM) is a competitive tool in image segmentation, and an active contour approach driven by adaptive local prefitting energy function based on Jeffreys divergence (APFJD) applied in image segmentation is effective while handling images with uneven intensity. [28] The LMD was developed by two mathematicians, Stanley Osher and James Sethian, in the 1980s. This method has been widely used in many disciplines and fields, and subsequently, a large amount of level set data has been developed, which makes the level set, and the application of sets in computing has become easier. It has already been applied in brain image segmentation and breast thermograms and is also used in data processing. The variation level set segmentation method, which is advantageous for dealing with a variety of topologies, is an important branch of medical applications. It is essential to transform the higher dimensional space into a lower dimensional space to describe the evolution and achieve medical image segmentation. The GMAC model was used in the variation-level set [29]. Later, to improve the accuracy of segmentation, it was combined with statistics [30].

In summary, in the field of medical image segmentation, a combination of methods is often used to complement each other's advantages. Owing to the complexity of the situation, the automatic breast ultrasound image segmentation method cannot accurately segment the breast ultrasound image, whereas the LSM can automatically process the topological changes of the tracking shape; thus, adaptive region growth and variation-based methods are adopted. The level set breast ultrasound image segmentation method makes the segmentation of breast ultrasound images more accurate.

2. Related Work

To improve image segmentation accuracy, Koshki et al. [31] proposed a two-step method to segment breast thermal images and obtained relatively accurate results. Niaz et al. [32] proposed an active contour method based on an improved combined local and global fitting function to address the breast tumor segmentation problem, and it was experimentally confirmed that the proposed method exhibited better performance when compared to previous state-of-the-art methods. Gonalves et al. [33] proposed a deep learning algorithm that used the learning of the probability map of key points in the loss function as a regularization term for robust learning of key point localization and analyzed a deep image segmentation and deep key point-based algorithm. A new algorithm was proposed for detecting model interactions that improve state-of-the-art performance and execution time for breast key point detection tasks. Liao et al.

[34] combined ultrasound image diagnosis and deep learning technology to identify benign and malignant tumors. The experimental results show that the accuracy of the region segmentation algorithm based on supervised blocks is the same as that of manual segmentation. Ho et al. [35] presented a deep multimagnification network trained by partial annotation for automated multiclass tissue segmentation using a set of patches from multiple magnifications in digitized whole slide images. They proposed architecture with multiencoder, multidecoder, and multiconcatenation that outperformed other single- and multimagnification-based architectures by achieving the highest mean intersection-over-union and could be used to facilitate pathologists' assessments of breast cancer. Huang et al. [36] proposed a new segmentation method by dense prediction and local fusion of superpixels for breast anatomy with scarce labeled data and achieved good results. Wang et al. [37] proposed a novel CNN with coarse-to-fine feature fusion to address the poor quality of breast ultrasound images and large variations in breast lesion size, shape, and location. Zhang et al. [38] proposed a multitask learning model based on a soft and hard attention mechanism for the simultaneous segmentation and binary classification of breast ultrasound images, and the experimental results demonstrated the effectiveness of the model. Salama and Aly [39] proposed a new framework for segmentation and classification of breast cancer images. Different models, including InceptionV3, DenseNet121, ResNet50, VGG16, and MobileNetV2, were applied to classify the mammographic image analysis society (MIAS), digital database for screening mammography (DDSM), and the curated breast imaging subset of DDSM (CBIS-DDSM) into benign and malignant, and the accuracy of breast cancer diagnosis results has been greatly improved. Hassanzadeh et al. [40] leveraged the potentials of dense and residual blocks, used the capability of evolutionary computation, and proposed an automatic evolutionary model to detect an optimum and accurate network structure and its parameters for medical image segmentation. Subsequently, Wang et al. [41] applied an iterative edge attention network in image segmentation with EA-net, whereas Han et al. [42] used a CNN for image segmentation with U-NeXt. There is a new method for the segmentation of medical images that has achieved good results.

Recently, in terms of mammary image segmentation, the accuracy of the algorithm was improved by combining region growing with other methods, which is crucial for the effective classification of mammary images. In this study, based on the segmentation algorithm combining RG and adaptive level set, the boundary points of the ARG are defined through the level set of the pixel neighborhood, so that the contour perimeter and area of the lesion region could be effectively identified. The Jaccard of Koshki et al. [31] is 0.9658, and the run time is 2.6 seconds, whereas the Dice of Wang et al. [37] is 0.8142, and that of Zhang et al. [38] is 0.8152. Our work is to calculate the value of Jaccard and Dice used in an ARG algorithm for level set optimization, and the results exhibited better Jaccard and Dice. The run time was fast, which is suitable for the segmentation effect of breast images.

3. Methodology

The region-growing method involves gathering the region formed by the pixel points based on their similar properties. This method starts with a single pixel and gradually merges to form the desired region. Initially, the seed pixel is selected from the seed point inside the lesion by considering it as the starting point, that is, the starting point of growth.

The qualified pixels in the domain are merged with the seed pixels to form a new seed pixel, and the merge operation in mathematics is used until it changes to the edge line and the unqualified points outside the edge line. Therefore, in regional growth, the selection of seed pixels, determination of growth rules, and termination conditions are the three key factors.

3.1. Regional Growth Principle. Regional growth utilizes the similarity between image pixels. It is a gray-value-based processing method that divides an image into similar regions according to specified criteria, as shown in Algorithm 1.

This is a process of region growth, that is, region growth is a process of aggregating into larger regions according to subregions under the above conditions, as shown in Figure 1. In the 4-neighbor connected pixels of the seed point pixel label 1, namely points 2, 3, 4, and 5, respectively. The gray value 12 of pixel point 4 is the closest to the gray value 10 of the seed point; thus, pixel point 4 is added to the segmentation area, and pixel 4 is used as a new seed point to perform the following process. During the second cycle, in the image to be analyzed, namely 2, 3, 5, 6, 7, and 8, respectively, the gray value of pixel 6 is 16, and the gray mean value of the already segmented area, comprising 1 and 4, is 10.5, pixel 6 gets added to the segmented area. The last image depicts the direction in which the region grows, from shallow to deep.

3.2. Steps of Regional Growth

3.2.1. Based on Regional Grayscale Difference

① Scan the pixels to find pixels that have not yet been attributed

② Check its neighborhood pixels by considering a pixel as the center, that is, compare the pixels in the neighborhood with that at the center one by one, and merge them if the grayscale difference is less than a predetermined threshold

③ Assume the newly merged pixel as the center, return to step ②, and check the neighborhood of the new pixel until the area cannot be expanded further

④ Return to step ①, continue to scan until all pixels are assigned, and end the entire growth process

Assume R represents the whole image, then the segmentation can be seen as the process of dividing the region R into n sub-regions: $R_1, R_2, \ldots, R_n$
And $\cap_1^n R_i = R$, R_i is a connected area, $i = 1, 2, 3, \ldots, n$
if $R_i \cap R_j$ is empty,
for any i, j; $i \neq j$;
If $P(Ri)$ = TURE,
for $i = 1, 2, 3, \ldots, n$;
then $P(R_i \cup R_j)$ = FALSE

Algorithm 1: Region growing.

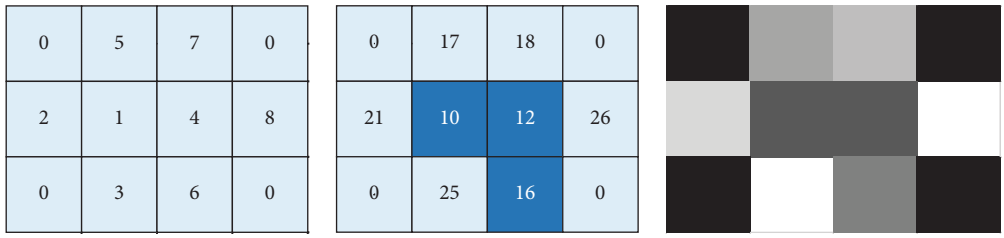


Figure 1: Region growing method: (a) pixel label, (b) gray value, and (c) corresponding image.

3.2.2. Based on the Statistical Properties of Grayscale Distribution in the Region

① Divide the pixels into small nonoverlapping areas

② Compare the cumulative grayscale histograms of adjacent regions and merge regions based on the similarity of the grayscale distribution

③ Set the termination criteria and merge the regions by repeating the operations in step ② until the termination criteria are met

3.2.3. Based on the Shape of the Area. Fixed grayscale is used to divide the image into sections. Set the perimeters P_1 and P_2 of the two adjacent areas and set the part where the grayscale difference between the two sides of the common boundary line of the two areas is less than the given value as L, if (T_1 is a predetermined threshold)

$$\frac{L}{\min\{P_1, P_2\}} > T_1. \quad (1)$$

Following the merging of the two regions as shown in (1), let B be the common boundary length of the two adjacent areas, and the image is divided into various areas with fixed grayscale, and the length of the part, where the grayscale difference between the two sides of the common boundary line of the two areas, is set, which is less than the given value, which is set to L, and T_2 is a predetermined threshold value.

$$\frac{L}{B} > T_2. \quad (2)$$

In summary, the first method merges the regions with lower contrast in the common boundary of two adjacent regions, and the second method merges the lower contrast portion of the common boundary of the two adjacent regions [43].

3.3. Adaptive Region Growth Algorithm. While processing the breast images, it was found that the determination of growth seed points, conditions for regional growth, and conditions for terminating regional growth are the key factors for regional growth. The similarity between growth points and similar regions can be judged based on the gray value, texture, and image information, such as color. Therefore, the number, position, and gray value of seeds in the RG algorithm are also important factors that affect the results of RG. The ARG algorithm means that not only the position of the seed needs to be selected but also the gray value of the starting point of growth needs to be selected. ARG uses the grayscale difference as the criterion for regional merging and then performs the regional merging method according to the similarity of the grayscale distribution in a small area, equalizes the grayscale histogram, and finally divides and merges to determine the finish condition.

In the region growing method, when a long point is selected, the growth is performed according to the similarity criterion of the gray histogram. Usually, the boundary between the discontinuous regions of the gray level, threshold processing boundary of the pixel property distribution, and segmentation of the direct search region are considered. Adaptive region growth is the process of aggregating pixels or subregions into larger regions according to predefined criteria, starting from a set of growth points, which can be

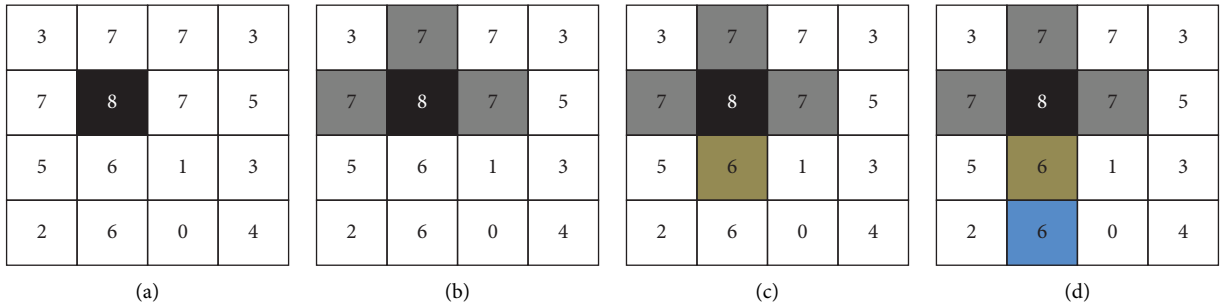


Figure 2: Schematic of iterative region growth results: (a) original image, (b) first RG result, (c) second RG result, and (d) third RG result.

either a single pixel or a certain growth point. A small area, adjacent pixels, or areas with properties similar to the growth point are merged with it. These new pixels are regarded as new growth points, and the process is repeated continuously until no pixels that meet the conditions can be included. This implies that the growth points can no longer grow because the growth of this region has been completed.

In other words, the conditions for RG are similarity criteria defined according to the continuity between pixel gray levels, and the conditions for region growth stop define a termination rule, conditions, and regional growth stops. In the algorithm, we define a variable, the maximum pixel gray-value distance, reg_maxdist. When the absolute value of the difference between the gray value of the pixel to be added and the average gray value of all pixels in the segmented area is less than or equal to reg maxdist, the pixel is added to the segmented area; in contrast, and the area increases and the algorithm ends.

As depicted in Figure 2, Figure 2(a) is the original image, and the numbers represent the grayscale of the pixels. Assuming a pixel with a gray value of eight as the initial growth point, denoted as $f(i, j)$, the circle represents the seed point. In the 8-neighborhood, the growth criterion is that the difference between the gray value of the point to be measured and the gray value of the growing point is one or zero. Figure 2(b) shows that after the first region growth, the difference between $f(i-1, j)$, $f(i, j-1)$, $f(i, j+1)$ and the gray value of the growth point is one, thus merged. Figure 2(c) shows that after the second growth, $f(i+1, j)$ is merged. Figure 2(d) shows that after the third growth, $f(i+1, j-1)$ and $f(i+2, j)$ are merged. Thus far, there are no pixels that meet the growth criteria, and the growth stops.

Therefore, the steps for realizing adaptive region growth are as follows: the processing process is depicted in Figure 3.

Step 1: Scan the image sequentially to find the first pixel that has not yet been attributed and set the pixel as $(x0, y0)$.

Step 2: By assuming $(x0, y0)$ as the center, consider four-neighborhood pixels (x, y) of $(x0, y0)$. If $(x0, y0)$ satisfies the growth criterion, merge (x, y) with $(x0, y0)$. In the same area, push (x, y) onto the stack at the same time.

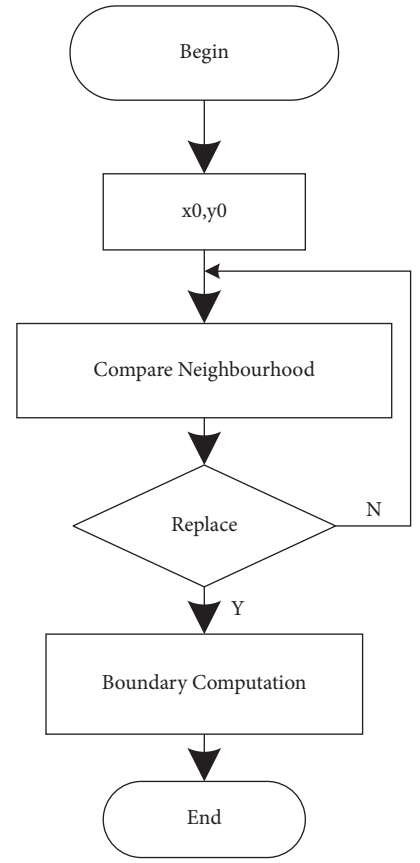


Figure 3: ARG flow chart.

Step 3: Take a pixel from the stack, treat it as $(x0, y0)$, and return to Step 2.

Step 4: When the stack is empty, go back to step 1.

Step 5: Repeat steps 1–4 until every point in the image has attribution and the growth is over.

In a word, the region growing method is simple to calculate and has a good effect on segmenting connected regions with the same features. However, due to the problems of noise and uneven greyscale, it is easy to generate holes and oversegmentation. Therefore, threshold method and region method are used for medical image segmentation. It is usually combined with other segmentation methods.

4. Optimization of Variation Level Set Segmentation

4.1. Variable Differential Principle. Assume that $E(u)$ represents two-dimensional energy, as shown in

$$E(u) = \iint F\left(x, y, u, u_x, u_y\right)\mathrm{d}x\mathrm{d}y. \tag{3}$$

We derived the Equation (3) to obtain the extreme value of energy, which is the minimum value, and applied the Euler-Lagrange Equation to obtain the basic equation for the level set, as shown in

$$\frac{\partial F}{\partial u} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial u_x}\right) - \frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{\partial F}{\partial u_x}\right) = 0. \tag{4}$$

In the nonlinear partial differential equation (PDE) composed of Euler-Lagrange, to address the extreme value of the minimum energy, it is necessary to refer to the time variable, which also transforms the static nonlinear PDE problem into a dynamic PDE problem and realizes the steady evolution of the solution [44].

The effect of the variable differential set treatment of the breast ultrasound image is depicted in Figure 4, in which, we can see that the ultrasound image shows the presence of noise. The image is not high, the edge is unclear, and the image is directly used. The edges of the identified lesion area are relatively broad.

4.2. Level Set Segmentation Principle. Assume that C is a closed curve, which can be expressed using the following equation:

$$C = \{(x, y) \mid u(x, y) = c\}. \tag{5}$$

In (5), we introduced a time variable, t in the solution of the original PDE, and the time function $u(x, y)$ is the level set, which is shown in

$$C(t) = \{(x, y) \mid u(x, y, t) = c\}. \tag{6}$$

We also introduced time variables separately while constructing original Equation (4). The derivative of Equation (4) is shown in

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \nabla u \cdot \left(\frac{\partial x}{\partial y}, \frac{\partial y}{\partial t}\right) = 0. \tag{7}$$

Assume the partial differential components of x, y, and partial derivatives concerning the time variable: the partial derivative of the curve corresponds to the time variable, and $\partial c/\partial t$ represents the variation in curve C.

Similarly, the form in which the total derivative of $u(t)$ can be expressed as $-\nabla u \cdot V$, here, we assume $\beta(F)$ as an energy functional, and the total derivative of $u(t)$ transforms to the following equation:

$$\beta(F) = N \cdot V, \tag{8}$$

$$\frac{\partial u}{\partial t} = \beta(F)|\nabla u|. \tag{9}$$

It can be observed from the above equation that F is a velocity function. (9) is a developed form of the level set PDE equation, which belongs to a class of solutions that apply the Hamilton-Jacobi equation to the variation differential problem, which requires the introduction of time variables to find the solution of the equation.

At this time, the solution of the equation does not affect the value of parameter C. Therefore, C is assigned zero. The solution of curve C can be transformed into a solution for the level set of the embedded function. Similarly, the evolution of the curve can be transformed into the evolution of the level set equation of the embedded function $u(x, y)$ under a given initial value $u_0(x, y)$.

In the evolution process of the initial curve C, the introduced time variable can determine the current curve $C(t)$ by taking the two-dimensional level set of $u(x, y, t)$ at any time, to obtain the pole that makes the energy function go. The solution of the Euler-Lagrange equation of small values terminates the evolution of curve C and finds a solution to [45]

$$C(0)\{(x, y) \mid u(x, y, 0) = u_0(x, y)\}. \tag{10}$$

In summary, the numerical calculation in this LSM may be unstable, or the solution process tends to be complicated owing to the improper selection of the embedded function $u(x, y)$. Therefore, the most commonly used embedding function is the signed distance function $-d((x, y), C)$, which indicates that it is outside curve C, and vice versa. Among them, the Euclidean distance represents the distance between point (x, y) and curve C, and its main advantage is that it is equal to one, and the rate of change of the function is uniform, which is conducive to the stability of the numerical calculation [46].

4.3. Variation Level Set Segmentation. When the level set indicates that curve C is a closed curve, the minimization of the equation of motion of the curve can be applied to the minimum energy of the capability functional function that should close the curve. (11) expresses the energy function of the Geodesic Active Contour model for level set image segmentation.

$$E(C) = \int_0^1 g(|\nabla[C(s)]|)\mathrm{d}s. \tag{11}$$

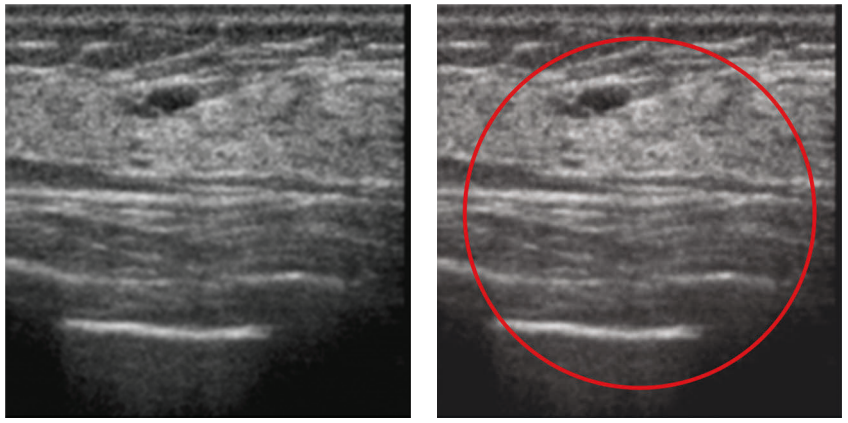
Considering the initial value and assuming that the curve parameters $C(0)$ and $C(1)$ are equal, (11) can be modified to

$$E(C(p)) = \int_0^1 g(C(s))C_p\mathrm{d}s. \tag{12}$$

Applying the LMD to (12), the PDE of the corresponding embedding function is shown in

$$\frac{\partial u}{\partial t} = [g(|\nabla I \times [C(s)]|)k - \nabla g \cdot N]|\nabla u|. \tag{13}$$

The Heaviside function is introduced in (13), and using Green's Equation, the energy function can be expressed in



(a) (b)

Figure 4: Result of variable let segmentation. (a) Original image and (b) after CV filtering.

$$E(u) = \iint_{\Omega} g(x, y)|\delta(u)\mathrm{d}x\mathrm{d}y|. \tag{14}$$

The energy differentia $\delta(u)$ is equal to $\mathrm{d}H(u)/\mathrm{d}u$, $H(u)$ is the Heaviside function by assuming that the energy u is larger than zero, and the Heaviside function value is equal to one. Using the variation method, the level can be set based on the type of the evolution.

$$\frac{\partial u}{\partial t} = \partial(u)\mathrm{div}\left(g\frac{\nabla u}{|\nabla u|}\right). \tag{15}$$

The regularization $\partial_{\varepsilon}(u)$ is the approximation in

$$\frac{\partial u}{\partial t} = \partial_{\varepsilon}(u)\mathrm{div}\left(g\frac{\nabla u}{|\nabla u|}\right). \tag{16}$$

Overall, the minimum solution of the closed curve C was converted into a solution of the minimum value of the energy functional. By introducing the embedded function, Heaviside function, and Green Equation to transform the functional, the solution of $E(C)$ is transformed into the solution problem of $E(u)$ and then the PDE is obtained by the LSM, and the steady-state solution is finally obtained by numerical calculation, which is the variation LSM [47].

In the process of curve evolution, the level set segmentation method must address the PDE for each iterative calculation and loop until convergence. The results are depicted in Figure 5. In addition, the energy function in the segmentation model of the LSM may be nonconvex, and the curve evolution is very sensitive to the initial contour, which is prone to local minimum results, further resulting in segmentation failure.

The GAC model was introduced as an active contour model to address this problem. Compared to the region-based model using level set evolution, the numerical iteration of its PDE is time consuming; therefore, the segmentation speed is slow, which is not ideal. However, the orientation of the evolution model based on gradient information in segmentation and the existence of a stable minimum solution make it easy to solve.

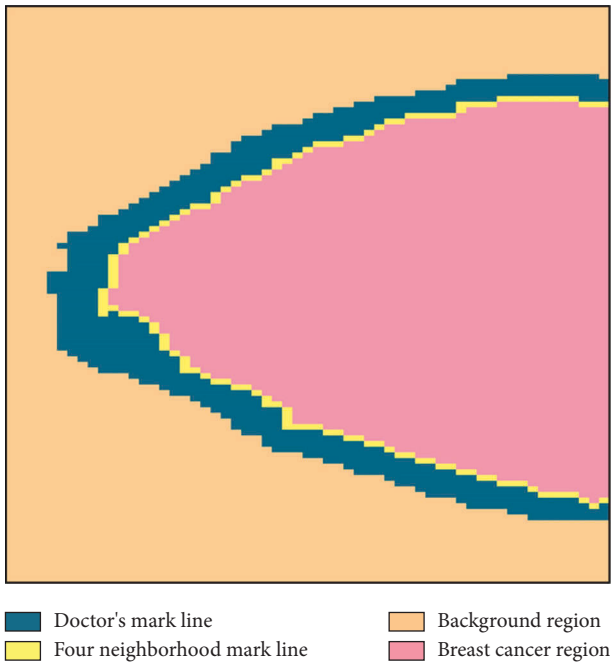


Figure 5: Level set optimization boundary.

To address the boundary problem in the traditional LSM in the past, especially in the local minimum energy function, which cannot achieve segmentation results, we have proposed a global variation segmentation method. In this method, the image area information is used to build a posterior probability-based energy model, followed by a very small variation in the overall framework. The combination of the GAC model uses gradient information to establish a global variation energy model and finally constructs a new integrated image segmentation model.

4.4. Variation Level Set Model. The perimeter of the shadow part was calculated, and the pixel neighborhood method was used to obtain the boundary points. We used the level set

optimization to test using templates a and b. According to the data in Figure 6, we can see that initially a template layer is developed, and the marked image and original image are acquired. Thereafter, the Otsu algorithm is used to measure the target and background using the maximum class variance method to set the threshold.

5. Breast Ultrasound Image Segmentation Algorithm

5.1. Differential Image Acquisition. Read the "marked image" and "unmarked image," because the "marked image" is marked in red. The Red channel is referred as the R channel, and the difference between the two images "−diff" (difference value) is obtained.

5.2. Image Segmentation. Using the OUST algorithm to binarize the diff image [48], a black-and-white image was obtained in the previous step.

The regional growth method was used to grow from two origins in the upper right corner and the lower left corner, respectively. The growth rules are as follows:

(1) Horizontal growth begins from the first pixel of each row

(2) Vertical growth starts from the first pixel in each column

(3) Stop the growth if it encounters a marked boundary value

(4) Stop all growth after all growth outside the marked area is complete

The area of the shadowed part was calculated. After executing the above three steps, only the pixel value of the shadow part was zero, all the points with a pixel value of zero in the image were traversed, and the total number was counted, which is the area of the shadow part.

In the neighborhood of templates a and b, if only one neighborhood is a boundary point, this point can be considered as a boundary point. Taking the four neighborhoods as an example, as depicted in Figure 6, all black points are traversed, and when there is a white point in the four neighborhoods of the black point, the point is determined as a boundary point and marked as orange. The number of orange points was counted to obtain the perimeter of the marked area.

6. Experimental Results and Analysis

6.1. Experimental Data and Platform. In the experiment, three datasets from various sources were employed. The first dataset(data set (1) was generated from the mini-MIAS mammography database, which divided the data into benign and malignant categories based on tumor severity. The second dataset(data set (2) was gathered from public dataset Dataset_BUSI_with_GT, which was collected and released. It includes thousands of breast ultrasound images, of which 780 were chosen as the experimental dataset. The third dataset (data set (3) was a two-dimensional image of over 200 breast instances acquired over 6 months using Toshiba's BSM31 equipment. Staff screening resulted in the identification of nine photographs for the study. The data set 3 is manually labeled into three sections: benign, malignant, and normal.

The collected images were run on an AMD Ryzen 74800H with 8 cores and 16 threads at 4.2 G Hz, the GPU was RTX 2060 6G, the operating system was Windows 11, the memory is 16G, the hard disk is 512G, using Python version 3.8 programming, the compiler uses VS Code, and the experimental results are obtained.

Using adaptive area production and variation level set optimization, the relative perimeter and area of the diseased area in breast images were obtained, and its shape was determined. Finally, the optimization results of four-neighborhood a and eight-neighborhood b in adaptive region growing were compared, along with the relative perimeter difference, perimeter-area ratio, and other parameters.

6.2. Evaluation. In this study, the Jaccard similarity coefficient (Jaccard) and Sørensen–Dice coefficient (Dice) are introduced to calculate the effectiveness of breast image segmentation. Jaccard was used to compare the similarities and differences between the limited sample sets. The larger the Jaccard coefficient value, the more similar the samples, and the higher the degree. The Jaccard was calculated as shown in Equation (17). The Jaccard similarity coefficient was defined.

$$\operatorname{acc}(x, y) = \frac{|x \cap y|}{|x| + |y| - |x \cap y|}. \tag{17}$$

When x equals y, the Jaccard coefficient is one. When A and B do not intersect, the Jaccard coefficient is zero, and the Jaccard distance represents the dissimilarity of the sample or set. The larger the Jaccard distance, the lower the sample similarity. Therefore, the Jaccard distance was used to describe the dissimilarity. However, the disadvantage is that it is only applicable to sets of binary data.

The Dice distance is mainly used to calculate the similarity between two sets (it can also measure the similarity of the strings). The Dice similarity coefficient of the two sets of x and y is defined as: where $|x|$ represents the cardinality of the set x (that is, the number of elements in the set) and $|y|$ represents the cardinality of the set y.

$$\operatorname{Dice}(x, y) = \frac{2 * |x \cap y|}{|x| + |y|}. \tag{18}$$

Jaccard and Dice calculations were performed on 18 sample images collected at the hospital. The calculation results are depicted in Figure 7. It can be observed from the figure that the two parameters of the six sets of data are both below 0.99, which can be judged as irregular images and malignant tumors.

Comparing the images of the public dataset, the calculation results of Jaccard and Dice are shown in Figures 8-9. It can be seen from the figure that 11 pairs of two parameters

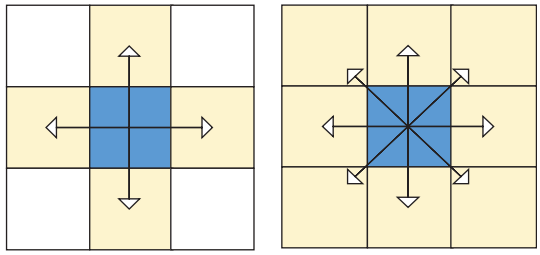


Figure 6: boundary model size. (a) Model a. (b) Model b.

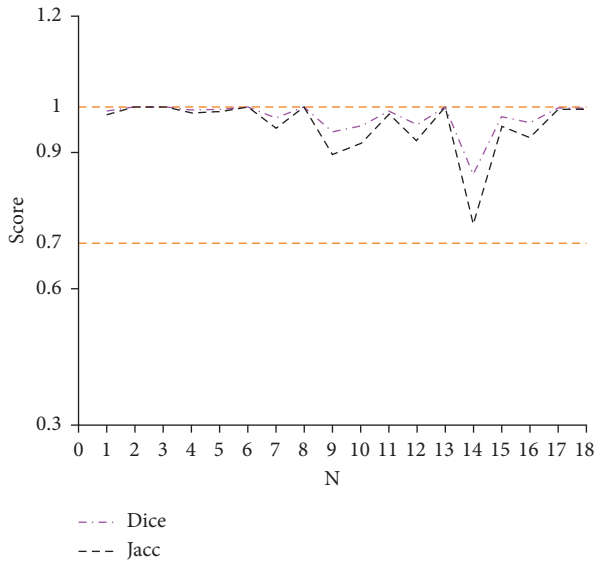


Figure 7: Perimeter comparison.

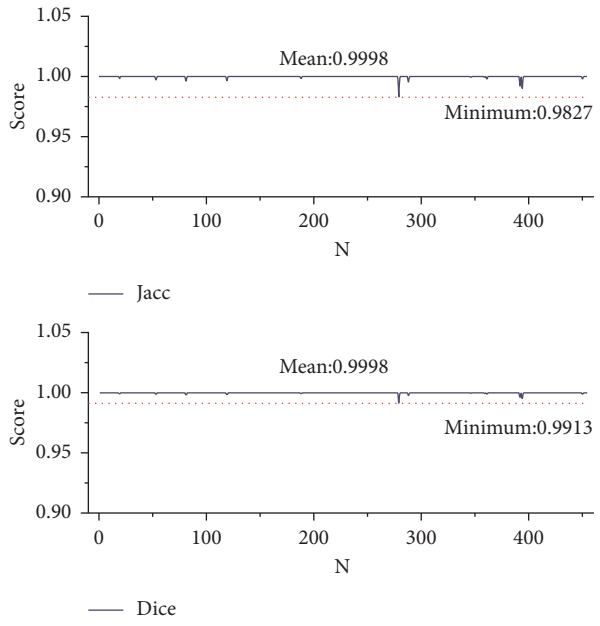


Figure 8: Calculation results of Jaccard and Dice from data set 1.

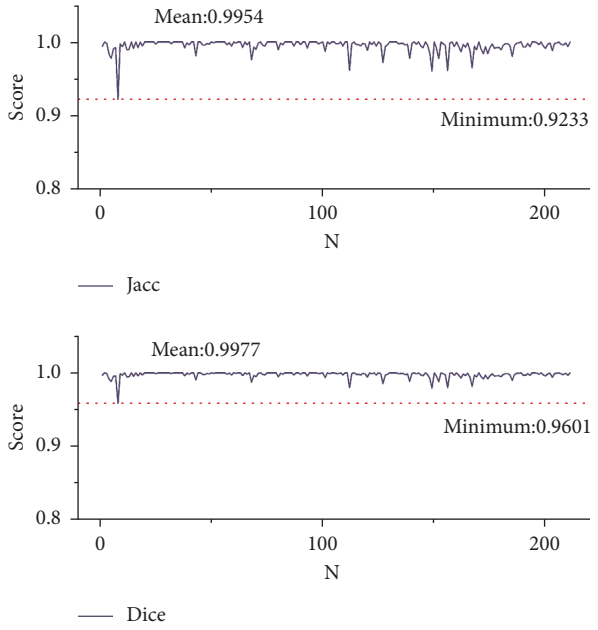


Figure 9: Calculation results of Jaccard and Dice from data set 2.

are below 0.99, which are images of the algorithm misjudged, and the misjudgment rate is 2.422%. The running time of each image on the public dataset is shown in Figure 10. In which, data set 1 is a subset of benign tumors in the public dataset and data set 2 is a subset of malignant tumors in the public dataset.

6.3. Results and Discussion. According to the comparison of different public data sets, nine sets of ultrasound images were simultaneously selected for testing at the same time. The outline of the lesion is represented using red, and the long and short diameters are represented using green lines, as shown in Figure 11.

In the experiment, a third group of double lesions was used as an example. To demonstrate the effectiveness of the processing results, the small lesions in the third group were selected and processed by adaptive region growth and variation level set optimization.

The degree histogram is shown on the right in Figures 11(a)–11(c) respectively give the local effect maps of

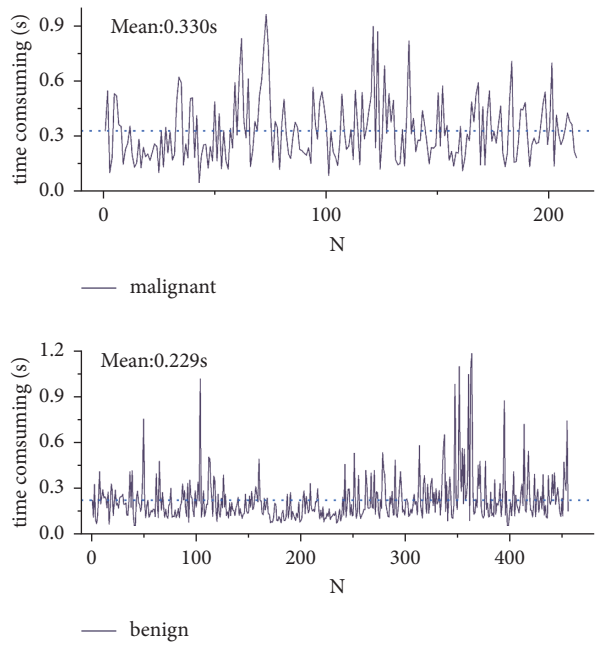


Figure 10: Run time results of data set 1 and data set 2.

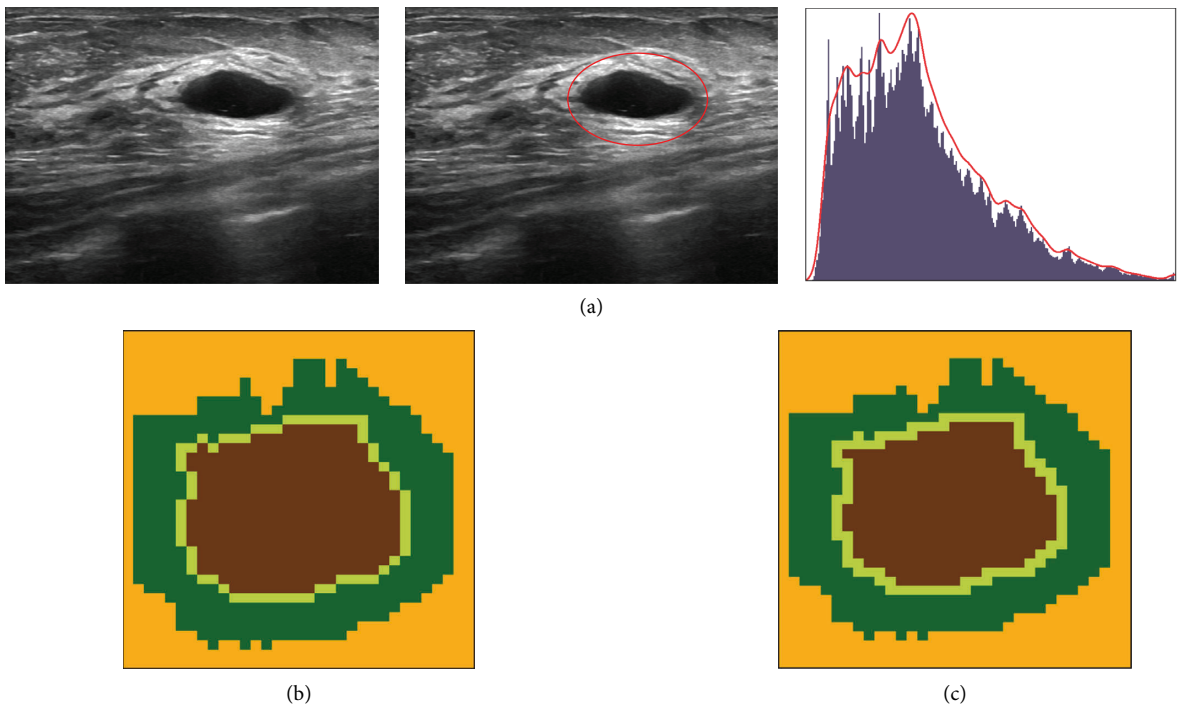


Figure 11: Test3 breast ultrasound image: (a, b) model a processing and (c) model b processing.

the neighborhoods processed by two different templates, (b) is the boundary termination point of the growth of the four-neighborhood optimization area of template a, (b) the pixel distribution on the left, and the right schematic diagram of the lesion, (c) is the boundary termination point for the growth of the eight-neighborhood optimization region of template b, (c) the pixel distribution on the left, and the schematic diagram of the lesion on the right.

From the comparison of the results in Figures 11(b) and 11(c), the boundary point in Figure 10(b) is a closed curve with strong connectivity. In Figure 11(c), the lesion point is mistakenly judged as the boundary point. Figure 11(b) has a better effect.

The results of the adaptive region growth optimized by the variable differential level set are applied to other lesions, and the images of the left lesions in the third

Table 1: Result of area and perimeter in data set 3.

Name	Region	Area	Model a perimeter	Model b perimeter
test1	Left	4904	240	341
	Right	5661	262	373
test2	Whole	143	55	76
test3	Left	3573	248	355
	Right	369	72	91
test4	Whole	1017	136	189
test5	Whole	672	169	211
test6	Whole	974	177	228
test7	Whole	1295	185	224
test8	Whole	3101	224	286
test9	Whole	1643	196	259

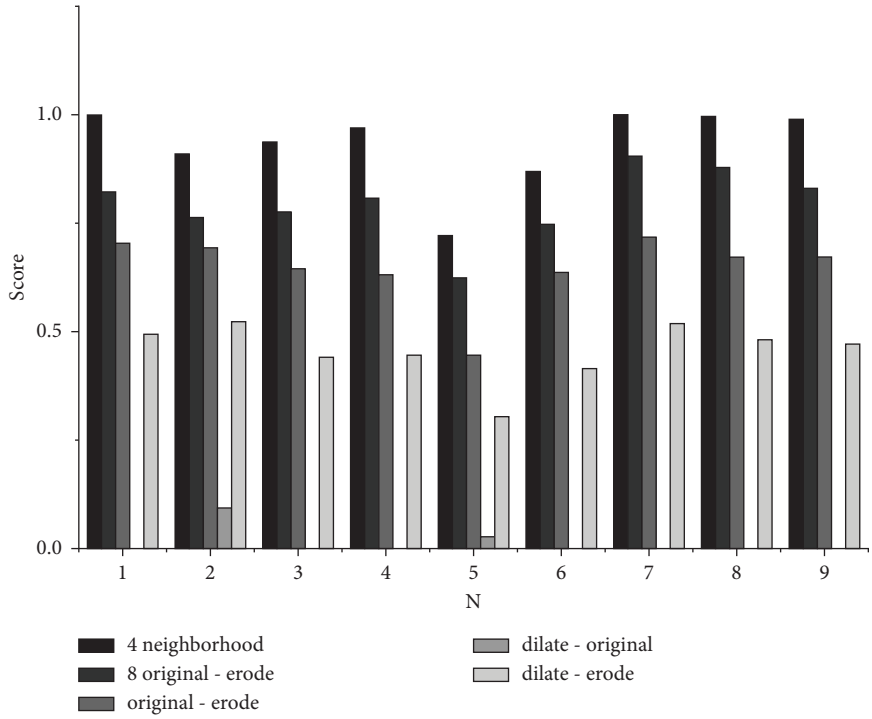


Figure 12: Dice comparison histogram in data set 3.

group are processed to obtain the image shown in Figure 11(a), and then the edges are corrected to obtain Figure 11(b).

At the same time, the perimeter and area of the nine groups were calculated according to the method of modifying the adaptive region growth by variable differential. The calculation results are listed in Table 1. This lays the foundation for the automatic labeling of lesions and identification of lesion morphology.

We evaluated five ways to find the circumference of breast cancer on the first dataset using Jaccard and Dice scored on data set 1 and dataset set 2. The five methods are 4-neighborhood, 8-neighborhood, original image minus erode image, dilate image minus original image, and dilate image minus erode image. The Dice and Jaccard scores for the various ways are shown in Figures 12 and 13.

For data set 3, we simultaneously compared the perimeters computed for the two different templates in Section 6.2, resulting in the image depicted in Figure 14. The comparison results show that the perimeter data calculated by template a is closer to the actual disease image. In addition, the comparison data of the area and perimeter of the application of template a and the method described in this study were also investigated. From Figure 14, it is evident that as the perimeter length increased, the accuracy of the detection results decreased.

According to Table 2, the run results showed a total of 24 breast tumors, which are malignant tumor data in public dataset 2, those images are Jaccard values below 0.99, or Dice values below 0.98. No is the image number, and value is the score with Dice or Jaccard. Combined with a comparison of clinical observations, the tumors in the 24 images detected above were all malignant tumors.

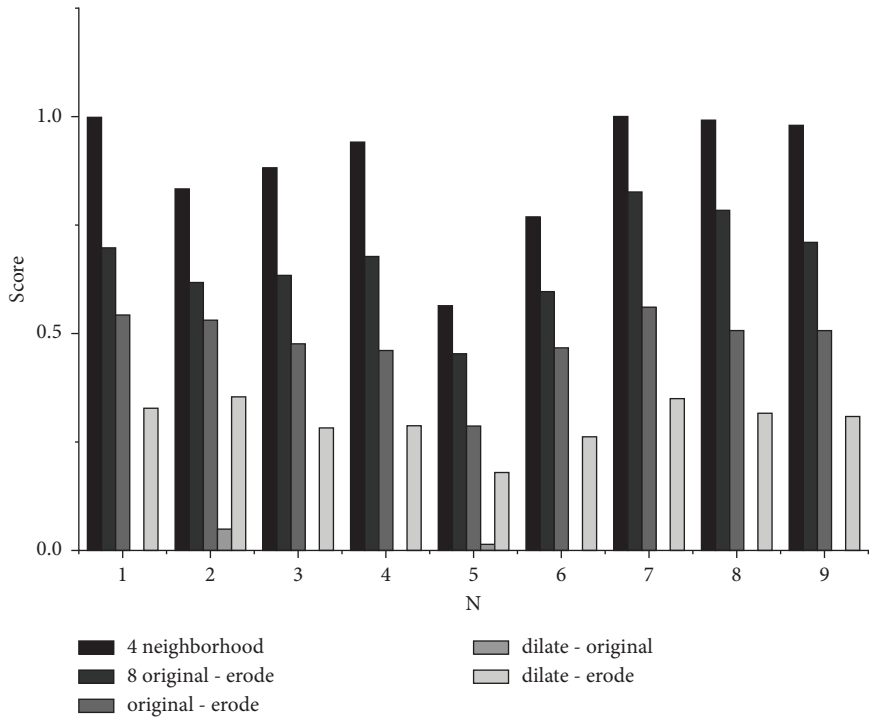


Figure 13: Jaccard comparison histogram in data set 3.

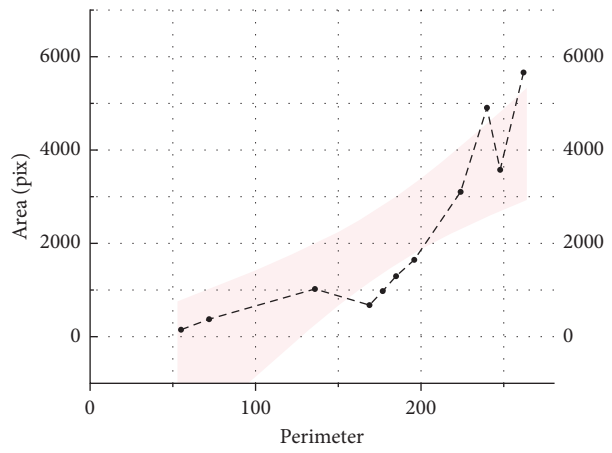


Figure 14: Perimeter comparison in data set 3.

Figure 15 depicts 13 images with Jaccard and Dice below 0.99 in the malignant tumor database. It can be seen from the figure that breast malignant tumors present an irregular shape, and the average Jaccard value of the public dataset is 0.9932.

The experimental results demonstrated that the clinical images and public dataset images were processed in this study, and the boundary points of the adaptive region growth were defined through the level set of the pixel neighborhood to effectively identify the contour perimeter and area of the lesion area. The results demonstrated that in the proposed method for breast ultrasound image segmentation based on adaptive region growing and changing level sets, the Jaccard and Dice values of the processed benign tumors were both greater than 0.99; therefore, breast ultrasound image segmentation using adaptive region growing and changing level sets is adopted. This method is suitable for the segmentation of mammary gland images and can address the problem of image classification in the case of large noise, blurred boundaries, and poor imaging quality of mammary gland images.

According to the data in Table 3, it shows that there are different methods of breast image segmentation indicators. Through the investigation of Dice and Jaccard, Jaccard and running time were compared to level set model for multiclass [31], and Dice was compared to Coarse-to-Fine Fusion CNN [37] and soft and hard

Table 2: Dice or Jaccard evaluation value of less than 0.99 for malignant tumors in data set 2.

No.	4	5	8	12	43	68	80	101	112	120	127	139
Dice	0.991741	0.988884	0.960118	0.994826	0.990654	0.987858	0.994806	0.993732	0.980671	0.994378	0.986136	0.988848
Jaccard	0.983618	0.978013	0.923295	0.989706	0.981481	0.976007	0.989666	0.987542	0.962074	0.988819	0.972651	0.977941
No.	148	149	152	156	162	166	167	169	172	174	185	203
Jaccard	0.992388	0.980077	0.988719	0.980632	0.99343	0.994326	0.982379	0.994867	0.992215	0.992126	0.990241	0.994195
Dice	0.98489	0.960932	0.97769	0.962	0.986945	0.988715	0.965368	0.989787	0.984549	0.984375	0.98067	0.988458

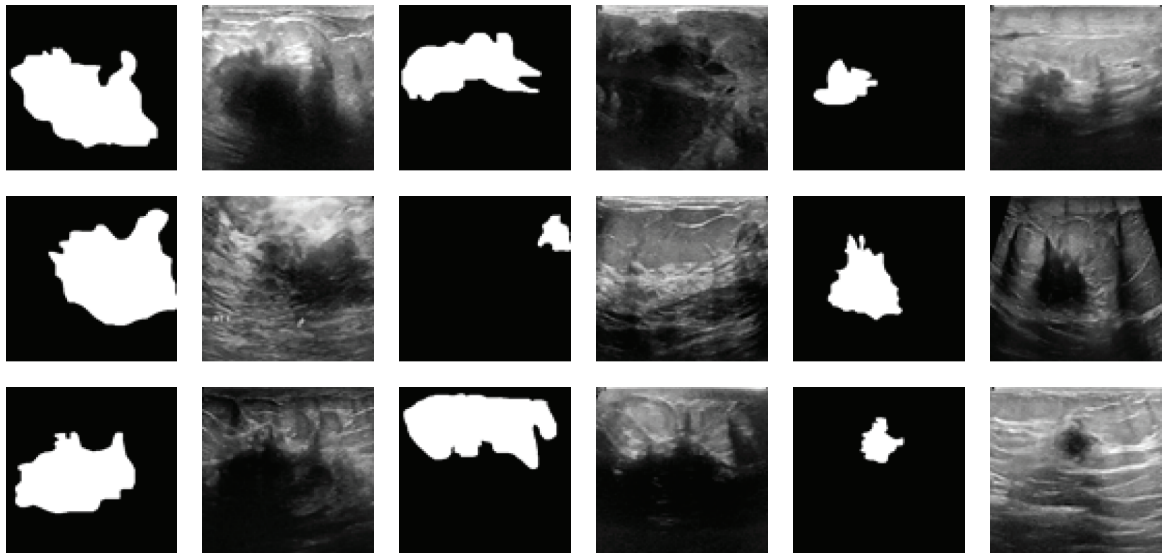


Figure 15: Correspondence map with a Dice or Jaccard evaluation value of less than 0.99 for malignant tumors in data set 2.

Table 3: Comparison of different methods of breast image segmentation indicators.

Method	Average Dice (%)	Average Jaccard (%)	Average run time (%)
Soft and hard attention multitask learning [38]	81.42	—	—
Level set model for multiclass [31]	—	96.58	2.60
Coarse-to-fine fusion CNN [37]	86.52	—	—
Our method for malignant public data subsets	99.99	99.97	0.23
Our method for benign public data subsets	99.77	99.54	0.19

attention multitask learning [38], respectively. By testing the surface of the public data set, the value of the method explained in this paper is better than that listed in level set model for multiclass [31] on Jaccard, which reduces the running time; when compared to the Dice index, the value of the Dice method in this paper is higher than that mentioned in Coarse-to-Fine Fusion CNN [37] and soft and hard attention multitask learning [38].

7. Conclusion

In this study, breast image segmentation based on global adaptive region growing and variation-level sets were combined for blurred and noisy edges of lesions in breast ultrasound images. By establishing the difference model between the breast ultrasound image and the labeled image, setting the R channel, selecting two seed points through the upper right and lower left, growing along the horizontal and vertical directions simultaneously, and then setting the neighborhood template and variation-level set optimization. When compared to the adaptive region growing method [13], this study optimizes the cut-off adjustment through variation level sets, compares the localized boundaries of the four-neighborhood template a and the eight-neighborhood template b, and provides the discrimination of the neighborhood template a. Result of level set optimization adaptive region growing algorithm is better. When compared to the level set model for multiclass [31], this study optimizes Jaccard; the value is 0.9999 in malignant and 0.9954 in benign, both of which are better than 0.9658. At the same time, the runtime is shorter, 0.233 s and 0.185 s in this study, which is shorter than 2.6 s. Considering the Dice, by comparing soft and hard attention multitask learning [38] and Coarse-to-Fine Fusion CNN [37], the scores are 0.9999 and 0.9977, respectively, both of which are better than 0.8142 and 0.8652.

In the future, we will obtain the segmentation effect of breast images and extract the perimeter and area of the lesion image, thereby laying a foundation for the accuracy of computer-aided diagnosis and lesion identification.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of Heilongjiang Province of China under grant no. LH2021F039.

References

[1] Y. Guo, A. Şengür, and J. W. Tian, "A novel breast ultrasound image segmentation algorithm based on neutrosophic similarity score and level set," *Computer Methods and Programs in Biomedicine Biomedical Research and Medical Practice*, vol. 123, pp. 43–53, 2016.

[2] J. L. Ma, L. Y. Bao, Q. Lou, and D. Kong, "Transfer learning for automatic joint segmentation of thyroid and breast lesions from ultrasound images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 17, no. 2, pp. 363–372, 2021.

[3] Y. Hu, Y. Guo, Y. Wang et al., "Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model," *Medical Physics*, vol. 46, no. 1, pp. 215–228, 2019.

[4] M. Xian, Y. Zhang, H. D. Cheng, F. Xu, B. Zhang, and J. Ding, "Automatic breast ultrasound image segmentation: a survey," *Pattern Recognition*, vol. 79, pp. 340–355, 2018.

[5] M. Xu, K. Huang, Q. Chen, and X. Qi, "Mssa-net: multi-scale self-attention network for breast ultrasound image segmentation," in *Proceedings of the 2021 IEEE 18th international symposium on biomedical imaging (ISBI)*, pp. 3193–3196, Nice, France, April 2021.

[6] W. D. Wang, H. Sheng, Y. L. Chen et al., "A fast, edge-preserving, distance-regularized model with bilateral filtering for oil spill segmentation of SAR images," *Journal of Oceanology and Limnology*, vol. 39, no. 4, pp. 1198–1210, 2021.

[7] N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," *Medical Physics*, vol. 26, no. 8, pp. 1642–1654, 1999.

[8] Y. Peng, S. Zhao, and J. Liu, "Segmentation of overlapping grape clusters based on the depth region growing method," *Electronics*, vol. 10, no. 22, p. 2813, 2021.

[9] R. Bhakkiyalakshmi, P. Ponnammal, and M. K. Srilekha, "Fast and adaptive detection of pulmonary nodules in thoracic CT images using a contextual clustering based region growing," *International Journal of Engineering & Technology*, vol. 7, no. 2.24, pp. 106–108, 2018.

[10] C. G. Lin, "A novel automated inspection approach based on adaptive region-growing image segmentation," *Journal of the Chinese Society of Mechanical Engineers, Series C: Transactions of the Chinese Society of Mechanical Engineers*, vol. 35, no. 1, pp. 57–65, 2014.

[11] C. Y. Ma, X. J. Ma, Y. G. Xu, L. Xiang, and K. Zhang, "Enhanced seeded region growing algorithm and its application in signal decomposition," *Measurement Science and Technology*, vol. 33, no. 9, Article ID 095111, 2022.

[12] E. A. Carlsen, C. B. Johnbeck, M. Loft et al., "Semiautomatic tumor delineation for evaluation of 64Cu-dotatate PET/CT in patients with neuroendocrine neoplasms: prognostication based on lowest lesion uptake and total tumor volume," *Journal of Nuclear Medicine*, vol. 62, no. 11, pp. 1564–1570, 2021.

[13] N. K. Babu, "Adaptive region growing image segmentation algorithms for breast MRI," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 8729–8732, 2019.

[14] X. L. Li and J. S. Chen, "Region adaptive adjustment strategy based on information entropy for remote sensing image segmentation," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-4-2021, pp. 69–74, 2021.

[15] F. Poux, C. Mattes, Z. Selman, and L. Kobbelt, "Automatic region-growing system for the segmentation of large point clouds," *Automation in Construction*, vol. 138, no. 6, Article ID 104250, 2022.

[16] A. R. Reddy and P. R. Reddy, "Chenna Reddy MRI brain tumor segmentation and prediction using modified region growing and adaptive SVM," *Soft Computing*, vol. 11, no. 1, pp. 1–14, 2021.

[17] Y. S. Li, Y. Liu, L. J. Huang, Z. L. Wang, and J. W. Luo, "Deep weakly-supervised breast tumor segmentation in ultrasound images with explicit anatomical constraints," *Medical Image Analysis*, vol. 76, Article ID 102315, 2022.

[18] Y. Liu, Y. Chen, B. Han, Y. Zhang, X. Zhang, and Y. Su, "Fully automatic Breast ultrasound image segmentation based on fuzzy cellular automata framework," *Biomedical Signal Processing and Control*, vol. 40, pp. 433–442, 2018.

[19] R. B. Fair, "Artificial neural network approach for multiphase segmentation of battery electrode nano-CT images," *Npj Computational Materials*, vol. 8, no. 1, pp. 1–11, 2022.

[20] A. Vakanski, M. Xian, and P. E. Freer, "Attention-enriched deep learning model for breast tumor segmentation in ultrasound images," *Ultrasound in Medicine and Biology*, vol. 46, no. 10, pp. 2819–2833, 2020.

[21] M. Taheri, M. Rastgarpour, and A. Koochari, "A novel method for medical image segmentation based on convolutional neural networks with SGD optimization," *Journal of Electrical and Computer Engineering Innovations*, vol. 9, no. 1, pp. 37–46, 2021.

[22] N. Regonda, S. K. Yadav, and M. Reddy, "Medical image segmentation using level set based active contour method," *International Journal for recent developments in science & technology*, vol. 1, no. 2, pp. 81–89, 2018.

[23] F. Ahmad, E. Noor, and I. Shafaf, "Amplification of pixels in medical image data for segmentation via deep learning object-oriented approach," *International Journal of Advanced Technology and Engineering Exploration*, vol. 8, no. 74, pp. 2394–5443, 2020.

[24] S. V. Shwetha, L. Dharmanna, B. S. Anami, and M. Rafi, "Design and analysis of an algorithm for breast tumor segmentation in mammogram and ultrasound images," *International Journal of Computer Theory and Engineering*, vol. 13, no. 4, pp. 108–117, 2021.

[25] C. Praveen Kumar and K. Rajendra Prasad, "Multi-ROI segmentation for effective texture features of mammogram

images," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 24, no. 8, pp. 2461–2469, 2021.

[26] N. Shrivastava and J. Bharti, "Breast tumor detection in MR images based on density," *International Journal of Image and Graphics*, vol. 22, 2022.

[27] B. Lei, S. Huang, H. Li et al., "Self-Co-Attention neural network for anatomy segmentation in whole breast ultrasound," *Medical Image Analysis*, vol. 64, pp. 101753–101832, 2020.

[28] P. Q. Ge, Y. Chen, G. N. Wang, and G. Weng, "An active contour model driven by adaptive local pre-fitting energy function based on Jeffreys divergence for image segmentation," *Expert Systems with Applications*, vol. 210, no. 12, Article ID 118493, 2022.

[29] D. P. Lestari, S. Madenda, E. Ernastuti, and E. P. Wibowo, "Comparison of three segmentation methods for breast ultrasound images based on level set and morphological operations," *International Journal of Electrical and Computer Engineering*, vol. 7, no. 1, pp. 383–391, 2017.

[30] Y. Chen, C. Liu, J. Wan, S. Zhao, X. Zhang, and Y. Lan, "Breast mass segmentation in mammographic images based on gmac model using variational level set method," *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia*, vol. 39, no. 1, pp. 35–40, 2016.

[31] A. Shamsi Koshki, M. Zekri, M. R. Ahmadzadeh, S. Sadri, and E. Mahmoudzadeh, "Extending contour level set model for multi-class image segmentation with Application to Breast Thermography Images," *Infrared Physics & Technology*, vol. 105, Article ID 103174, 2020.

[32] A. Niaz, A. A. Memon, K. Rana et al., "Inhomogeneous image segmentation using hybrid active contours model with application to breast tumor detection," *IEEE Access*, vol. 8, pp. 186851–186861, 2020.

[33] T. Gonalves, W. Silva, and M. J. Cardoso, "Deep image segmentation for breast keypoint detection," *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 54, no. 1, pp. 35–39, 2020.

[34] W. X. Liao, P. He, J. Hao et al., "Automatic identification of breast ultrasound image based on supervised block-based region segmentation algorithm and features combination migration deep learning model," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 984–993, 2020.

[35] D. J. Ho, D. V. Yarlagadda, T. M. D'Alfonso et al., "Deep multi-magnification networks for multi-class breast cancer image segmentation," *Computerized Medical Imaging and Graphics*, vol. 88, no. 7, Article ID 101866, 2021.

[36] Q. Huang, Z. Miao, S. Zhou, C. Chang, and X. Li, "Dense prediction and local fusion of superpixels: a framework for breast anatomy segmentation in ultrasound image with scarce data," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–8, 2021.

[37] K. Wang, S. Liang, S. Zhong, Q. Feng, Z. Ning, and Y. Zhang, "Breast ultrasound image segmentation: a coarse-to-fine fusion convolutional neural network," *Medical Physics*, vol. 48, no. 8, pp. 4262–4278, 2021.

[38] G. Zhang, K. Zhao, Y. Hong, X. Qiu, K. Zhang, and B. Wei, "SHA-MTL: soft and hard attention multi-task learning for automated breast cancer ultrasound image segmentation and classification," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 10, pp. 1719–1725, 2021.

[39] W. M. Salama and M. H. Aly, "Deep learning in mammography images segmentation and classification: automated CNN approach," *Alexandria Engineering Journal*, vol. 60, no. 5, pp. 4701–4709, 2021.

[40] T. Hassanzadeh, D. Essam, and R. Sarker, "An evolutionary DenseRes deep convolutional neural network for medical image segmentation," *IEEE Access*, vol. 8, pp. 212298–212314, 2020.

[41] K. Wang, X. Zhang, X. Zhang, Y. Lu, S. Huang, and D. Yang, "EANet: iterative edge attention network for medical image segmentation," *Pattern Recognition*, vol. 127, Article ID 108636, 2022.

[42] Z. Han, M. Jian, and G. G. Wang, "ConvUNeXt: an efficient convolution neural network for medical image segmentation," *Knowledge-Based Systems*, vol. 253, Article ID 109512, 2022.

[43] S. Y. Lu, H. Huang, P. Liang, G. Chen, and L. Xiao, "Hepatic vessel segmentation using variational level set combined with non-local robust statistics," *Magnetic Resonance Imaging*, vol. 36, pp. 180–186, 2017.

[44] D. Courault, L. Hossard, V. Demarez et al., "STICS crop model and Sentinel-2 images for monitoring rice growth and yield in the Camargue region," *Agronomy for Sustainable Development*, vol. 41, no. 4, pp. 49–213, 2021.

[45] Y. W. Wang, Y. M. Pan, K. J. A. Feng, and Y. K. Feng, "Medical image segmentation based on global variational level set," in *Proceedings of the 2011 International Conference on Computer Science and Information Technology (ICCSIT 2011)*, pp. 28–33, Chengdu, China, June 2011.

[46] C. Wang, W. L. Wang, S. C. Pan, and F. Y. Zhao, "A local curvature based adaptive particle level set method," *Journal of Scientific Computing*, vol. 91, no. 1, p. 3, 2022.

[47] S. Feng, "Effective document image binarization via a convex variational level set model," *Applied Mathematics and Computation*, vol. 419, Article ID 126861, 2022.

[48] S. J. Lee, Y. Chung, C. Kim, R. Shrestha, and W. Kim, "Thermographic inspection of CLP defects on the subsurface based on binary image," *International Journal of Precision Engineering and Manufacturing*, vol. 23, no. 3, pp. 269–279, 2022.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5690879, 9 pages
https://doi.org/10.1155/2022/5690879

Research Article

Measurement Model for Medical Image Feature Matrix Similarity Based on CNN

Lili Wang

School of Information Engineering, Harbin University, Harbin 150076, China

Correspondence should be addressed to Lili Wang; wanglili2007@hrbu.edu.cn

Received 2 June 2022; Revised 18 July 2022; Accepted 20 August 2022; Published 5 September 2022

Academic Editor: Xiaofeng Li

The original similarity measurement model is easy to ignore the processing of image details, resulting in poor accuracy of similarity measurement. In the paper, we propose a similarity measurement model for the medical image feature matrix based on the convolutional neural network (CNN). First, the Gaussian convolution kernel is used to obtain the global and local feature data of medical images, and the corresponding data set is formed. Second, the convolution layer of CNN is introduced, and the image feature matrix is obtained by the convolution layer. Finally, the similarity measurement model of the medical image feature matrix is constructed. The results show that the image similarity measurement effect of this model is better when the test process is divided into three parts: global, local, and detail. The highest error rate of the proposed algorithm is only about 0.21, which takes less time, and the overall fitting degree can reach about 91%. Compared with traditional methods, the accuracy of image similarity measurement is higher and the use effect is better.

1. Introduction

With the continuous development of visualization technology and computer graphics, the research on the medical image is also deepening. The basis of medical image research is feature analysis. Therefore, at present, taking medical images as the research object, the integration of research on the extraction method of medical image feature points has become a hot research direction [1]. At the same time, there have been many studies on the similarity measurement of image features, which provide more useful information for the in-depth analysis and application of medical images. Many achievements have been made in medical image feature analysis and similarity measurement at home and abroad. Abdar et al. [2] proposed a medical image classification and fusion model, which mainly uses the binary residual feature fusion method to solve the uncertainty of the algorithm. This method makes a detailed analysis of the medical image features and obtains a good image classification effect. However, this method takes a long time; Shu et al. [3] proposed an initial spatial convolution block to collect the context information related to the decoding stage, extract and accumulate features from different paths and establish associations between structural features and semantic features for medical image segmentation. However, the practical applicability of this method is relatively weak. Poudel and Lee [4] established a multi-scale global feature map and used the attention mechanism to suppress noise and bad features, resulting in the complete recovery of context feature dependence, and completed the feature extraction of medical images, which has a high recall and precision. However, it takes a long time. Wenping et al. [5] proposed an improved random walk node similarity measurement method: a random walk similarity measurement method based on relative entropy. The transition probability distribution of the node is constructed according to the transition probability of the node arriving at the influential node in the network after the multi-step random walk. The relative entropy of the transition probability distribution of two nodes is calculated to obtain the difference score between node pairs in the network. Then the similarity matrix between network nodes is obtained. RE-model algorithm performs well in symmetry, network propagation, and community discovery, but the measurement effect is not ideal. Feng, et al. [6] established the isometric isomorphism model

of multi-scale spatial entities from the viewpoint of metric geometry. The geometric and topological similarity measurement of graphics is expressed in multiple scales for planar and linear spatial entities with different complexities. This method can simultaneously take into account the changes in geometry and topology of multi-scale spatial entities and conform to the multi-scale abstraction law of spatial entities, but the similarity measurement takes a long time. Wang et al. [7] proposed a spatiotemporal trajectory similarity measurement method using Hausdorff distance. Starting from the three features of spatiotemporal trajectory, a spatiotemporal trajectory reorganization strategy for similarity measurement is proposed. The traditional idea of similarity measurement with a point as the center is changed to track segment as the center, and the track spatiotemporal clustering experiment is carried out with micro-blog check-in track data and taxi GPS track data. This method can effectively calculate the similarity of spatiotemporal trajectories and meet the needs of spatiotemporal trajectories clustering, but the measurement accuracy is not ideal.

The original similarity measurement method is easy to ignore the analysis of image details, resulting in unsatisfactory calculation results. Therefore, this paper proposes a similarity measurement model of the medical image feature matrix based on CNN. Starting from the global and local, the feature points of the medical image are extracted in an orderly, and CNN is introduced to construct the feature matrix, to complete the construction of the medical image feature matrix similarity measurement model. To ensure the use effect of the similarity measurement model based on CNN designed in this study, after the model is built, the test analysis method is used to verify the model in this paper. The contributions of this paper are as follows: (1) Starting from the global and local, the feature points of medical images are extracted orderly, taking into account the details of medical images, which provides the basis for improving the accuracy of the algorithm calculation; (2) CNN is used to construct the similarity measurement model of the feature matrix to improve the performance of the model; (3) Using different data sets, the performance of this model is tested under different indicators to verify the efficiency of this model.

2. Proposed Model

2.1. Acquired Feature Matrix of Medical Images. Image features can be divided into global features and local features. The former mainly reflects the statistical information of the image, while the latter describes the image as a whole based on local structure, texture, and other details [8, 9]. Aiming at the deficiency of the original similarity measurement model, the acquisition of the image feature matrix is optimized to ensure the effectiveness of acquiring image feature information. In the process of image feature acquisition, scale-invariant feature mode is selected to extract the relevant information in a medical image. This method can deal with translation, rotation, scaling, brightness change, and other external factors. It also maintains certain stability for visual changes and applies to the scaling and rotation changes of human organs in the medical image [10].

A medical image equalization algorithm with segmentation factors is introduced. Firstly, the medical image of the input image is divided into two sub-medical images according to the threshold. In this process, the average brightness of the output image is retained, and the disadvantage of brightness drift of the output image is overcome. Suppose the input image is X, the medical image $H(k)$ of the grayscale of the input image k is defined as

$$H(k) = n_k, \text{for } k = 0, 1, \ldots, L-1, \tag{1}$$

where $n(k)$ is the frequency of the k grayscale pixel in the image, and L is the highest grayscale in the image. The probability density function of image $p(k)$ is

$$p(k) = \frac{H(k)}{N}, \text{for } k = 0, 1, \ldots, L. \tag{2}$$

According to the threshold τ, the medical image grayscale $[0, L-1]$ of the input image is divided into two parts: low square and high-quality medical image, which are expressed by $H_{\text{low}}(k)$ and $H_{\text{up}}(k)$. The range of their grayscale is $[0, \tau-1]$ and $[\tau, L-1]$. Probability density function of sub medical image n_{low} and n_{up} are the total number of pixels in low-quality medical image area and high-quality medical image area. The probability density function of a medical image is shown in (3) and (4):

$$P_{\text{low}}(k) = \frac{H_{\text{low}}(k)}{n_{\text{low}}}, \text{for } k = 0, 1, \ldots, \tau-1, \tag{3}$$

$$P_{\text{up}}(k) = \frac{H_{\text{up}}(k)}{n_{\text{up}}}, \text{for } k = \tau, \tau+1, \ldots, L-1, \tag{4}$$

$$\tau = \lfloor \text{median} \cdot r \rfloor. \tag{5}$$

As mentioned in the introduction, the literature [4] method emphasizes the high-frequency pixels of the image, which is easy to cause the loss of details of the output image, and the phenomenon that the low-frequency pixels are swallowed by the adjacent high-frequency pixels. Through equations (6) and (7), the modified medical image equalization algorithm modifies the low-quality medical image and the high-quality medical image and overcomes the problems of the literature [4] method.

$$\text{new\_}p_{\text{low}}(k) = \log\left[p_{\text{low}}(k) + 1\right] \quad k = 0, 1, \ldots, \tau-1, \tag{6}$$

$$\text{new\_}p_{\text{up}}(k) = \log\left[p_{\text{up}}(k) + 1\right] \quad k = \tau, \tau+1, \ldots, L-1. \tag{7}$$

Medical image equalization is the last step to modify the medical image equalization algorithm. The cumulative probability density function for medical image equalization in the literature [4] algorithm is $c(k)$.

$$c(k) = \sum_{i=0}^{k} p(i), \text{for } k = 0, 1, \ldots, L-1. \tag{8}$$

The transformation function $f(k)$ is used to remap the grayscale k of the input image. The definition of the transformation function $f(k)$ is shown in the following equation:

$$f(k) = X_0 + (X_{L-1} - X_0) \cdot c(k), \tag{9}$$

where X_0 and X_{L-1} are the minimum value and the maximum value of the image grayscale. From (9), it can be seen that the grayscale of the input image through the literature [4] method is remapped to the entire dynamic range $[X_0, X_L - 1]$.

Consistent with the literature [4] method, the sub-medical image equalization equation of the modified medical image equalization algorithm is shown in (10). The medical image of the modified grayscale can be regarded as the cropped pixels that are reassigned to the medical image before the medical image is equalized. Finally, the final enhanced image is obtained by equalizing the sub-medical image.

$$f(k) = \begin{cases} X_0 + (X_{\tau-1} - X_0) \cdot \sum_{i=0}^{k} \text{new\_}p_{\text{low}}(i), \text{for } k = 0, 1, \ldots, \tau - 1, \\ X_\tau + (X_{L-1} - X_\tau) \cdot \sum_{i=\tau}^{k} \text{new\_}p_{\text{up}}(i), \text{for } k = \tau, \tau + 1, \ldots, L - 1. \end{cases} \tag{10}$$

The modified medical image equalization algorithm enhances the input image through medical image segmentation, correction, and equalization. Compared with the literature [4] algorithm, the modified medical image equalization algorithm has a higher signal-to-noise ratio on the premise of preserving image details.

The image size of the medical image is set under different sizes, which can be obtained by convolution of the original image and Gauss:

$$A(x, y, z) = G(x, y, z) * p(x, y), \tag{11}$$

where (x, y) is the original position of the image, z is the scale spatial factor, $G(x, y, z)$ is the two-dimensional Gaussian function, p is the original medical image, and A is the spatial scale of the medical image. When acquiring feature points of medical images, the DoG operator [11] is used to process medical images. The operator can be expressed as

$$D(x, y, z) = (G(x, y, iz) - G(x, y, z)) * p(x, y). \tag{12}$$

This operator is used to calculate the points on the image, and the calculated values are connected to obtain the eigenvalue curve of the image [12, 13]. Then, the fitting function is used to accurately obtain the position and scale of image feature points, and unstable feature points are removed, to enhance the stability and antimanufacturing ability of image feature points, as shown in Figure 1.

Following the above process, feature points of medical images are extracted orderly, and the feature points in the images are integrated and stored in the medical image database [14].

2.2. Modeling of Similarity Measurement for the Constructing Medical Image Feature Matrix. According to the medical image features acquired in 2.1, the CNN is used to construct the medical image feature matrix, which is used as the basis of the measurement model.

In this study, CNN is used to classify the extracted feature points. This part is mainly related to the convolution layer of the neural network [15]. Therefore, the image data processing process of the convolution layer in this model is set as follows:

$$f_1(a) * f_2(a) = \int_{-\infty}^{\infty} f_1(a) f_2(b - a)\text{da}, \tag{13}$$

where $f_1(a)$ and $f_2(a)$ refer to the image information to be processed, and $*$ is the convolution calculation symbol. To improve the accuracy of feature classification, a Gaussian convolution [16, 17] check is used to process the image data in this part, and the classification results of feature points are obtained. The processing results are output in the form of a matrix.

Suppose the output feature points matrix is set as a set, then $w = \{w_1, w_2, \ldots, w_n\}$, in which n refers to the number of medical images [18], and w_i refers to the feature matrix corresponding to the medical image i. Then the similarity measurement equation of the image matrix can be obtained by using Mahalanobis distance:

$$d(w_i, w_j)_u = \sqrt{(w_i - w_j)^t u (w_i - w_j)}, \tag{14}$$

where t is the measurement time, w_i, w_j is the corresponding feature vector in the image i , j, $d(w_i, w_j)_u$ is the distance from image w_i to image w_j. The more similar the two are, the smaller the value of $d(w_i, w_j)_u$ is. In this model, u is the similarity measurement matrix of two images [19]. To ensure the validity of its numerical value, it is set as a symmetric positive definite matrix c.

$$u = yy^t, \tag{15}$$

where y is the feature quantity of image classification.

Then the characteristics of the above equation can be reflected as

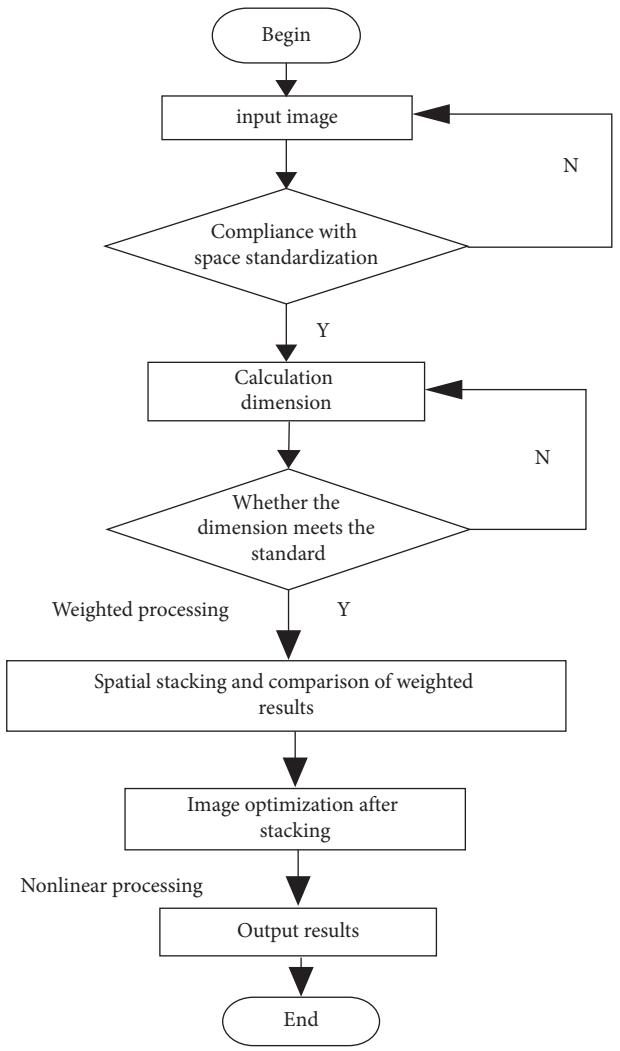


Figure 1: Medical image feature acquisition process.

$$d\left(w_i, w_j\right)_u = \sqrt{\left(w_i - w_j\right)^t y y^t \left(w_i - w_j\right)}. \tag{16}$$

Integrate (14) and (16) to obtain the following similarity measurement model, as follows:

$$\begin{aligned}\sum_{i,j}\left|d_{i,j} - d'_{i,j}\right| &= \sum_{i,j}\left(w_i - w_j\right)^t \left(w_i - w_j\right) \\ &\quad -\left(w_i - w_j\right)^y y y^t \left(w_i - w_j\right) \\ &= \sum_{i,j}\left(w_i - w_j\right)^t \overline{yy}^t \left(w_i - w_j\right) \geq 0.\end{aligned} \tag{17}$$

To prevent the above equation from overfitting during image processing [20], it is rewritten as

$$R_S = -\left(w_i - w_j\right)^t \overline{yy}^t \left(w_i - w_j\right), \tag{18}$$

where R_S is the model constraint, y is the measurement model component. Then the final model processing equation is

$$y^* = \arg\min\left(w_1 * R_1 + w_2 * R_2 + w_3 * R_3\right), \tag{19}$$

where y^* is the optimal similarity measurement model, R_1 is constraint condition of the subject, R_2 is the constraint condition of the main feature, and R_3 is regular constraint. For the optimization of this model, it is necessary to solve many of the above items. Therefore, in the process of calculation, the above equation should be used for nonlinear problem planning, to obtain the optimal similarity measurement model.

This part is connected with the abovementioned medical image feature extraction part to complete the construction process of the feature similarity measurement model of the medical image. So far, the similarity measurement model of the medical image feature matrix based on CNN has been designed, as shown in Figure 2.

3. Experimental Analysis and Results

3.1. Data Sets and Evaluating Indicator. MURA data set and MRNet data set were used for analysis. MURA data set: one of the largest X-ray database at present, there are a large number of muscle and bone X-ray images, including shoulder, humerus, elbow, forearm, wrist, palm, and fingers. The images in this dataset are manually marked by doctors. MRNet data set: this data set is the MRI data of the knee joint, including 1370 MRI images and 1104 abnormal examination images such as anterior cruciate ligament tear and meniscus tear, which is suitable for the analysis and research of medical image features. From the above two data sets, one thousand medical image data are selected as the data source. First, the selected data are preprocessed to avoid noise, redundant information, and other interference, and then two thousand data are divided into 4 points for cross-validation. The performance of this model is fully verified through many experiments.

The test platform used in this test consists of two parts. Because of the particularity of the signal of the medical image acquisition site, the medical image acquisition equipment cannot take real photos. Therefore, it only shows the same prototype as the test equipment. To ensure the effectiveness of this test, the first step of this experiment is the global similarity measurement of medical images. In the second part, the local comparison is used to improve the detail of the design model test in this paper.

Error rate: in the calculation of medical image feature data, the function objective is optimized through spatial constraints. To verify the accuracy of the proposed method, the error rate is selected as an index for comparative analysis.

$$U = \frac{G}{E} \times 100\%. \tag{20}$$

Measurement time: to verify the classification efficiency of the proposed method, the classification time in different iteration times is selected as the indicator.

$$T = \frac{UX}{N}. \tag{21}$$

Fitting degree of measurement: to verify the accuracy of the method, a section of data is input and different methods are selected for fitting degree analysis. The fitting degree

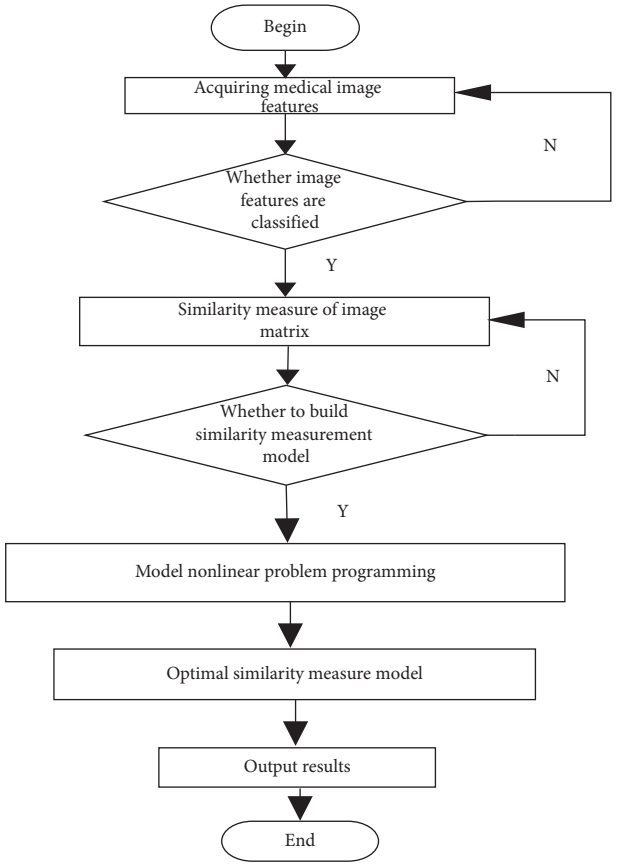


Figure 2: Similarity measurement model design process.

directly determines the classification performance of the method.

$$R^2 = 1 - \frac{Q}{L}. \tag{22}$$

3.2. Results and Discussion. In this test, any image is selected as the key image of similarity measurement, and the methods in Literature [2], Literature [3], Literature [4], and the design model in this paper are used to perform similarity measurement on the images in the image database and obtain similar images. Through the image extraction results, the differences between the methods in Literature [2], Literature [3], Literature [4], and the design model in this paper are obtained, as shown in Figure 3.

Through the above settings, the image extraction results of literature [2] method, literature [3] method, literature [4] method, and the design model in the paper are shown in Figure 4.

According to the above experimental results, the simulation measurement accuracy of the design model in this paper is better than that of the methods in Literature [2], Literature [3], and Literature [4]. Similar medical images with high accuracy can be obtained by using the model designed in this paper. The methods in Literature [2], Literature [3], and Literature [4] have poor accuracy for image similarity measurement and cannot extract all medical images with feature information in the database.

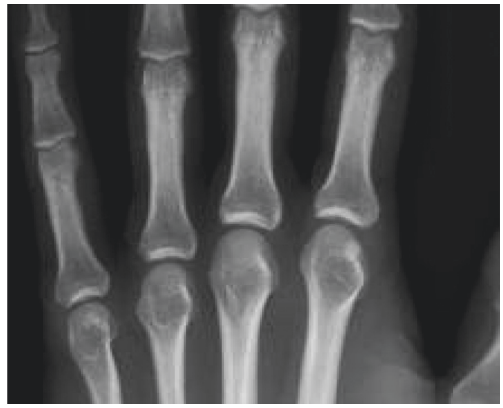


Figure 3: Test image.

The universal joint graph is selected as the test sample for local test. It is known that there are 150 images with the following joint characteristics in the database. The design model in this paper and the methods in Literature [2], Literature [3], and Literature [4] are used to perform similarity measurement on the image data in the database to extract images with the same features. In this part of the test, the above step is repeated for a total of 5 times, and the accuracy of image extraction is calculated, as shown in Figure 5.

The following test results are obtained through multiple experiments. To ensure the effectiveness of the test results, only two decimal places of the calculation results are reserved, as shown in Table 1.

According to Table 1. In the local similarity measurement test, the image extraction accuracy of the design model in this paper is better than that of the methods in Literature [2], Literature [3], and Literature [4]. It has a high similarity measurement ability for the images in the medical image database. By integrating the experimental results of the above two parts, it can be seen that the similarity measurement model of the medical image feature matrix based on CNN designed in this paper can achieve good results in both global similarity measurement and local similarity measurement. Therefore, the similarity measurement model of the medical image feature matrix based on CNN designed in this paper is better than the methods in Literature [2], Literature [3], and Literature [4] and has higher measurement accuracy, as shown in Figure 6.

The features extracted in this test are shown in the circle in Figure 6. Through the above settings, the methods in Literature [2], Literature [3], Literature [4], and the image extraction results of the design model in this paper are shown in Figure 7.

It can be seen from the above experimental results that the detail processing effect of the design model in this paper is better than that of the methods in Literature [2] and Literature [3]. The image of small details can be effectively obtained by using the model designed in this paper. The methods in Literature [2] and Literature [3] have a poor effect on obtaining the details of the image and cannot obtain the image consistent with the details of the test image.

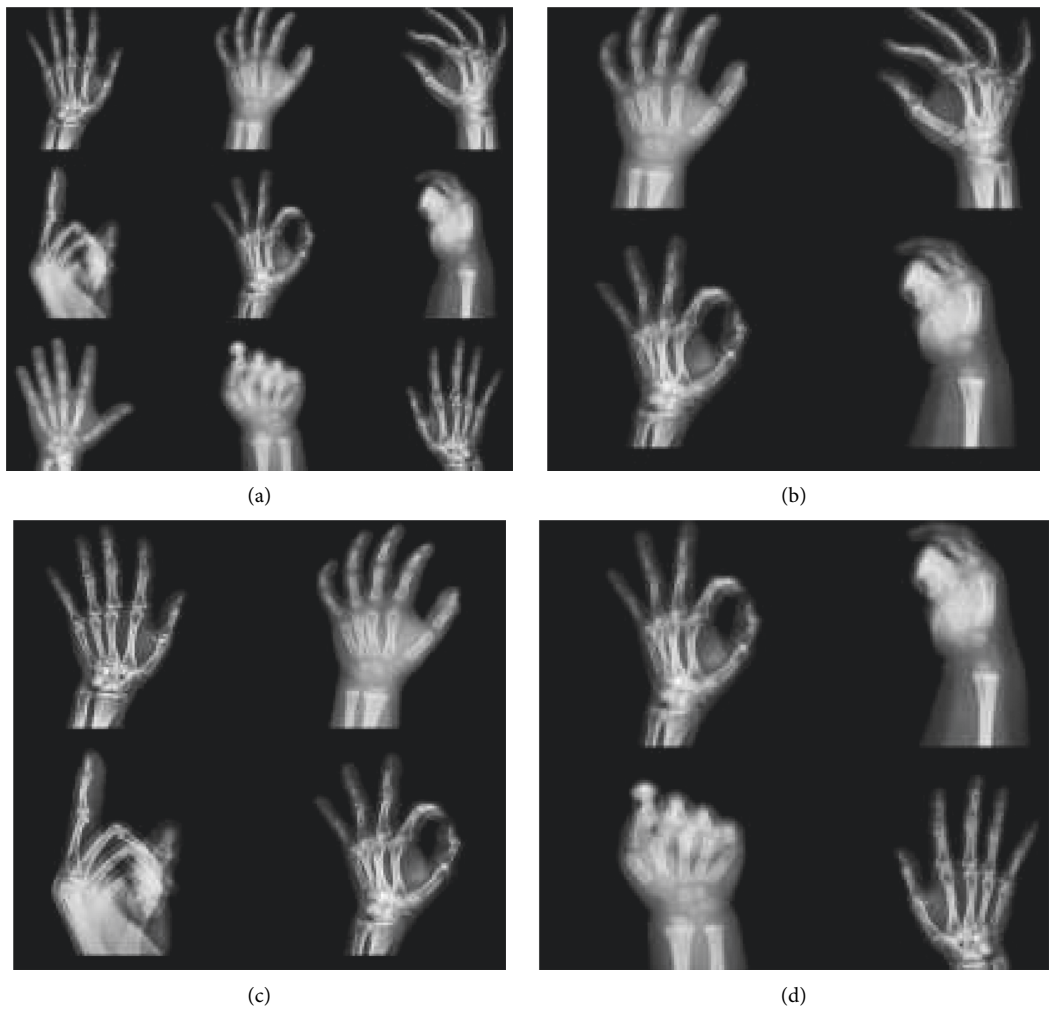


Figure 4: Comparison results of global similarity measurement. (a) Proposed method. (b) Literature [2] method. (c) Literature [3] method. (d) Literature [4] method.

Figure 5: Partial test image.

Table 1: Comparison results of local similarity measurement.

Number of tests	Literature [2] method/ %	Literature [3] method/ %	Literature [4] method/ %	Proposed method/%
1	92.02	90.62	87.20	98.25
2	92.62	81.31	91.23	98.74
3	92.15	94.43	89.36	98.64
4	92.25	84.66	88.95	98.37
5	92.44	94.65	87.64	98.87

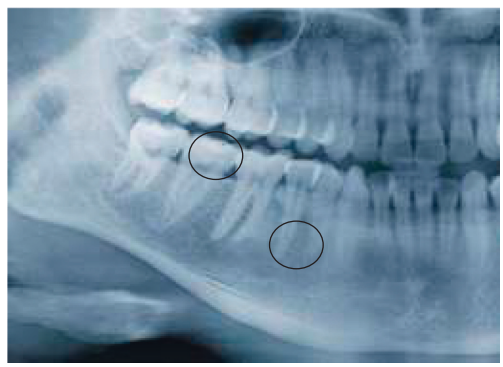


Figure 6: Dental test image.

To test the accuracy of this method in similarity measurement of medical image features, the proposed method is compared with the traditional methods in Literature [2], Literature [3], and Literature [4]. With 200 medical images used in the database, four methods are used to analyze the error rate, As shown in Figure 8.

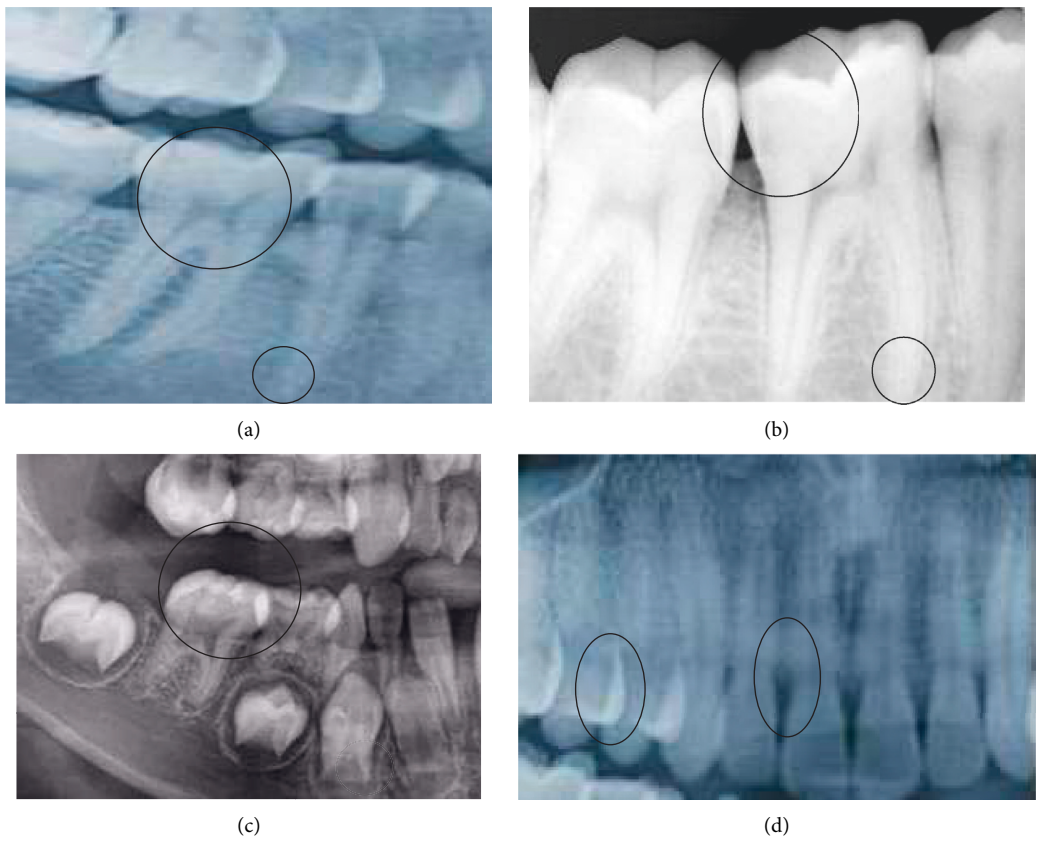


Figure 7: Comparison results of detail similarity measurement. (a) Proposed method. (b) Literature [2] method. (c) Literature [3] method. (d) Literature [4] method.

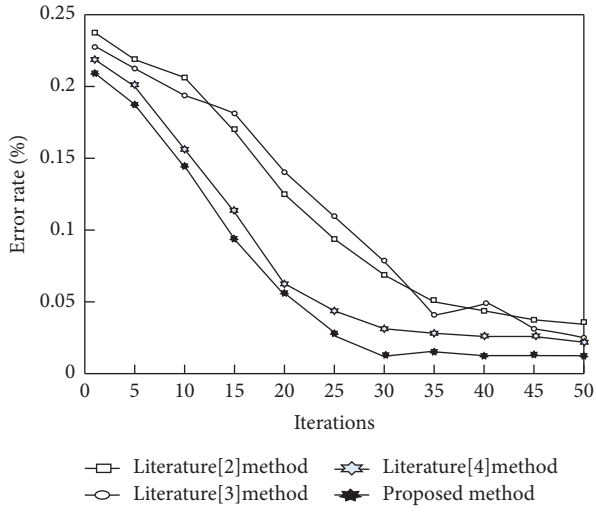


Figure 8: Comparison of error rates of different methods.

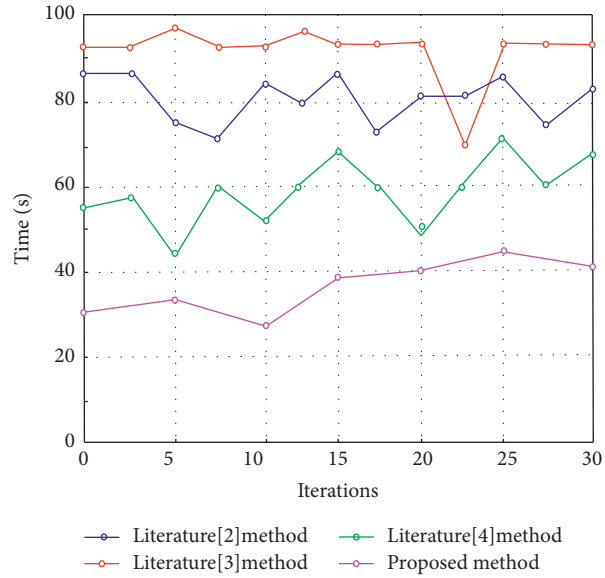


Figure 9: Comparison of measurement time of different methods.

As can be seen from Figure 8, the error rate of the method proposed in this paper is smaller than that of the other three methods in different iteration times. The highest error rate is only about 0.21, and the minimum error rate can be 0.01. The decrease in error rate is the largest. The minimum error rate of literature [2] method, literature [3] method, and literature [4] method is also between 0.03 and 0.05. Therefore, the proposed method has obvious advantages in the accuracy of image feature detection.

To further verify the measurement efficiency of the proposed method, the measurement time is taken as the experimental index to compare the four methods. In the

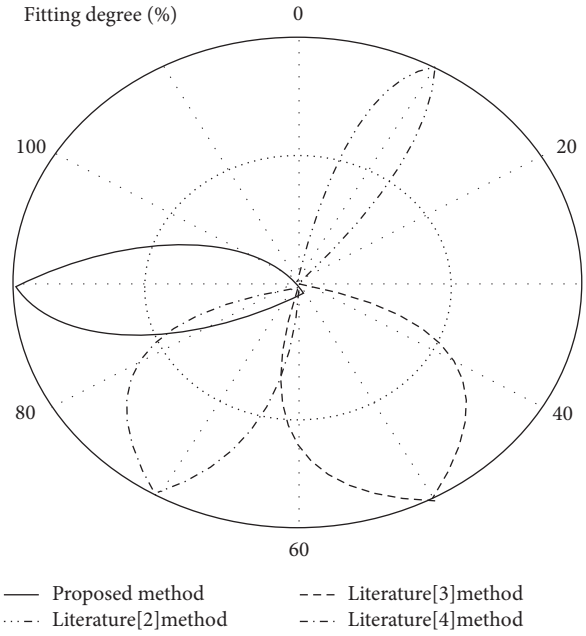


Figure 10: Comparison of the fitting degree of different methods.

experiment, 200 medical images in the database are selected, and the number of iterations is 30, as shown in Figure 9.

According to Figure 9, the measurement time of the method proposed in this paper is the lowest at the 10^{th} iteration, only about 25 s, and the highest at the 25^{th} iteration, about 43 s. However, the minimum time of the method in Literature [2], Literature [3], and Literature [4] is 73 s, 71 s, and 43 s, respectively, which are higher than the maximum time of the proposed method. And the measurement time variation of this method is smaller than that of the other three methods. Therefore, it can be considered that the proposed method has high measurement efficiency.

Taking the fitting degree as the test index, the comparative experiments of the four methods are performed. The experimental results obtained according to the experimental indexes in Section 3.1 are shown in Figure 10.

Figure 10 shows that the fitting degree of the proposed method in this paper is the highest among the four methods, which can reach about 91%, and it reaches 100% when the fitting degree is the highest. However, the fitting degree of the methods in Literature [2], Literature [3], and Literature [4] is relatively low. In particular, the fitting degree of the methods in Literature [2] is only about 15%, and the fitting degree of the methods in Literature [3] and Literature [4] is about 51% and 71%, respectively, which is quite different from the fitting degree of the methods in this paper. Therefore, it can be seen that the method proposed in this paper has obvious advantages in fitting degree and meets the actual needs.

4. Conclusion

With the continuous development of medical image imaging technology, there are gradually more ways to obtain medical images. Medical image has the advantages of low cost, simple operation, and easy preservation, as well as enhancing people's health awareness, thus it has been widely used in medical treatment in recent years. To further analyze medical images and obtain more useful information, this paper proposes a similarity measurement model of medical image feature matrix based on CNN. The global and local information of the medical image is analyzed, the image feature points are extracted orderly, and the CNN is input to construct the feature matrix. Based on the CNN operation, the similarity measurement model of the medical image feature matrix is constructed, and the model design is completed. The results show that the proposed model has a good similarity measurement effect in global, local, and detailed aspects, and the maximum error rate of this model is only about 0.21, the minimum operation time is only about 25 s, and the overall fitting degree can reach about 91%, which has significantly superior performance compared with traditional methods.

In the research of medical images, it is an important direction for the development of medical images in the future to realize the diagnosis of disease according to the different locations and the similarity of location. In future research into medical images, it is necessary to not only focus on medical knowledge but also learn from the corresponding model recognition and bionics knowledge, to integrate multiple patterns, obtain better similarity measurement methods, and make more breakthroughs. In this study, we only optimize the insufficient use of the previous model, and the research depth is insufficient. In the future, we need to continue to find better calculation methods to improve the use effect of the similarity measurement model.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] X. Liu, L. Yang, J. Chen, S. Yu, and K. Li, "Region-to-boundary deep learning model with multi-scale feature fusion for medical image segmentation," *Biomedical Signal Processing and Control*, vol. 71, no. PA, Article ID 103165, 2022.

[2] M. Abdar, M. A. Fahami, and S. Chakrabarti, "BARF: a new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification," *Information Sciences*, vol. 577, no. 11, pp. 353–378, 2021.

[3] Y. Shu, J. Zhang, B. Xiao, and W. Li, "Medical image segmentation based on active fusion-transduction of multi-stream features," *Knowledge-Based Systems*, vol. 220, no. 4, Article ID 106950, 2021.

[4] S. Poudel and S. W. Lee, "Deep multi-scale attentional features for medical image segmentation," *Applied Soft Computing*, vol. 109, no. 5, Article ID 107445, 2021.

[5] Z. Wenping, L. Shaoqian, and M. Fang, "A random walk similarity measurement model based on relative entropy,"

Journal Of Nanjing University (Natural Science), vol. 55, no. 06, pp. 984–999, 2019.

[6] X. Feng, N. Jiqiang, and L. Hao, "Establishment of the similarity metric model of multi-scale spatial object using isometry," *Geomatics and Information Science of Wuhan University*, vol. 44, no. 9, pp. 1399–1406, 2019.

[7] P. Wang, N. Jiang, Y. Wan, and Y. Wang, "Measuring similarity of spatio-temporal trajectory using Hausdorff distance," *Journal of Computer-Aided Design & Computer Graphics*, vol. 31, no. 4, pp. 647–658, 2019.

[8] M. Z. Ahmed and C. Mahesh, "An efficient image based feature extraction and feature selection model for medical data clustering using deep neural networks," *Traitement du Signal*, vol. 38, no. 4, pp. 1141–1148, 2021.

[9] H. Ayadi, M. T. Khemakhem, M. Daoud, J. X. Huang, and M. Ben Jemaa, "MF-Re-Rank: a modality feature-based Re-Ranking model for medical image retrieval," *Journal of the Association for Information Science and Technology*, vol. 69, no. 9, pp. 1095–1108, 2018.

[10] K. Fu and W. Xiajingbo, "Cloud model similarity measurement method based on mutual membership," *Journal of Beijing University of Technology*, vol. 38, no. 4, pp. 405–411, 2018.

[11] H. Guan and M. Liu, "Domain adaptation for medical image analysis: a survey," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1173–1185, 2022.

[12] P. Srinivasu, P. B. Durga, and R. Dhuli, "Multimodal medical image fusion based on content-based decomposition and PCA-Sigmoid," *Current medical imaging*, vol. 18, no. 5, pp. 546–562, 2022.

[13] V. M. Bashyam, J. Doshi, G. Erus et al., "Deep generative medical image harmonization for improving cross-site generalization in deep learning predictors," *Journal of Magnetic Resonance Imaging*, vol. 55, no. 3, pp. 908–916, 2021.

[14] A. Andrushia, K. Sagayam, H. Dang, M. Pomplun, and L. Quach, "Visual-saliency-based abnormality detection for MRI brain images—alzheimer's disease analysis," *Applied Sciences*, vol. 11, no. 19, p. 9199, 2021.

[15] P. Szwargulski, M. Moddel, N. Gdaniec, and T. Knopp, "Efficient joint image reconstruction of multi-patch data reusing a single system matrix in magnetic particle imaging," *IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 932–944, 2019.

[16] J. Peng, X. Zhang, W. Hui et al., "Improving the measurement of semantic similarity by combining gene ontology and co-functional network: a random walk based approach," *BMC Systems Biology*, vol. 12, no. S2, pp. 18–116, 2018.

[17] T. Lan, J. Liu, H. Qin, and L. L. Xu, "Time-domain global similarity method for automatic data cleaning for multi-channel measurement systems in magnetic confinement fusion devices," *Computer Physics Communications*, vol. 234, pp. 159–166, 2019.

[18] Y. Ding, C. Zhang, M. Cao et al., "ToStaGAN: an end-to-end two-stage generative adversarial network for brain tumor segmentation," *Neurocomputing*, vol. 462, pp. 141–153, 2021.

[19] D. Li, W. Yu, K. Wang, D. Jiang, and Q. Jin, "Speckle noise removal based on structural convolutional neural networks with feature fusion for medical image," *Signal Processing: Image Communication*, vol. 99, pp. 116500–116511, 2021.

[20] S. Wang, C. Li, R. Wang et al., "Annotation-efficient deep learning for automatic medical image segmentation," *Nature Communications*, vol. 12, no. 1, pp. 5915–5847, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 4825720, 9 pages
https://doi.org/10.1155/2022/4825720

Research Article

Segmentation for Human Motion Injury Ultrasound Medical Images Using Deep Feature Fusion

Jingmeng Sun and Yifei Liu

Physical Education Department, Harbin Engineering University, Harbin 150001, China

Correspondence should be addressed to Yifei Liu; liuyifei@hrbeu.edu.cn

Received 26 May 2022; Revised 14 July 2022; Accepted 29 July 2022; Published 29 August 2022

Academic Editor: Xiaofeng Li

Image processing technology assists physicians in the analysis of athletes' human motion injuries, not only to improve the accuracy of athletes' injury detection but also to improve the localization and recognition of injury locations. It is important to accurately segment human motion injury ultrasound medical images. To address many problems such as poor effect of traditional ultrasonic medical image segmentation algorithm for a sports injury. Therefore, we propose a segmentation algorithm for human motion injury ultrasound medical images using deep feature fusion. First, the accurate estimated value of human posture is extracted and combined with image texture features and image gray value as the target feature value of the ultrasonic medical image of human motion injury. Second, the image features are deeply fused by an adaptive fusion algorithm to enhance the image resolution. Finally, the best segmentation value of the image is obtained by the trained support vector machine to realize the accurate segmentation of human motion injury ultrasonic medical image. The results show that the average accuracy of the posture accurate estimation of the proposed algorithm is 95.97%; the segmentation time of the human motion injury ultrasound medical image of the proposed algorithm is below 150 ms; and the convergence of the algorithm is completed when the number of iterations is 3. The maximum segmentation error rate is 2.68%. The image segmentation effect is consistent with the ideal target segmentation effect. The proposed algorithm has important application value in the field of ultrasonic medical diagnosis of sports injury.

1. Introduction

Human motion analysis is to use relevant methods to track and capture human motion parameters, so as to reconstruct human structure and posture and achieve the goal of estimating and identifying human structure and posture [1]. With the continuous progress of image processing technology [2], using image processing technology to assist doctors in human motion injury analysis of athletes is one of the focuses of scholars at home and abroad in recent years. Using image processing technology to process the ultrasonic medical image of the injured part of human motion athletes not only can improve the injury detection of athletes but also can improve the ability to locate and identify the injured position. Therefore, the ultrasonic medical image segmentation algorithm for human motion injury has very important research significance.

Yan et al. [3] proposed an image segmentation algorithm based on a level set. First, the algorithm obtains the value of the image symbol function through the calculation of the objective function, determines the relevant constraints, and establishes the image segmentation model. Finally, the image contour segmentation is completed based on the established model and the set relevant initial curve, but the algorithm is too complex and has the practical problem of long segmentation time. Liu et al. [4] proposed the ore image segmentation algorithm based on U-Net and Res-Unet convolution networks. First, the algorithm implements gray value and median filtering to complete image denoising and uses the histogram equalization method to extract the image target position. And then, based on U-net and Res-U-net convolution network, the image contour segmentation model is constructed; finally, the image segmentation is realized through the image segmentation model. However,

in practical application, it is found that the convergence speed of this algorithm is slow and the practical application effect is poor. Sediqi and Lee [5] proposed an image semantic segmentation algorithm based on new sampling and context convolution. The algorithm preserves the spatial information of the image based on the guidance filter; then, the local context convolution method is used to cover the target in the image, and the accurate boundary description is carried out; and finally, the accurate segmentation of the image is realized by depicting the results. However, this algorithm has a high segmentation error rate and poor practical application effect. Lin and Li [6] proposed a parallel region segmentation algorithm for high-resolution remote sensing images based on a minimum spanning tree. The algorithm divides the image into several pixel blocks according to the conventional mosaic algorithm. Based on the multicore parallel method, the image segmentation rules are obtained, and the minimum spanning tree model is established. Finally, the model is solved by the parallel merging method to realize image segmentation. However, the segmentation time of this method is poor, which is quite different from the ideal application effect. Xue et al. [7] proposed an image segmentation algorithm based on improved FCN-8s. First, the algorithm establishes the image data set, extracts the image features according to the multiscale feature extraction method, and then establishes the convolution network model of the image. Finally, the image segmentation is realized through the output of the model. However, this method has a slow convergence speed and is difficult to achieve the expected goal.

The above image segmentation algorithm fails to fuse the image pixel values. Therefore, the above algorithm has the problems of long segmentation time, slow convergence speed, high segmentation error rate, and poor segmentation effect. In order to address these problems, this paper proposes a human motion injury ultrasonic medical image segmentation algorithm based on deep feature fusion. The contributions of this paper are as follows: (1) according to the accurate estimation of human posture, image texture features, and image gray value, the image target feature value is obtained, which improves the accuracy of feature extraction. (2) This study uses the advantages of fast speed and high precision of support vector machine to realize accurate image segmentation and puts forward the quality and efficiency of image segmentation. (3) Multiple data sets are selected for testing to prove that this algorithm has certain advantages in cutting time, convergence speed, segmentation error rate, segmentation effect, and so on.

2. Methodology

Before the image segmentation of human motion injury ultrasonic medical image [8, 9], it is necessary to give priority to estimating the athlete's motion posture and take it as the feature of the medical ultrasonic image to assist in the segmentation of the medical ultrasonic image.

2.1. Analysis of Human Motion Trajectory of Athletes. The high-definition camera is used to capture the athlete's motion process, obtain the athlete's motion video image, and complete the motion trajectory analysis of the athlete's human parts on this basis [10, 11].

2.1.1. Trajectory Analysis of Human Single Part. Set the single part appearance test item of human as $\Omega_a(b)$, and $\Phi_a(b_t, b_{t+1})$ is similarity measure, as to complete the motion trajectory expression of a single part of the human. The process is as follows:

$$\begin{cases} \Omega_a(b) = \sum_{i=1}^{C}\left(\Omega(b_i) + \sum_{n=1}^{D-1}\beta_n\left\|b_i \neq b_i^n\right\|\right), \\ \Phi_a(b_t, b_{t+1}) = \\ \exp\left(-\dfrac{\alpha^2(T(b_t), T(b_t + b_{t+1}))\left\|\widehat{b}_t - b_{t+1}\right\|_2^2}{\theta^2}\right), \end{cases} \tag{1}$$

where the image frame of the human single part motion image is described by t and $(t+1)$. β_n is Lagrange multiplier; D denotes the total number of multipliers; and C is the single part appearance feature set. The appearance of athletes' human parts is described by $\Omega(b_i)$, and b is the position of athletes' human parts in the image frame. α^2 is chi-square operator. $T(b)$ is local HOG feature coefficient [12]; θ is a relevant parameter; and $\widehat{b}_t$ is the predicted location of the $(t+1)$ image frame through streamer prediction.

According to the above calculation results, set the athlete's single part status configuration parameter as χ_i^e to obtain the athlete's single part trajectory. The process is as follows:

$$E_e(\chi^e) = \sum_{i=1}^{L}\Omega_a(\chi_i^e) + \sum_{i=1}^{L-1}\Phi_a(\chi_i^e, \chi_{i+1}^e), \tag{2}$$

where $E_e(\chi^e)$ is the athlete's single part trajectory, χ^e is video sequence, L denotes the number of video frames, the optimization function of athlete's single part trajectory [13] is described by E_e, and i is the constant.

2.1.2. Analysis of Athlete's Symmetrical Part Trajectory. Based on the analysis of the trajectory of single parts of the human, the trajectory of symmetrical parts of athletes is analyzed, so as to realize the analysis of the trajectory of athletes' human motion. Since the positions of symmetrical parts are left and right symmetrical, it is necessary to obtain their positions through the position coordinates of x, y. Set the athlete's compound coordinate position as $f = (x, y)$ to obtain the appearance feature of the athlete's symmetrical part. The specific results are as follows:

$$\Omega_\varepsilon(f) = \frac{\left(\Phi_a(f,x) + \Phi_a(f,y)\right) \times \left(\wedge(f,x)^T \cdot \wedge(f,x)\right)}{\left(1 + g^{-|f,x-f,y|/K}\right)}, \tag{3}$$

where the expression of the athlete's symmetrical part is marked in the form of $\Phi_a(f,x)$ and $\Phi_a(f,y)$. g is the penalty function of the symmetrical part; K denotes the penalty coefficient; and $\wedge$ stands for the color histogram of the human motion image.

Finally, based on the above calculation results, the motion track of athletes' symmetrical parts is obtained, and the process is as follows:

$$E_\varepsilon(\chi^\varepsilon) = \sum_{i=1}^{L} \Omega_\varepsilon(\chi_i^\varepsilon) + \sum_{i=1}^{L-1} \Phi_\varepsilon(\chi_i^\varepsilon, \chi_{i+1}^\varepsilon), \tag{4}$$

where χ^ε is video sequence, χ_i^ε represents status configuration set of symmetrical parts of athletes, and E_ε is the target optimization function.

2.2. Establishment of an Accurate Estimation Model of Athletes' Motion Posture. According to the analysis results of athletes' human motion trajectory, an accurate estimation model of athletes' motion posture is constructed to realize the accurate estimation of athletes' motion posture. The athletes' motion posture can be expressed by three-dimensional coordinates with p nodes. To reduce the fuzziness of athletes' motion posture estimation results, it is assumed that the athletes' motion posture to be estimated can be linearly represented by a set of predefined basic postures. The athlete's motion posture model in the video can be expressed in the following equation:

$$S_t = \sum_{i=1}^{k} c_{it} B_i, \tag{5}$$

where t represents the value of frame t of the athlete's motion video, B_i represents the basic posture of athletes, k represents the number of atoms of the basic posture B_i, and c_{it} is the coefficient of base posture B_i.

The two-dimensional posture of each frame of the athlete's motion video can be obtained by translating and rotating the corresponding three-dimensional posture. Therefore, the projection relationship between the athlete's two-dimensional posture and its corresponding athlete's three-dimensional posture can be expressed by the following equation:

$$W_t = \left(R_t S_t + T_t 1^T\right), \tag{6}$$

where R_t and T_t represent the camera translation vector and rotation matrix, respectively, and T represents matrix transpose.

Supposed that W, R, and T represent the set of W_t, R_t, and T_t of all frames, respectively.

$\theta = \{C, R, T\}$ is the overall parameter of the athlete's three-dimensional posture coordinates. $L(\theta; W)$ is the loss function, expressed as follows:

$$L(\theta; W) = \frac{\nu}{2} \sum_{i=1}^{n} \left\| W_t - R_t \sum_{i=1}^{k} c_{it} B_i - T_t 1^T \right\|_F^2, \tag{7}$$

where $\| * \|_F$ is the F norm of the representative matrix and ν represents the projection variance of athletes' three- to two-dimensional posture.

According to the above results, the athlete's motion posture model is optimized, and the optimization results are

$$S_t = \frac{\nu}{2} \sum_{i=1}^{k} c_{it} R_{it} B_i. \tag{8}$$

Set the fixed components of the model as $\eta = 26$ and local mixing type as $\mu = 4$, and A is the top of the model. δ is edge set of connected parts of human; $B = \left\{b_i|_{i=1}^{K}\right\}$ denotes location collection of human parts; and χ_i is human part status configuration, so as to establish the accurate estimation model of athlete's posture [14]. The process is as follows:

$$M(\chi) = \frac{\sum_{\chi_i \in \chi} \Omega(\chi_i) + \sum_{\chi_i, \chi_j \in \chi} \gamma\left(\chi_i, \chi_j\right)}{A} \delta, \tag{9}$$

where $M(\chi)$ is the established accurate athlete's posture prediction model, $\gamma(\chi_i, \chi_j)$ is the athlete injury scoring item, $\Omega(\chi_i)$ stands for appearance features of human parts, and the part compatibility of the top of a human $A = (v_i, v_j)$ is marked in the form of b_i and b_j.

2.3. Accurate Estimation Process of Athletes' Posture. The specific accurate estimation process of an athletes' posture is shown in Figure 1.

According to the data in Figure 1, the accurate estimation process of athlete's posture is to capture the athlete's motion process through the high-definition camera, obtain the athlete's motion video image, and complete the image acquisition. On this basis, the human position and the initial position of each part of the human in a single frame image are estimated. According to the relevant estimation results, the relevant constraint relationship of human part posture estimation is established, so as to build the relevant posture estimation model. By obtaining the motion trajectory of human parts, the model is optimized, and the accurate estimated value of athlete posture is output.

2.4. Segmentation Algorithm of Athletes' Injury Ultrasonic Image. Taking the accurate estimated value of athletes' posture obtained above as the target feature of athletes' injury ultrasonic image, SVM [15] is used to segment athletes' injury ultrasonic image.

2.4.1. Image Segmentation Recognition. According to the determined image features and high-resolution CT scanning algorithm, the high-resolution information sampling equation of medical ultrasound image is obtained [16]. The process is as follows:

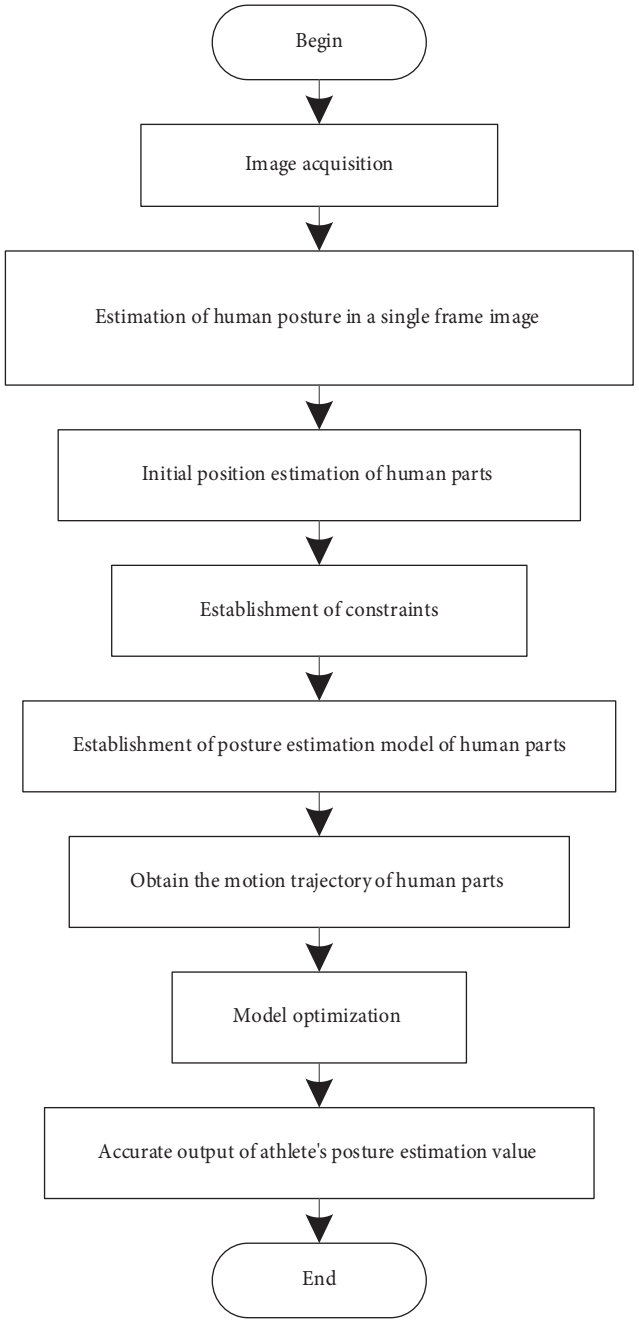


Figure 1: Accurate estimation process of athletes' posture.

$$\begin{cases} \dot{m}(r) = -\mathrm{Gm}(r) + \mathrm{Hf}(n(r-\lambda)), \\ \dot{n}(r) = -\mathrm{On}(r) + \mathrm{Pm}(r-\varphi), \end{cases} \tag{10}$$

where pixel acquisition point of ultrasonic medical image is marked in the form of G, H, O, and P. $m(r)$ is pixel length of the injured part of the athlete in the image; $n(r)$ is the width; and the domain decomposition value of information feature fusion for human motion injury is marked in the form of λ and φ.

Because the acquired image features are not a kind of athlete's posture estimation value, it is necessary to deeply fuse the image features. Set the similarity feature of the ultrasonic medical image as $w(m,n)$; $N\mathrm{cut}(m,n)$ is the image gray histogram; and (x, y) denotes the gray pixel set, so as to establish feature decomposition model of ultrasonic image [17]. The process is as follows:

$$F_{fj}z(x, y_m) = \frac{1}{\sqrt{x}Q}\left(\left(\frac{Q}{2} + \frac{xQ}{2}\right) - y_m\right), \tag{11}$$

where $z(x, y_m)$ is image texture output, Q stands for feature decomposition coefficient, $F_{fj}z(x, y_m)$ is the established feature decomposition model, and m is the constant.

According to the adaptive fusion algorithm, the pixel value of medical ultrasound image is fused, and the deep feature fusion output of the medical ultrasound image is completed. The calculation equation is as follows:

$$R(a) = \left[J(a) - \frac{S_{zj}}{\max(t(a), t_0)}\right] + S_{zj}, \tag{12}$$

where $R(a)$ is feature fusion output value of medical ultrasound image, S_{zj} is image pixel subset, $J(a)$ represents the gray image [18, 19], t_0 is the image initial structure similarity, and $(t(a), t_0)$ denotes the similarity degree.

2.4.2. Proposed Algorithm. The process of ultrasonic medical image segmentation algorithm for human motion injury is as follows:

Input: ultrasonic medical image of human motion injury

Output: image segmentation results

(1) Based on the deep feature fusion results of human motion injury ultrasonic medical image, the corresponding constraints are formulated to reduce the segmentation surface error of ultrasonic image. First, the fusion evaluation result of the human motion injury image is set as v, and the image matching evaluation function is marked as $v(g_i)$. Based on this, the distance between the pixel and the cluster center in the human motion injury image is calculated, and the pixel gray output of the visual graphics of the human motion injury image is obtained, which is used as the pathological characteristic value of the ultrasonic image. The calculation equation is as follows:

$$\vartheta = \arccos\left\{\left[\frac{\overrightarrow{Z_iW_k} \times \overrightarrow{Z_iZ_j}}{\left\|\overrightarrow{Z_iW_k}\right\| \cdot \left\|\overrightarrow{Z_iZ_j}\right\|}\right]_{\max}\right\}, \tag{13}$$

where ϑ is the pixel gray output value is marked. W_i is the gradient direction position value of the image, and $\overrightarrow{Z_iZ_j}$ denotes the image rendering vector.

(2) According to the above calculation results, the fusion region constraint is carried out on the visual graphics of the ultrasonic image, and the threshold segmentation value of SVM is obtained. The calculation equation is as follows:

$$\xi(m,n) = \frac{1}{T(i)} \exp\left(-\frac{c(m,n)}{\rho^2} \right), \tag{14}$$

where $\xi(m,n)$ is the optimal segmentation threshold, $T(i)$ denotes the best match value of the image template, ρ^2 is a constraint parameter, and $c(m,n)$ stands for the feature point of the pixel. Finally, the feature fusion value is minimized by a priori energy function to obtain the best segmentation output value. The calculation equation is as follows:

$$f(x) = \left(r_\rho x - r_\rho x_1 \right), \tag{15}$$

where $f(x)$ is the optimal segmentation and output of ultrasonic medical images of human motion injury, r_ρ is statistical features of ultrasonic images, and x is image edge subset.

(3) According to the optimal segmentation value of SVM, the ultrasonic pathological feature is adaptively segmented, so as to realize the accurate segmentation of human motion injury ultrasonic medical images [20].

The process of ultrasonic medical image segmentation algorithm for sports injury based on deep feature fusion is shown in Figure 2.

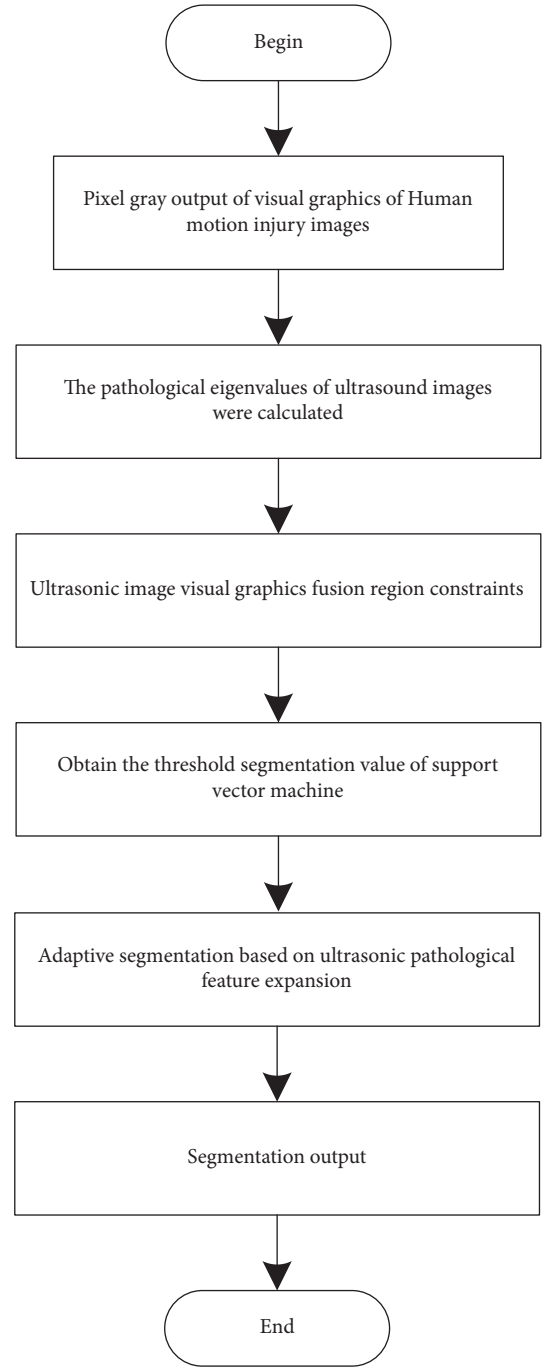


Figure 2: The proposed algorithm process.

3. Experimental Analysis and Results

3.1. Data Set. In this paper, MedPix data set and MURA data set data sets are selected for the experiment. MedPix data set is a free and open online access database, which contains medical images, teaching cases, and clinical topics and integrates image and text metadata, including more than 12,000 patient cases, 9,000 topics, and nearly 59,000 images. Our main target audience includes doctors and nurses, full-time medical personnel, medical students, nursing students, and others who are interested in medical knowledge. The content materials are organized according to the location of the disease (organ system) and pathological category and patient data and through image classification and image title. The collection can be searched by patient symptoms and signs, diagnosis, organ system, image form and image description, keywords, contributing authors, and many other search options. MURA data set is one of the largest public ray image data sets produced by the Stanford machine learning working group. MURA is a data set of musculoskeletal radiographs. It contains a total of 14,863 studies in 12,173 patients and 40,561 multiview ultrasound radiographs, including fingers, elbows, forearms, hands, humerus, shoulders, and wrists.

Integrated the data in the two data sets and input all the experimental sample data into the simulation software. Randomly select 2,000 images in two data sets to test the segmentation error rate. Eighty percent of the data is used for training, and 20% of the data is used for testing. After many tests, the optimal operating parameters of the simulation software are obtained, and relevant experiments are carried out under these parameters, so as to ensure the authenticity and reliability of the experimental results.

3.2. Experimental Standard. The proposed algorithm of human motion injury based on deep feature fusion, algorithm (algorithm 1) proposed by literature [3], algorithm (algorithm2) proposed by literature [4], algorithm (algorithm 3) proposed by literature [5], algorithm (algorithm 4) proposed by literature [6], and algorithm (algorithm 5) proposed by literature [7], were used for testing. In the process of image segmentation, the segmentation

Table 1: Accuracy of athletes' posture accurate estimation.

Number of images/piece	Estimated accuracy (%)					
	Proposed algorithm	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Algorithm 5
100	95.63	85.63	85.63	85.64	75.41	69.36
200	96.84	87.49	84.75	87.45	86.39	75.64
300	97.56	86.33	90.12	71.36	84.75	75.88
400	96.31	85.94	87.11	74.56	82.31	74.63
500	95.22	82.37	78.63	82.35	74.56	85.12
600	98.17	84.19	85.26	79.63	72.35	87.65
700	96.33	85.67	74.16	78.41	74.18	85.22
800	95.14	85.61	71.33	71.63	84.36	84.79
900	92.37	84.75	79.63	85.36	76.33	86.31
1,000	96.17	83.69	85.23	80.12	84.15	84.54
Average value	95.97	85.16	82.19	79.65	79.48	80.91

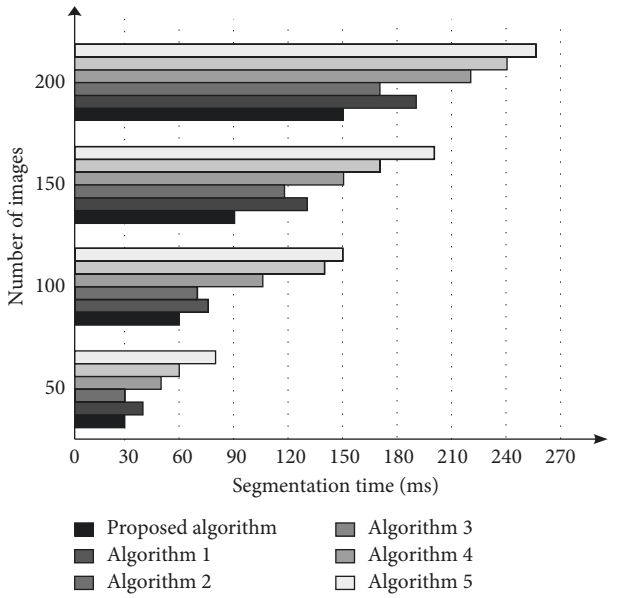


Figure 3: Comparison of segmentation time test results of different segmentation algorithms.

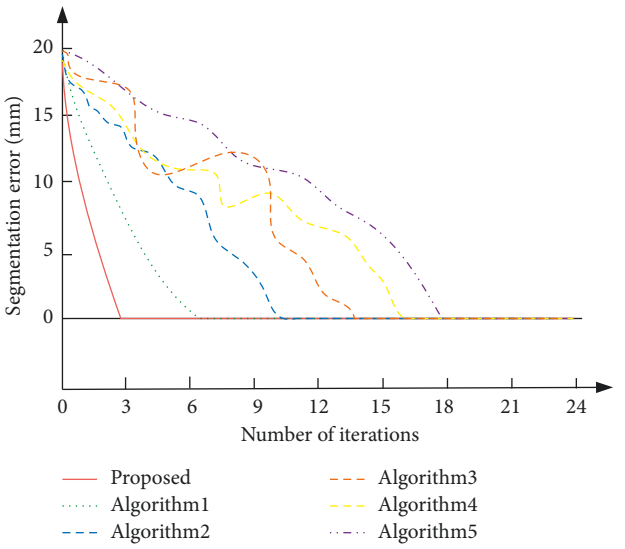


Figure 4: Comparison of convergence test results.

performance of segmentation algorithm is the key to reflect the effectiveness of segmentation algorithm. When using the proposed algorithm, algorithms 1–5 to carry out image segmentation, several test indexes such as accuracy, segmentation time, convergence speed, segmentation error, and segmentation effect are selected to test the segmentation performance of the above six image segmentation algorithms.

(1) Accuracy of athlete's posture estimation: this indicator refers to the probability of accurately estimating an athlete's posture. The calculation equation is as follows:

$$f = \frac{z_i}{z_j} \times 100\%, \tag{16}$$

where z_i refers to the amount of data about the accurately estimate athlete's posture and z_j refers to the total amount of athlete's data sample.

(2) Segmentation time: this indicator is only the sum of the time taken to complete the segmentation steps of ultrasonic medical images of human motion injury. The calculation equation is as follows:

$$t = \sum_{i=1}^{n} T_i, \tag{17}$$

where T_i represents the time taken for the i-th division step.

(3) Convergence speed: this indicator refers to the number of iterations when the segmentation error of ultrasonic medical images of human motion injury is the lowest. The less the number of iterations, the faster the convergence speed.

(4) Segmentation error rate: this index refers to the ratio of the number of correctly segmented images to the total number of experimental samples. The calculation equation is as follows:

$$e = \frac{o}{p} \times 100\%, \tag{18}$$

Table 2: Comparison of segmentation error rate test results of different segmentation algorithms.

Number of images/piece	Segmentation error rate test result (%)					
	Proposed algorithm	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Algorithm 5
100	0	0	0	0.11	0.27	0.42
200	0	0	0.10	0.47	0.61	1.04
300	0	0.22	0.48	1.13	1.55	2.08
400	0.25	0.39	0.99	1.88	2.03	3.43
500	0.47	0.58	1.57	2.59	3.21	4.57
600	0.89	1.05	2.03	3.01	4.08	5.29
700	1.23	1.64	2.84	3.99	5.12	6.43
800	1.79	2.08	3.42	4.58	6.25	7.14
900	2.37	2.51	4.05	5.01	7.39	7.92
1,000	2.68	3.16	4.92	5.76	8.01	8.50

where o refers to the number of correctly segmented images, and p refers to the total amount of experimental samples.

(5) Segmentation effect test: the closer the segmentation results of different algorithms are to the ideal segmentation results, the better the segmentation effect of human motion injury ultrasonic medical image.

3.3. Results and Discussion. The accuracy of athletes' posture accurate estimation of the proposed algorithm, algorithm 1, algorithm 2, algorithm 3, algorithm 4, and algorithm 5 are compared. The comparison results are shown in Table 1.

According to the data in Table 1, the average accuracy of an athlete's posture estimation of the proposed algorithm is 95.97%. The average accuracy of an athlete's posture estimation of algorithm 1 is 85.16%. The average accuracy of an athlete's posture estimation of algorithm 2 is 82.19%. The average accuracy of an athlete's posture estimation of algorithm 3 is 79.65%. The average accuracy of an athlete's posture estimation of algorithm 4 is 79.48%. The average accuracy of an athlete's posture estimation of algorithm 5 is 80.91%. Compared with the experimental comparison algorithm, the accuracy of an athlete's posture estimation of algorithm in this paper is higher, and the estimation effect is better.

In the process of image segmentation, the segmentation time is the key index to display the segmentation performance of the segmentation algorithm. The above six image segmentation algorithms are used to carry out image segmentation and test the image segmentation time of the six algorithms, as shown in Figure 3.

It is seen in Figure 3 that in the process of image segmentation, the longer the segmentation time, the worse the segmentation performance of the segmentation algorithm, and the shorter the segmentation time, the higher the segmentation performance. By analyzing Figure 3, it can be seen that with the increase in detection times, the segmentation time of the six image segmentation algorithms has increased to varying degrees. The segmentation time of ultrasonic medical image of sports injury in the proposed algorithm is less than 150 ms, which is 45 ms, 19 ms, 71 ms, 90 ms, and 105 ms lower than algorithms 1–5, respectively. The proposed algorithm has the shortest image segmentation time among the six segmentation algorithms. This is because the proposed algorithm uses an adaptive fusion algorithm to fuse the image pixel values during image segmentation. Therefore, the segmentation time of the proposed algorithm in image segmentation is better than the other five image segmentation algorithms.

When carrying out image segmentation, the convergence speed of the segmentation algorithm is an important process to detect the segmentation performance of the segmentation algorithm. The above six image segmentation algorithms are used to carry out image segmentation, and the segmentation convergence speed of the six algorithms is tested, as shown in Figure 4.

According to Figure 4, in the process of image segmentation, the faster the convergence speed, the better the image segmentation performance, and vice versa. By analyzing the experimental data in Figure 4, it can be seen that the proposed algorithm completes the convergence of the algorithm when the number of iterations is 3, and the number of convergence is 4, 8, 11, 14, and 15 lower than algorithms 1–5, respectively. In general, the convergence speed of other algorithms is lower than that of the proposed algorithm, and the convergence speed of algorithm 4 is the most unstable. It can be proved that the proposed algorithm has a high convergence speed when implementing image segmentation.

In the process of image segmentation, the level of image segmentation error can directly affect the segmentation performance of the segmentation algorithm. When using the proposed algorithm, algorithms 1–5 to perform image segmentation, the segmentation error rates of six image segmentation algorithms are tested. It is shown in Table 2.

The larger the segmentation error, the worse the segmentation performance of the segmentation algorithm, and vice versa. It can be seen from Table 2 that the more images to be segmented, the greater the segmentation error measured by the six algorithms. The maximum segmentation error rate of the proposed algorithm is 2.68%, which is 0.48%, 2.24%, 3.08%, 5.33%, and 5.82% lower than algorithms 1–5, respectively. Overall, the test result of the proposed algorithm is the image segmentation algorithm with the smallest error among the six image segmentation

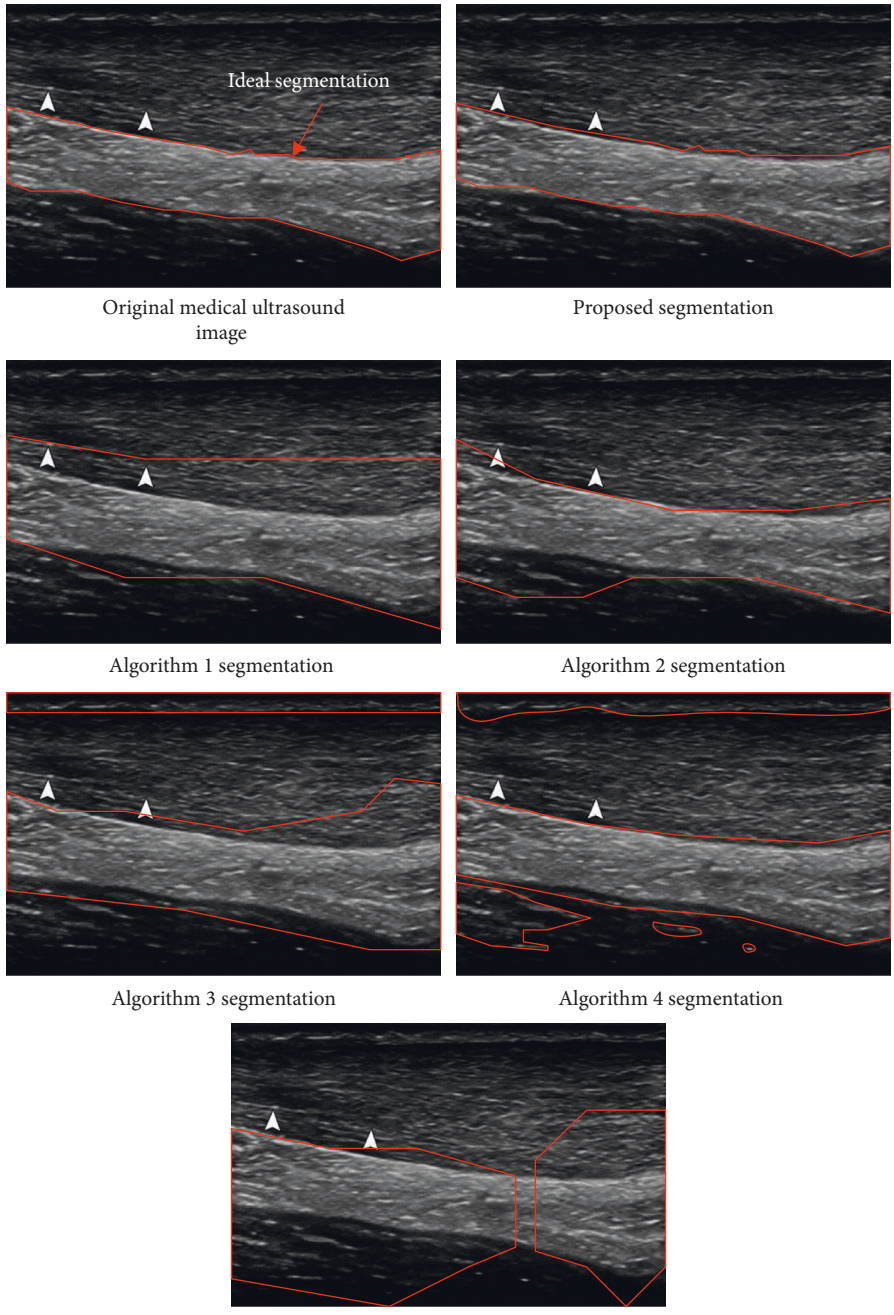


Figure 5: Comparison of segmentation effect test results of different segmentation algorithms.

algorithms. It can be proved that the segmentation error of the proposed segmentation algorithm is small, and the segmentation algorithm is effective.

Randomly select an image in the MedPix data set for the segmentation effect test. Based on the above test results, the segmentation effects of the above six image segmentation algorithms are tested, and the test results are shown in Figure 5.

According to the analysis of Figure 5, the image segmentation effect of the proposed algorithm is consistent with the ideal target segmentation effect, while the segmentation effects detected by other segmentation algorithms are lower than the test results of the proposed algorithm, in which algorithms 1–3 are inconsistent with the ideal segmentation effect, while algorithms 4 and 5 also have excessive segmentation. It can be proved that the proposed algorithm has an excellent segmentation effect in image segmentation, which proves that the segmentation algorithm has high effectiveness and superior segmentation performance.

4. Conclusions

With the continuous progress of image processing technology, the image segmentation algorithm has become one of the important technologies to assist doctors in diagnosis.

Aiming at the problems existing in the traditional image segmentation algorithm, a human motion injury ultrasonic medical image segmentation algorithm based on deep feature fusion is proposed. The algorithm trains SVM according to the image feature fusion results and completes the segmentation of human motion injury image through the eigenvalue output of SVM. The experimental results show that the average accuracy of the athlete's posture of the algorithm is 95.97%. The segmentation time of ultrasonic medical images of human motion injury in this paper is less than 150 ms. The convergence of the algorithm is completed when the number of iterations is 3. The maximum segmentation error rate is 2.68%. The image segmentation effect is consistent with the ideal target segmentation effect. However, there are some problems in the design process of the algorithm. The algorithm has some errors in enhancing the effect of image segmentation. In view of this problem, we will continue to optimize the segmentation algorithm until the algorithm is perfect, so as to further improve the quality of the ultrasonic medical image segmentation algorithm for human motion injury.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest with any financial organizations regarding the material reported in this manuscript.

References

[1] H. S. Munawar, A. W. Hammad, and S. T. Waller, "A review on flood management technologies related to image processing and machine learning," *Automation in Construction*, vol. 132, no. 8, Article ID 103916, 2021.

[2] Q. Yun and C. Leng, "Intelligent control of urban lighting system based on video image processing technology," *IEEE Access*, vol. 8, Article ID 155506, 2020.

[3] S. Yan, X. C. Tai, J. Liu, and H. Y. Huang, "Convexity shape prior for level set-based image segmentation method," *IEEE Transactions on Image Processing*, vol. 29, pp. 7141–7152, 2020.

[4] X. Liu, Y. Zhang, H. Jing, L. Wang, and S. Zhao, "Ore image segmentation method using U-Net and Res_Unet convolutional networks," *RSC Advances*, vol. 10, no. 16, pp. 9396–9406, 2020.

[5] K. M. Sediqi and H. J. Lee, "A novel upsampling and context convolution for image semantic segmentation," *Sensors*, vol. 21, no. 6, p. 2170, 2021.

[6] W. Lin and Y. Li, "Parallel regional segmentation method of high-resolution remote sensing image based on minimum spanning tree," *Remote Sensing*, vol. 12, no. 5, p. 783, 2020.

[7] J. Xue, Y. Wang, and A. Qu, "Image segmentation method for Lingwu long jujubes based on improved FCN-8s," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 37, no. 5, pp. 191–197, 2021.

[8] R. Sami, S. Soltane, and M. Helal, "Microscopic image segmentation and morphological characterization of novel chitosan/silica nanoparticle/nisin films using antimicrobial technique for blueberry preservation," *Membranes*, vol. 11, no. 5, 303 pages, 2021.

[9] R. Schmitz, F. Madesta, M. Nielsen et al., "Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture," *Medical Image Analysis*, vol. 70, no. 12, Article ID 101996, 2021.

[10] W. Li, Z. Luo, and X. Xi, "Movement trajectory recognition of sign language based on optimized dynamic time warping," *Electronics*, vol. 9, no. 9, pp. 1400–1412, 2020.

[11] H. Xue, D. Q. Huynh, and M. Reynolds, "A location-velocity-temporal attention LSTM model for pedestrian trajectory prediction," *IEEE Access*, vol. 8, Article ID 44576, 2020.

[12] G. Shuqing and S. Yucong, "Traffic sign recognition based on HOG feature extraction," *Journal of Measurements in Engineering*, vol. 9, no. 3, pp. 142–155, 2021.

[13] A. Dmd, B. Sjr, and C. Ds, "Multi objective optimization of fused deposition modeling process parameters with desirability function - ScienceDirect," *Procedia CIRP*, vol. 99, pp. 707–710, 2021.

[14] H. Yang, B. Zhou, Q. Wei, X. Wang, X. Xu, and R. Zhang, "Accurate attitude estimation of HB2 standardmodel based on QNCF in hypersonic wind tunnel test," *Chinese Journal of Aeronautics*, vol. 33, no. 1, pp. 64–72, 2020.

[15] F. Muñoz-Muñoz and A. Rodrigo-Mor, "Partial discharges and noise discrimination using magnetic antennas, the cross wavelet transform and support vector machines," *Sensors*, vol. 20, no. 11, 3180 pages, 2020.

[16] M. C. Klein and E. Roberts, "Automatic error control during forward flux sampling of rare events in master equation models," *The Journal of Chemical Physics*, vol. 152, no. 3, Article ID 035102, 2020.

[17] H. Rastgoftar, E. M. Atkins, and I. V. Kolmanovsky, "Scalable vehicle team continuum deformation coordination with eigen decomposition," *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2514–2521, 2022.

[18] Y. Alammari, I. Iovkov, S. Berger, J. Saelzer, and D. Biermann, "Adhesion area estimation using backscatter image gray level masking of uncoated tungsten carbide tools," *Wear*, vol. 476, no. 1, Article ID 203666, 2021.

[19] D. Zerbst, C. Liebold, T. Gereke, A. Haufe, S. ClauB, and C. Cherif, "Modelling inhomogeneity of veneer laminates with a finite element mapping method based on arbitrary grayscale images," *Materials*, vol. 13, no. 13, p. 2993, 2020.

[20] J. Xi, J. Chen, Z. Wang et al., "Simultaneous segmentation of fetal hearts and lungs for medical ultrasound images via an efficient multi-scale model integrated with attention mechanism," *Ultrasonic Imaging*, vol. 43, no. 6, pp. 308–319, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5096298, 11 pages
https://doi.org/10.1155/2022/5096298

Research Article

Multithreshold Microbial Image Segmentation Using Improved Deep Reinforcement Learning

Minghui Zhou

College of Electronic Information, Jilin Communications Polytechnic, Changchun 130000, Jilin, China

Correspondence should be addressed to Minghui Zhou; zhmhui@jljy.edu.cn

Received 29 May 2022; Revised 4 July 2022; Accepted 9 August 2022; Published 23 August 2022

Academic Editor: Xiaofeng Li

Image segmentation technology can effectively extract the foreground target in the image. However, the microbial image is easily disturbed by noise, its greyscale has the characteristics of nonuniform distribution, and several microorganisms with diverse forms exist in the same image, resulting in insufficient accuracy of microbial image segmentation. Therefore, a multithreshold microbial image segmentation algorithm using improved deep reinforcement learning is proposed. The wavelet transform method is used to remove the noise of the microbial image, the threshold number of the microbial image after denoising is determined by calculating the number of peaks of the grey histogram, and the foreground target of the microbial image is enhanced by the mean iterative threshold segmentation method, the preliminary segmentation of the microbial image is realized, the multithreshold microbial image segmentation model based on ResNet-Unet is constructed, and the cavity convolution and dual Q network mechanism are introduced to improve the segmentation model. The preliminary segmented microbial image is input into the improved segmentation model to realize the segmentation of the multithreshold microbial image. The results show that the proposed algorithm can effectively remove the noise of microbial images. With the increase in the number of thresholds, the peak signal-to-noise ratio, structural similarity, and feature similarity show an upward trend, and the loss rate of the model is less than 0.05%. The minimum running time of the algorithm is 3.804 s. It can effectively and quickly segment multithreshold microbial images and has important application value in the field of microbial recognition.

1. Introduction

In people's daily life, microorganisms are everywhere, in the recognition of microorganism images, how to accurately obtain the features of microorganism images is the core problem that needs to be addressed [1]. Extracting image features and obtaining microbial targets by image segmentation is a commonly used effective method. However, the microbial image segmentation is often susceptible to various factors, resulting in poor microbial image target extraction [2, 3]. First, microorganisms are diverse, most of them have nonrigid features, their morphology shows nonregularity, their shapes are very different and their sizes vary. Second, in the microbial images, the difference degree between the image target and background greyscale is small. For multitarget microbial images, the spatial distribution of microorganisms is crossed and overlapping, and conventional segmentation techniques may produce a pseudocontour phenomenon [4]. In addition, noise is another important factor affecting the imaging effect of microbial images. Excessive noise will cause the lack of detailed information in microbial images, especially when there is a large number of bright signals, the microbial target will be submerged in the background, resulting in image quality degradation [5]. These factors can adversely affect image segmentation and lead to the degradation of image segmentation quality and efficiency. Therefore, it is important to study a new multi-threshold microbial image segmentation method.

To address the important research topic of multithreshold microbial image segmentation, Li et al. [6] implemented the segmentation process of greyscale image thresholding by constructing a quantum circuit, which can complete the acquisition of the histogram of pixel distribution and determine the image threshold by obtaining the maximum interclass variance of each greyscale level, to

achieve image segmentation. However, the single threshold value determined by this method is not suitable for the segmentation of a wide variety of microbial images. Li et al. [7] proposed to improve the traditional fuzzy C-mean algorithm for poor image segmentation by using a parallel Lévy-grey-wolf optimization algorithm to improve the image segmentation by obtaining the best initial centre. However, this method still results in a serious loss of image detail information. Cai et al. [8] proposed the use of an improved particle swarm optimization algorithm to achieve segmentation processing of thresholded images, and the method obtained a small number of image thresholds, which has limitations in applying it to the segmentation of a variety of microbial images. Jia et al. [9] proposed to improve the multiverse optimization algorithm using the Leyland flight mechanism to improve the accuracy of image segmentation by accurately determining the number of image thresholds. However, this method fails to achieve accurate learning of image information due to premature convergence. Sun et al. [10] proposed an image small target recognition method based on a deep reinforcement learning method using a multilayer convolutional neural network to achieve image target segmentation. However, the convolutional neural network constructed by this method suffers from gradient loss, which reduces the image segmentation performance and thus has a serious impact on the accurate recognition of image targets. To improve the quality and efficiency of multithreshold microbial image segmentation, a new multithreshold microbial image segmentation algorithm using improved deep reinforcement learning is proposed, and the contributions of this paper are as follows:

(1) The wavelet transform method is used to remove the microbial image noise, which improves the image quality and reduces the interference of noise to the subsequent image segmentation

(2) The foreground target of microbial images is enhanced to achieve the initial segmentation of microbial images and lay the foundation for the subsequent multithreshold microbial image segmentation

(3) To improve and introduce a reinforcement learning mechanism to build the relevant segmentation model to achieve the ultimate aim of image segmentation quality and efficiency

2. Methodology

2.1. Multithreshold Microbial Image Segmentation. The process of multithreshold microbial image segmentation algorithm using improved deep reinforcement learning is shown in Figure 1.

Analysis of Figure 1 shows that the microbial image is preprocessed to remove the noise contained in the image. The calculation of the number of image thresholds is completed by analyzing the number of wave peaks in the grey histogram to determine the number of thresholds. The foreground target of the microbial image is enhanced to achieve the initial segmentation of the background and foreground. The image segmentation model based on ResNet-Unet is constructed, and the improvement of the segmentation model is completed by introducing the null convolution and double Q network mechanisms, and the initial segmentation results are input to the improved model to obtain the segmentation results of the relevant multithreshold microbial images.

2.2. Microbial Image Denoising Based on Wavelet Transform. In this paper, the wavelet transform method is used to realize the noise reduction of microbial image. Its principle is as follows: set $f(i,j), i, j = 1, 2, \cdots N$ as the original microbial image, in which the value of N is among the integer power of 2, (i, j) is the pixel point of the image, and the grey value of the pixel is expressed as $f(i, j)$. When a large amount of noise is covered in the original multithreshold microbial image, the noisy microbial image can be described by the following equation [11]:

$$v(i,j) = f(i,j) + b(i,j), \tag{1}$$

where $b(i, j)$ is the noise in the microbial image, which meets the Gaussian distribution, the mean value is 0 is, σ^2 is the variance which is independent just as $f(i, j)$. It is processed using the wavelet method, which is described by the following equation:

$$Z_{v_{i,j}} = Z_{f_{i,j}} + Z_{b_{i,j}}, \tag{2}$$

where for noisy microbial images, $Z_{v_{i,j}}$ is the wavelet coefficient. For the original microbial image, $Z_{f_{i,j}}$ is wavelet coefficient. The noise in the microbial image, $Z_{b_{i,j}}$ is wavelet coefficient.

The multithreshold microbial image noise reduction process based on wavelet transform is as follows:

Step 1: wavelet transform is used to process noisy microbial images $v(i, j)$. After s layers of orthogonal redundancy decomposition, the wavelet coefficient can be determined, expressed as $W_{v_{i,j}}(s, j)$, and $j = 1, 2, \cdots s$.

Step 2: the estimation of noise variance σ^2 is realized from different decomposition levels and directions [12]. The following calculation equation is as follows:

$$\widehat{\sigma}_{b_{i,j}}(s,j) = \frac{\text{median}\left(\left|W_{v_{i,j}}(s,j)\right|\right)}{0.6634}, \tag{3}$$

Step 3: the parameters of the noise threshold are determined. For the original microbial image, $\sigma_{f_{i,j}}$ is the variance of $Z_{f_{i,j}}$. A satisfies a Gaussian distribution whose variance Equation can be described by the following equation. $Z_{v_{i,j}}$ satisfies Gaussian distribution, the calculation equation is as follows:

$$\widehat{\sigma}^2_{v_{i,j}}(s,j) = \frac{\sum_{i,j=1}^{N(j)} Z^2_{v_{i,j}}(s,j)}{N^2(j)}. \tag{4}$$

According to $\sigma^2_{v_{i,j}} = \sigma^2_{f_{i,j}} + \sigma^2_{b_{i,j}}$, it can be determined that:

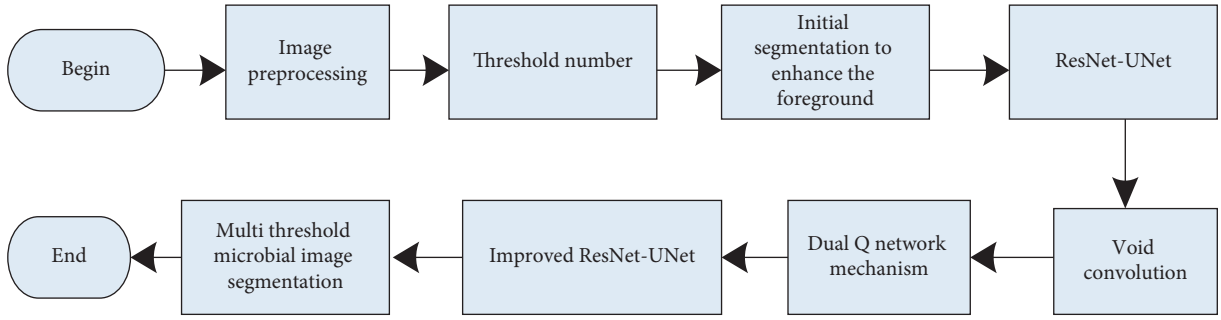


Figure 1: A multithreshold microbial image segmentation process.

$$\widehat{\sigma}_{f_{i,j}}(s,j)=\left[\max\left(\widehat{\sigma}^2_{v_{i,j}}(s,j)-\widehat{\sigma}^2_{b_{i,j}}(s,j)\right)\right]^{1/2}. \quad (5)$$

Step 4: determine the threshold coefficient, expressed as β. Through β, it can update the thresholds of different high-frequency sub-bands in different decomposition layers. The calculation Equation of β is expressed as follows:

$$\beta=\frac{[\log(L_k/j)]^{1/2}}{4.08}, \quad (6)$$

where for wavelet coefficient after j layers of decomposition, L_k is the length of layer k.

Step 5: determine the threshold calculation Equation, described as follows:

$$T(s,j)=\beta\widehat{\sigma}^2_{b_{i,j}}(s,j)\left(\widehat{\sigma}_{f_{i,j}}(s,j)\right)^{-1}. \quad (7)$$

The wavelet coefficient of the denoised image is determined by the wavelet soft threshold processing method, and its equation is described as follows:

$$\widetilde{Z}_{v_{i,j}}(s,j)=WST\left(Z_{v_{i,j}}(s,j)\right). \quad (8)$$

where $WST(\cdot)$ is the soft threshold processing function. The method to obtain the original microbial image $f(i,j)$ is to adopt the inverse wavelet transform based on determining $\widetilde{Z}_{v_{i,j}}(s,j)$.

Based on the above-given process, the noise reduction processing of microbial images can be obtained and realized.

2.3. Determination of Multithreshold for Microbial Images. In the multithreshold segmentation of the microbial image, the selection of threshold number has a direct impact on the segmentation effect of the microbial image [13]. The accurate threshold number can effectively avoid the misclassification of false peaks in the histogram of the microbial image and improve the segmentation effect of the microbial image. For the microbial image, obtain its grey histogram, expressed as h. Suppose U is the number of grey levels contained, i is the grayscale, and $i \in [0, U-1]$. The mean value of frequency is expressed as h_{ave} whose solution can be obtained by $h_{ave}=\text{mean}(h)$. In this paper, the threshold amount of microbial image is determined by the following steps:

Step 1: the denoise-colour or microbial image is obtained, perform the grey operation on it, and obtain the grey histogram.

Step 2: according to the obtained grey histogram, the frequency comparison relative to the grey value is completed in corresponding order. For any grey value, when its frequency is greater than that of the adjacent grey value, the determination of the wave crest can be realized, that is, the frequency relative to the grey value.

Step 3: the histogram peak can be obtained through step 2, but there are some pseudopeaks in it. To ensure that the false peak misclassification problem does not occur, set D as the peak spacing, and consider that there is only one peak in D area.

Step 4: most of the pixels in the microbial image are covered by each target. To accurately determine the wave crest [14], it is necessary to further screen the wave crest, and the screening rule is set that its grey value should be higher than h_{ave}.

Step 5: set A_i as the wave peak, which can be expressed by the frequency of the grey value of i, and A^n refers to the number of wave peaks. The solution Equation is described as follows:

$$A^n=\text{count}\left(A_i^k\right). \quad (9)$$

where $A_i^k - D < A_i^k < A_i^k + D$ and $A_i^k > h_{ave}$. k refers to the frequency, and $k \in [1, 2, \ldots, n]$. A_i^k refers to the wave peak with the frequency as k.

Step 6: thus, the threshold number of microbial images can be obtained, expressed as m. The following solution equation is described as follows:

$$m=A^n-1. \quad (10)$$

2.4. Foreground Target Enhancement of Microbial Image Using Mean Iterative Threshold Segmentation. The mean iterative threshold segmentation method is widely used in multithreshold image segmentation because of its low computational difficulty [15, 16]. Therefore, this paper uses this algorithm to enhance the foreground target of

multithreshold microbial images. Inputting the enhanced microbial image into the image segmentation model of improved deep reinforcement learning can effectively eliminate the characteristics of worthless background and avoid the influence of image background on multithreshold microbial image target segmentation. The algorithm can reduce the complexity of image segmentation and reduce the parameters of the multithreshold microbial image segmentation model and the number of training samples.

The principle of this method is to divide the microbial image into two categories, namely, foreground and background, according to the different grey attributes of microbial image foreground and background. The grey levels of each category show different characteristics. The classification of image pixels is realized by threshold T. This method can distinguish the foreground and background of microbial images. The specific steps are as follows:

Step 1: set the original threshold value T. The average grey value of the microbial image is taken as its initial value.

Step 2: the foreground and background of the microbial image are distinguished by threshold value T, which are expressed as $E1$, $E2$. When the grey value of the pixel is higher than the threshold value T, it is classified into $E1$. Otherwise, it is classified as $E2$.

Step 3: get the solution of the mean grey value of all the pixels of $E1$, $E2$, which are expressed as μ_1 and μ_2.

Step 4: according to the threshold number m of microbial image m determined in Section 2.2, the threshold value T is adjusted by the following equation:

$$T = 2m^{-1}(\mu_1 + \mu_2). \tag{11}$$

Step 5: repeat steps 2, 3, and 4, and the algorithm ends when the average change of the last two times is lower than the set value.

Step 6: thus, the new threshold value T' can be obtained. Through T', the distinction between the foreground and background of the microbial image is realized.

2.5. Multithreshold Microbial Image Segmentation Model Using ResNet-UNet. Resnet-34 is a residual convolution neural network. Its structure has 34 layers and includes three parts: independent convolution module, residual convolution module, mean pooling, and full connection module [17]. The independent convolution module consists of a size of 7×64 channel convolution kernel of 7 and 3×3. The former can realize the convolution operation of the input image, and the latter is used to perform the maximum pool processing of the image. The residual convolution part contains four residual convolution blocks with differences. For any residual convolution block, it is realized by twice convolution, batch regularization, and ReLu. Taking the residual convolution block as a processing unit, complete the residual processing of the residual convolution module with 3, 4, 6, and 3 different times, respectively, and then "smooth" the feature vector. Through mean pooling and full connection, the microbial image is segmented and the segmented image is obtained, to reduce the impact of network gradient disappearance on network performance [18].

The UNet network is composed of the front and rear parts, in which the former is used to obtain microbial image features, and the latter can realize upsampling of image information [19]. According to the dimension difference of the feature map, it can be divided into 5 layers from top to bottom, and 3 layers $\times$ 3. The convolution and ReLu adjustable linear unit are used as the structure of each layer. For the adjacent two layers, the feature extraction structure reduces the size of the feature image and improves the feature dimension through feature cutting and downsampling. The upsampling structure enlarges the feature size of the microbial image through feature difference and upsampling and reduces its dimension. For the same layer, the front and back structure of the net network is completed by feature replication and splicing. Then, after the last processing of the upper sampling structure, execute 1×1. Convolution and ReLu operation can realize the segmentation of microbial image and obtain the target segmentation image. The ResNet-UNet designed in this paper takes ResNet-34 residual convolution neural network as the feature image extraction network of microbial image. Based on the basic principle of the UNet network, the resolution restoration of microbial images is realized through upsampling, feature replication, and splicing.

The specific steps are as follows: take the microbial image processed in Section 2.3 as the input of the ResNet-UNet network, and its size is 512×512. The characteristic map of the microbial image is obtained by ResNet -34 network, but its network structure needs to be adjusted, its independent convolution and residual convolution modules are retained, and the mean pooling and full connection modules are deleted. To ensure the tight connection between ResNet -34 network and the U-Net network, the number of convolution channels of the fourth residual convolution module of the ResNet-UNet network needs to be designed to have the same number of convolution channels as the convolution at the end layer of ResNet -34 network. According to the principle of the UNet network, the difference processing of the image is realized through upsampling to achieve the purpose of image resolution restoration. When there are features with the same size in upsampling and extracting the features of the image, they need to be spliced and fused, which can be realized through feature replication and splicing processing. After that, three tests were carried out twice 3×3 convolutions, batch regularization, and ReLu processing. The sigmoid function is applied to the last layer convolution of the upsampling structure to ensure that the ResNet-UNet network has a strong learning ability and improves the classification accuracy. Then the nonlocal operation is used to realize the integration of microbial image information. Through 1×1 convolution and upsampling processing, the initial segmentation of the multithreshold microbial image is realized.

2.6. Improvement of Multithreshold Microbial Image Segmentation Model

2.6.1. Void Convolution. To ensure that the ResNet-UNet network has a strong feature extraction ability and make the obtained microbial image feature map contain complete image details, this paper improves the ResNet-UNet network, introduces void convolution, and replaces the original convolution based on the original mean pooling and the deletion of the full connection module. While ensuring that the network parameters do not increase, the receptive field of the image is improved so that the details of the microbial image are not lost. For void convolution, it can be regarded as porous convolution, i.e., inserting pixels with a 0-pixel value between the pixels of the original convolution core, to increase the size of the convolution core and improve the image receptive field [20, 21].

2.6.2. Dual Q Network Mechanism. The improved ResNet-UNet network realizes the panoramic perception of multithreshold microbial image features, obtains the microbial image feature map, takes it as the training data set of reinforcement learning, fully considers the dual Q network mechanism, and introduces the target Net as a network copy of the depth Q value. The target value function is determined through the ResNet-UNet network and provided to the depth Q-value network [22]. While continuously adjusting it, the target Net, give play to the decision-making performance of the deep Q-value network and complete the feature extraction of the microbial map and the collaborative operation of parameters. Through the positive feedback mechanism, the error of the collaborative operation process is adjusted, to establish the best segmentation mechanism for the multithreshold microbial image. The objective function gradient function equation is as follows:

$$\begin{aligned}\nabla_\theta J(\pi_\theta) &= \int_A^S \int_A^S \rho^\pi(s) \nabla_\theta \pi_\theta(s,a) Q^\pi(s,a) \mathrm{d}a\mathrm{d}s \\ &= E_{s\cdot\rho^\pi, a\cdot\pi_\theta}\left[\nabla_\theta \log \pi_\theta(a|s) Q^\pi(s,a)\right],\end{aligned} \tag{12}$$

where state information and its collection are represented as s and S, the set of action a is A; the strategy is π, the network parameter is θ. $Q^\pi(s,a)$ is used to describe the expected reward of choosing an action a when the status is s, the probability distribution function is expressed as ρ. According to its value, the determination of the optimal action in the current state can be realized, E is the reward for the best action in the current status.

According to the objective function gradient function of equation (12), the deterministic strategy expression is obtained, use function μ and according to the action information, actions can be obtained. μ is regarded as the optimal action strategy $a_t = \mu(s_t|\theta^\mu)$. The quantitative analysis of multi-threshold microbial image segmentation can be realized by the following equation:

$$J(\mu_\theta) = \int_A^S \rho^\mu(s) \int_A^S r(s, \mu_\theta(s)) \mathrm{d}s = E_{s\cdot\rho^\mu}[r(s, \mu_\theta(s))]. \tag{13}$$

Under competitive conditions, equation (13) has poor stationarity. Therefore, this paper performs the first-order derivation. At this time, the deterministic strategy gradient can be described by the following equation:

$$\begin{aligned}\nabla_\theta J(\mu_\theta) &= \int_A^S \rho^\mu(s) \int_A^S Q^\mu(s,a) \nabla_\theta \mu_\theta(s)|_{a=\mu_\theta} ds \\ &= E_{s\cdot\rho^\pi}\left[Q^\mu(s,a) \nabla_\theta \mu_\theta(s)|_{a=\mu_\theta}\right].\end{aligned} \tag{14}$$

The compatibility of equation (14) is outstanding. It can complete the segmentation of various multithreshold microbial images in continuous adaptive learning and obtain the best-segmented image. Due to the variety and diversity of microorganisms, there are many characteristics of microorganisms [23]. Therefore, in this paper, the strategy network μ is regarded as the actor. The function (s,a) is fitted through the value network and regarded as critical, to complete the recognition and segmentation of different multithreshold microbial images. The expression of the objective function is described by the following equation:

$$J(\theta^\mu) = E_{\theta^\mu} r_1 + E_{\theta^\mu} \gamma r_2 + E_{\theta^\mu} \gamma^2 r_3 + \ldots E_{\theta^\mu} \gamma^2 r_n. \tag{15}$$

2.7. Data Sets and Evaluation Index. Experimental data sets were used Earth Microbiome Project, Human gut microbiomes, and Human Microbiome Project. Earth Microbiome Project: EMP, founded in 2010, is a large-scale crowdsourcing work aimed at analyzing the global microbial community. The general premise is to study from the perspective of the microbial community itself. The researchers used amplicon sequencing, metagenomics, and metabolomics to analyze 200000 samples from these communities, including 100000 microbial images, which is very suitable for multithreshold microbial image segmentation. Human gut microbiomes: human gut microbiomes refer to all microorganisms inhabiting the human gastrointestinal tract. The different roles of intestinal microbiota in human health and disease have been rerecognized. Metagenomic research has changed our understanding of the taxonomy and functional diversity of human microbiota, such as promoting the maturation of the immune system and providing a direct barrier against pathogen colonization. The data set includes a series of studies on human intestinal microbiota, including images, papers, and patents. In this paper, two thousand microbial images are used for experimental tests to verify the application effect of different algorithms. Human Microbiome Project: the overall mission of the HMP is to generate resources to facilitate the characterization of the human microbiota to further our understanding of how the microbiome impacts human health and disease. The initial phase of the project, HMP1, established in 2008, characterized the microbial communities of 300 healthy individuals, across several different sites on the human body: nasal passages, oral cavity, skin, gastrointestinal tract, and urogenital tract. 16S rRNA sequencing was performed to characterize the complexity of microbial communities at each body site and to begin to ask to investigate whether there is a core healthy microbiome.

Metagenomic whole genome shotgun sequencing provided insights into the functions and pathways present in the human microbiome.

A total of five thousand microbial images were selected in the earth microbiome projects data set, human gut microbiomes data set, and Human microbiome project data set, the discount coefficient is 0.95, and the learning rate is 10^{-3}. The value of strategy reward is [−1, 1], and the reward of action is −1. The parameters of the multithreshold microbial image segmentation model are adaptively selected to achieve a better effect of model training. Among them, four thousand images are used to train the segmentation model, and the other images are used for image testing. Different algorithms are applied to the segmentation of microbial images in the data set to analyze the image segmentation effect of this algorithm.

(1) Image denoising ability: due to the interference of noise factors when collecting microbial images, the visual effect of the obtained microbial images is poor. Therefore, the clearer the image after denoising, the stronger the denoising ability.

(2) Peak signal-to-noise ratio (PSNR): it is an index for evaluating image quality. The calculation Equation of this index is as follows:

$$\mathrm{PSNR} = 10 \times \log_{10}\left(\frac{(2^n - 1)^2}{\mathrm{MSE}}\right), \tag{16}$$

where MSE refers to the mean square error between an original image and a denoised image, and n refers to the number of bits of the sample value.

(3) Structural similarity (SSIM) is an important index to judge the contrast of segmented multithreshold microbial images and whether the structure is suitable and complete.

$$\%\ \mathrm{SSIM}(x, y) = \left[l(x, y)^{\alpha}\right]\left[c(x, y)^{\beta}\right]\left[s(x, y)^{\gamma}\right], \tag{17}$$

where x and y refer to the two input microbial images, $l(x, y)$ refers to the brightness comparison function, $c(x, y)$ refers to contrast comparison function, $s(x, y)$ refers to structure comparison function, and α, β, γ are different constants.

(4) After using the algorithm to segment the microbial image, the similarity of the segmented image features compared with the original microbial image can be reflected by feature similarity (FSIM). The value range of FSIM is [0, 1], and the closer its value is to 1, the closer their characteristics are. The calculation Equation of this index is as follows:

$$\mathrm{FSIM} = \frac{\sum_{x\in\Omega} S_L(x) PC_m(x)}{\sum_{x\in\Omega} PC_m(x)}, \tag{18}$$

where $PC_m(x)$ refers to phase consistency judgment function, and $S_L(x)$ refers to coupling functions of different terms.

(5) Loss rate: it refers to the rate of loss in the process of segmentation of microbial image by an algorithm. The calculation equation of this index is as follows:

$$A = (1 - (t - i)) \times 100\%, \tag{19}$$

where t refers to the planned quantity of microbial image experiment, and i refers to the total amount of microbial images.

3. Results and Discussion

In this paper, three microbial images randomly selected from the Earth microbiome projects data set, human gut microbiomes data set, and human microbiome Project data set are denoised first. By comparing the microbial images before and after denoising, the image denoising ability of this algorithm is analyzed as shown in Figure 2.

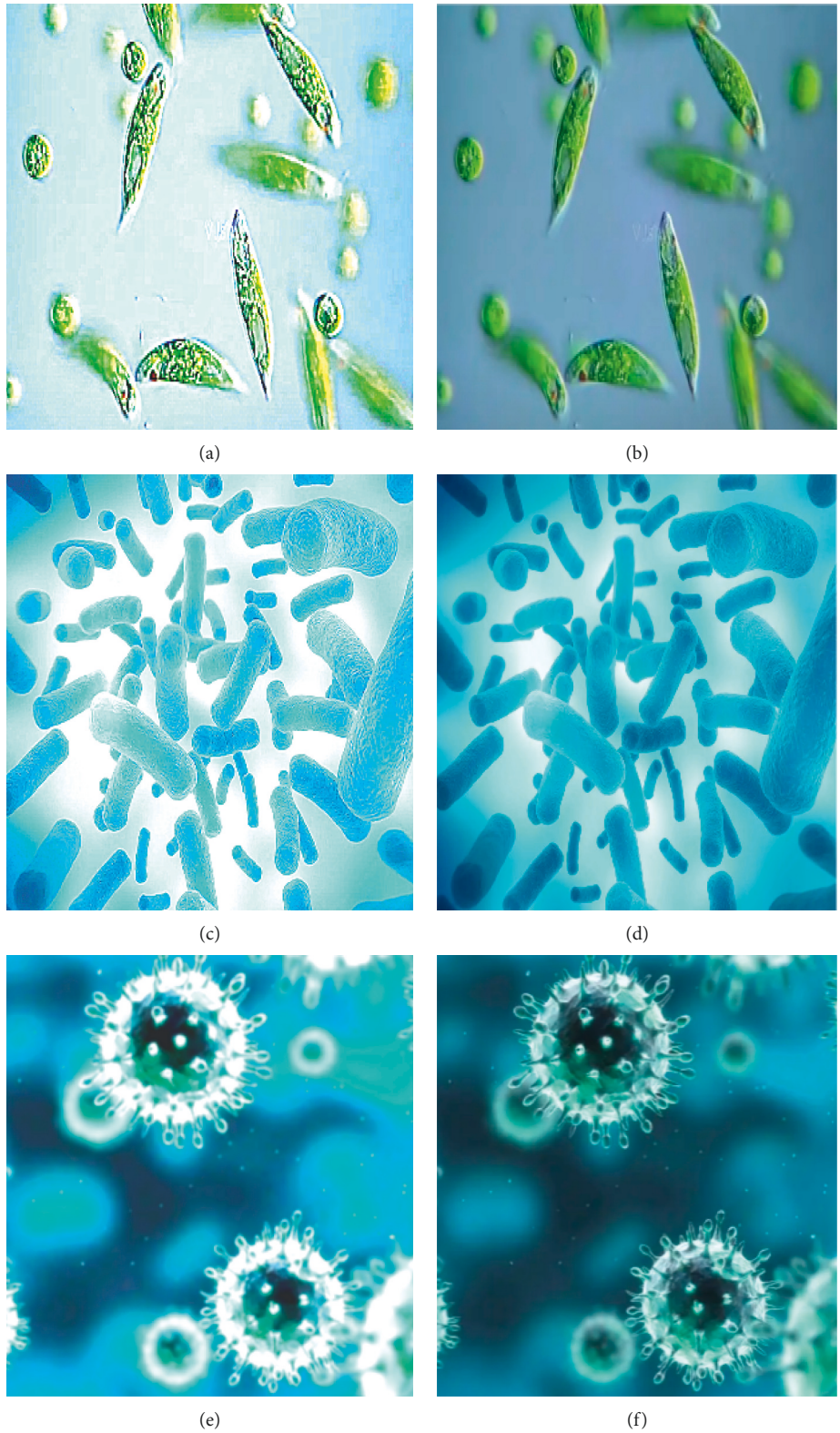
According to Figure 2. The three selected microbial images have a lot of noise, which has the problems of too high brightness and too large contrast, resulting in the loss of some details of the microbial image, which seriously affects the visual effect of the microbial image. After denoising with this algorithm, the supersaturation of the microbial image is suppressed, the brightness and contrast of the image are restored, and the microbial image information is complete and rich in detail. Therefore, it can be determined that the algorithm in this paper has image noise reduction performance, and the noise reduction effect is outstanding.

The threshold number of the microbial image has a direct impact on its image segmentation effect. This algorithm is used to determine the threshold number of Earth Microbiome Project data set, Human gut microbiomes data set, and Human Microbiome Project data set. The Influence of different threshold numbers on the microbial image segmentation effect is analyzed by the PSNR index as shown in Figure 3.

According to the data in Figure 3, PSNR index is used to reflect the visual effect of the segmented microbial image, and the image distortion decreases with the increase of the PSNR index. This algorithm is used to determine the threshold number of microbial image. With the increasing number of thresholds of the microbial image, the PSNR index of the microbial image after segmentation shows an increasing trend. When the threshold number of the Earth Microbiome Project data set increases to 6, the growth trend of the PSNR index tends to be stable. Thus, it can be determined that when the number of thresholds is 6, the image can achieve the best segmentation effect; similarly, it can be determined that the threshold numbers of the Human gut microbiomes data set and Human Microbiome Project data set are 8 and 4, respectively.

According to the relationship between the threshold number of different microbial images and SSIM, the image segmentation effect of this algorithm is analyzed. It is shown in Figure 4.

According to Figure 4, with the increasing number of threshold values of microbial images, SSIM indexes show an upward trend. When the number of threshold values is



(a) (b) (c) (d) (e) (f)

Figure 2: Comparison of microbial images before and after denoising. (a) Microbial image A before denoising. (b) Microbial image B after denoising. (c) Microbial image A before denoising. (d) Microbial image B after denoising. (e) Microbial image A before denoising. (f) Microbial image B after denoising.

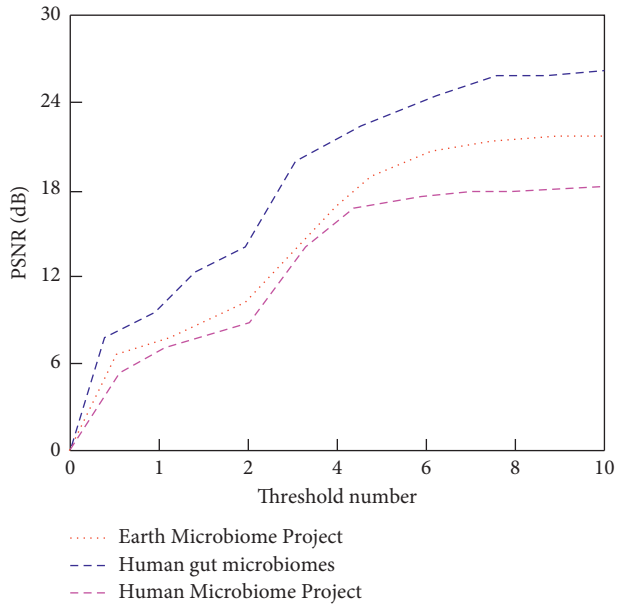


Figure 3: Relationship between threshold number and PSNR.

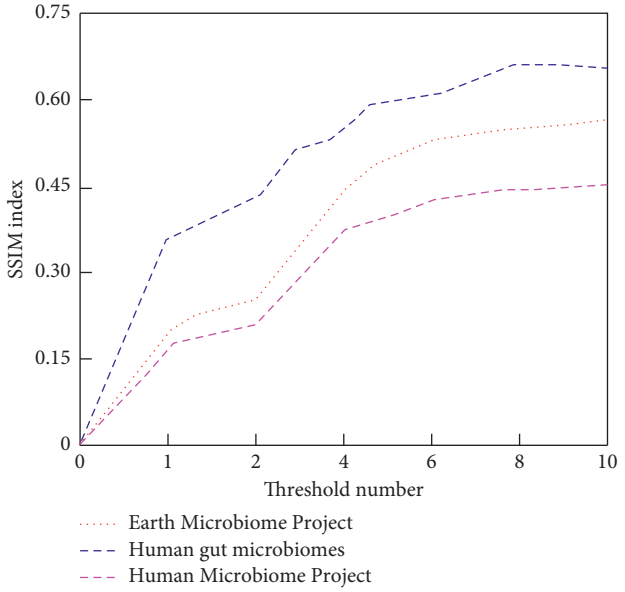


Figure 4: Relationship analysis between threshold number and SSIM.

small, SSIM indexes increase rapidly, the number of threshold values increases from 2 to 4, and SSIM indexes increase significantly. When the number of Human Microbiome Project data set threshold values reaches 4, SSIM indexes increase slowly. When the number of Human gut microbiomes data set threshold reaches 8, the SSIM index remains unchanged. When the number of Human Microbiome Project data set threshold increases to 6, the change in the SSIM index is very small. It can be seen that the threshold number of different data sets is different. When the threshold number of Earth Microbiome Project data set, Human gut microbiomes data set and Human Microbiome Project data set is 6, 8, and 4, respectively, it can have the best contrast.

By analyzing the relationship between FSIM index and threshold number, this paper realizes the analysis of the multithreshold microbial image segmentation effect. It is shown in Figure 5.

The FSIM index showed an upward trend with the increasing number of thresholds. When the number of thresholds increases from 1 to 2, the FSIM increase of the three data sets is small. When the number reaches 4, the three FSIM curves show a rapid growth trend and continue to increase the threshold number. The FSIM index of the Earth Microbiome Project data set changes slightly. After the threshold number of the Human gut microbiomes data set is 8, the FSIM does not change, and after the threshold number of the Human Microbiome Project data set is 6, the FSIM index remains stable. From the above-given analysis, when the threshold number of three data sets is 6, 8, and 4, respectively, the characteristics of the segmented image can be closer to the original microbial image.

Based on the microbial images of the Earth microbiome projects data set, human gut microbiomes dataset, and human microbiome Project data set, the image segmentation accuracy of this algorithm is verified by the loss rate. Therefore, this paper verifies the image segmentation accuracy of this algorithm by analyzing the model loss rate index under the conditions of training and testing. It is shown in Figure 6.

The training image set is used to complete the training of the multithreshold microbial image segmentation model, and the test image set is input into the trained model. The loss rate indexes of training and test image set gradually decrease with the increasing number of iterations. After 60 iterations, the segmentation model begins to converge, and the loss rate curve under the test condition fluctuates at the beginning of the iteration, which is due to the wide variety of microorganisms and the small variety of microorganisms in the training image set. However, this model can better learn the feature information of the microbial image and reduce the loss rate of the model, which is less than 0.05%. The experimental results show that this algorithm can effectively reduce the loss rate of the model and improve the segmentation accuracy of the multithreshold microbial image.

The quantum-based image segmentation algorithm in Li et al. [6], the parallel Lévy grey wolf optimized segmentation algorithm in Li et al. [7], the segmentation algorithm based on the Intelligent Algorithm in Cai et al. [8], the Lévy flight improved multithreshold segmentation algorithm in Jia et al. [9], and the target recognition algorithm of deep reinforcement learning in Sun et al. [10] are used as comparison algorithms. The above-given algorithm and proposed algorithm were used to segment the three data sets, and the advantages of this proposed algorithm for image segmentation were analyzed by normalized absolute error (NAE), mean square error (MSE), and running time indexes. The results are shown in Table 1.

According to the data in Table 1, different algorithms are used to segment three data sets. The number of thresholds determined by the algorithms in Li et al. [6], Li et al. [7], and Cai et al. [8] is less than other algorithms, which seriously affects the image segmentation performance of the

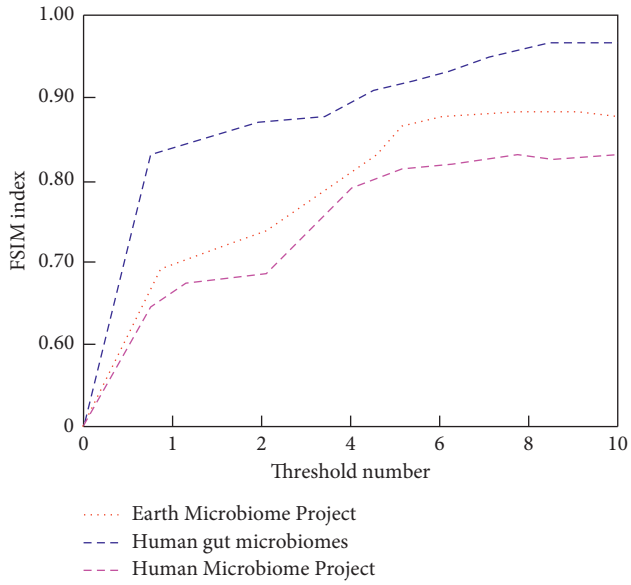


Figure 5: Relationship analysis between threshold number and FSIM.

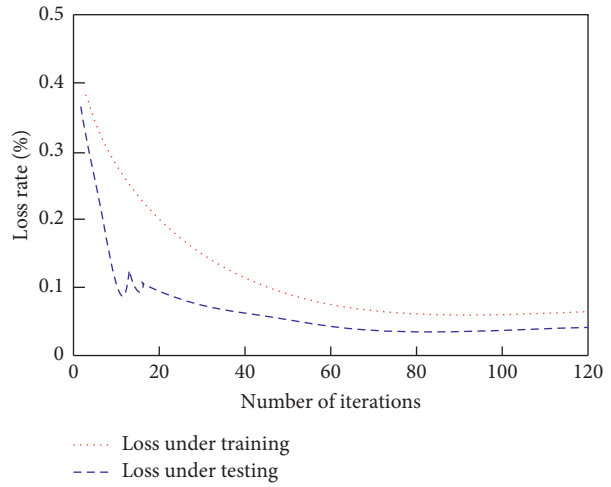


Figure 6: Image segmentation accuracy analysis of proposed algorithm.

Table 1: Comparison of segmentation performance of different algorithms in different data sets.

Algorithms	Data sets	Threshold number	NAE	MSE	Running time (s)
Li et al. [6]	Earth microbiome project	154	0.976	78.432	15.886
	Human gut microbiomes	137	0.925	70.537	17.1406
	Human microbiome project	107	0.826	64.311	15.34
Li et al. [7]	Earth microbiome project	134, 170	0.77	77.532	10.0427
	Human gut microbiomes	52, 113, 170, 181	0.646	68.443	10.7601
	Human microbiome project	73, 110, 136	0.561	62.431	9.7423
Cai et al. [8]	Earth microbiome project	59, 148, 152	0.75	78.531	4.7921
	Human gut microbiomes	58, 125, 186, 172	0.621	75.811	4.7178
	Human microbiome project	75, 128	0.561	69.583	4.6341
Jia et al. [9]	Earth microbiome project	61, 141, 165, 173, 224, 100	0.476	60.830	5.3121
	Human gut microbiomes	53, 120, 181, 169, 212, 141, 137, 101	0.501	57.041	5.6212
	Human microbiome project	83, 112, 150, 136	0.361	40.560	5.2353
Sun et al. [10]	Earth microbiome project	48, 138, 155, 182, 235, 100	0.503	53.880	4.6748
	Human gut microbiomes	69, 111, 177, 163, 202, 125, 114, 98	0.457	46.336	5.2346
	Human microbiome project	88, 108, 124, 147	0.357	38.483	4.2728
Proposed	Earth microbiome project	63, 134, 185, 197, 225, 100	0.316	34.015	4.0863
	Human gut microbiomes	66, 118, 181, 177, 215, 116, 128, 103	0.21	30.586	4.1408
	Human microbiome project	91, 105, 132, 148	0.288	29.043	3.804

algorithm. This is the reason why its NAE and MSE indexes are higher than other algorithms, and the operation efficiency of the above-given algorithms is low. The algorithms in Jia et al. [9], Sun et al. [10], and the algorithm in this paper is used to segment three data sets, the same threshold number can be determined. However, the NAE and MSE indexes of this algorithm are the lowest and can realize the efficient operation of image segmentation. The reason is that this algorithm can quickly and accurately learn the characteristic information of microbial image through the improved image segmentation model of deep reinforcement learning, and the loss of microbial image information is low. The experimental results show that the microbial image segmentation ability of this algorithm is outstanding and has significant operational advantages.

4. Conclusions

To address the problem of poor performance of multithreshold microbial image segmentation existing in traditional methods, this paper proposes a multithreshold microbial image segmentation algorithm using improved deep reinforcement learning. The results show that the proposed algorithm can improve the accuracy of multithreshold microbial image segmentation and has application performance. It can effectively address the problems of loss of image detail information and poor image visualization caused by noise interference in microbial images. The improved deep reinforcement learning segmentation model can efficiently learn microbial features of different morphologies and shapes and has better generalization performance. However, this paper did not verify the image detail information and image visual effect during the experimental process, which led to a decrease in the persuasive power of the experimental results. Therefore, this problem will be investigated in depth in the future as a way to improve the comprehensive performance of the proposed algorithm.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] Y. Wang, D. Chang, and Y. Zhao, "A new blind image denoising method based on asymmetric generative adversarial network," *IET Image Processing*, vol. 15, no. 6, pp. 1260–1272, 2021.

[2] C. Shang, D. Zhang, and Y. Yang, "A gradient-based method for multilevel thresholding," *Expert Systems with Applications*, vol. 175, no. 1, pp. 114845–114855, 2021.

[3] G. Toz and P. Erdoğmuş, "A novel hybrid image segmentation method for detection of suspicious regions in mammograms based on adaptive multi-thresholding (HCOW)," *IEEE Access*, vol. 9, pp. 85377–85391, 2021.

[4] J. Meng, X. Lv, L. Fu, and Q. Wu, "Rice disease image recognition based on improved multi-scale stack autoencoder," *Journal of Agricultural Science*, vol. 13, no. 1, pp. 18–29, 2020.

[5] W. Wang, W. Wang, and Z. Hu, "Retinal vessel segmentation approach based on corrected morphological transformation and fractal dimension," *IET Image Processing*, vol. 13, no. 13, pp. 2538–2547, 2019.

[6] P. Li, T. Shi, Y. Zhao, and A. Lu, "Design of threshold segmentation method for quantum image," *International Journal of Theoretical Physics*, vol. 59, no. 2, pp. 514–538, 2020.

[7] M. Q. Li, L. P. Xu, S. Gao, N. Xu, and B. Yan, "Remote sensing image segmentation based on a robust fuzzy C-means algorithm improved by a parallel Lévy grey wolf algorithm," *Applied Optics*, vol. 58, no. 17, pp. 4812–4822, 2019.

[8] Y. X. Cai, Y. Y. Xu, T. R. Zhang, and D. Li, "Threshold image target segmentation technology based on intelligent algorithms," *Computer Optics*, vol. 44, no. 1, pp. 137–141, 2020.

[9] H. Jia, X. Peng, W. Song, C. Lang, Z. Xing, and K. Sun, "Multiverse optimization algorithm based on Lévy flight improvement for multithreshold color image segmentation," *IEEE Access*, vol. 7, pp. 32805–32844, 2019.

[10] W. Sun, D. Yan, J. Huang, and C Sun, "Small-scale moving target detection in aerial image by deep inverse reinforcement learning," *Soft Computing*, vol. 24, no. 8, pp. 5897–5908, 2020.

[11] M. Wang, G. Pan, and Y. Liu, "A novel imperialist competitive algorithm for multithreshold image segmentation," *Mathematical Problems in Engineering*, vol. 2019, no. 1, pp. 1–18, 2019.

[12] X. Wang, X. Zhao, Y. Zhu, and X. Su, "NSST and vector-valued C–V model based image segmentation algorithm," *IET Image Processing*, vol. 14, no. 8, pp. 1614–1620, 2020.

[13] Y. Jin, X. B. Jiang, Z. K. Wei, and Y. Li, "Chest X-ray image denoising method based on deep convolution neural network," *IET Image Processing*, vol. 13, no. 11, pp. 1970–1978, 2019.

[14] S. Dhar and M. K. Kundu, "Interval type-2 fuzzy set and theory of weak continuity constraints for accurate multiclass image segmentation," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 9, pp. 2151–2163, 2020.

[15] S. Ghosh, I. Das, N. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation," *ACM Computing Surveys*, vol. 52, no. 4, pp. 1–35, 2020.

[16] J. Qin, X. Shen, F. Mei, and Z. Fang, "An Otsu multithresholds segmentation algorithm based on improved ACO," *The Journal of Supercomputing*, vol. 75, no. 2, pp. 955–967, 2019.

[17] J. Xiong, D. Yu, S. Liu, L. Shu, X. Wang, and Z. Liu, "A review of plant phenotypic image recognition technology based on deep learning," *Electronics*, vol. 10, no. 1, pp. 81–92, 2021.

[18] M. M. Omwenga, D. Wu, Y. Liang, L. Yang, D. Huston, and T. Xia, "Cognitive GPR for subsurface object detection based on deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11594–11606, 2021.

[19] H. Wang, Z. Yang, W. Zhou, and D. Li, "Online scheduling of image satellites based on neural networks and deep reinforcement learning," *Chinese Journal of Aeronautics*, vol. 32, no. 4, pp. 1011–1019, 2019.

[20] R. Ke, A. Bugeau, N. Papadakis, M. Kirkland, P. Schuetz, and C.-B. Schönlieb, "Multi-task deep learning for image segmentation using recursive approximation tasks," *IEEE Transactions on Image Processing*, vol. 30, pp. 3555–3567, 2021.

[21] H. Jeckel and K. Drescher, "Advances and opportunities in image analysis of bacterial cells and communities," *FEMS Microbiology Reviews*, vol. 45, no. 4, p. fuaa062, 2021.

[22] J. Loo, M. F. Kriegel, M. M. Tuohy et al., "Open-source automatic segmentation of ocular structures and biomarkers of microbial keratitis on slit-lamp photography images using deep learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 88–99, 2021.

[23] T. Liu, J. Liu, Y. Ma et al., "Spatial feature fusion convolutional network for liver and liver tumor segmentation from CT images," *Medical Physics*, vol. 48, no. 1, pp. 264–272, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 6475808, 13 pages
https://doi.org/10.1155/2022/6475808

Research Article

Framework for Classification of Chest X-Rays into Normal/COVID-19 Using Brownian-Mayfly-Algorithm Selected Hybrid Features

Roshima Biju,[1] Warish Patel,[2] K. Suresh Manic ,[3] and Venkatesan Rajinikanth [4]

[1]*Research Scholar, Department of Computer Science Engineering, Parul University, Vadodara, Gujarat 391760, India*
[2]*Department of Computer Science Engineering, Parul University, Vadodara, Gujarat 391760, India*
[3]*Department of Electrical and Communication Engineering, National University of Science and Tech, Muscat, Oman*
[4]*Department of Electronics and Instrumentation Engineering, St. Joseph's College of Engineering, OMR, Chennai 600119, Tamil Nadu, India*

Correspondence should be addressed to Venkatesan Rajinikanth; v.rajinikanth@ieee.org

Received 5 March 2022; Revised 5 June 2022; Accepted 18 June 2022; Published 11 August 2022

Academic Editor: Xiaofeng Li

The improvements in computation facility and technology support the development and implementation of automatic methods for medical data assessment. This study tries to extend a framework for efficiently classifying chest radiographs (X-rays) into normal/COVID-19 class. The proposed framework consists subsequent phases: (i) image resizing, (ii) deep features extraction using a pretrained deep learning method (PDLM), (iii) handcrafted feature extraction, (iv) feature optimization with Brownian Mayfly-Algorithm (BMA), (v) serial integration of optimized features, and (vi) binary classification with 10-fold cross validation. In addition, this work implements two methodologies: (i) performance evaluation of the existing PDLM in the literature and (ii) improving the COVID-19 detection performance of chosen PDLM with this proposal. The experimental investigation of this study authenticates that the effort performed using pretrained VGG16 with SoftMax helped get a classification accuracy of >94%. Further, the research performed using the proposed framework with BMA selected features (VGG16 + handcrafted features) helps achieve a classification accuracy of 99.17% on the chosen X-ray image database. This outcome proves the scientific importance of the implemented framework, and in the future, this proposal can be adopted to inspect the clinically collected X-rays.

1. Introduction

Due to different grounds, the occurrence speed of diseases in humankind is gradually rising, and timely screening and treatment are necessary to reduce the infection/death rates. In the current era, many advanced life-saving facilities are available in healthcare centers to treat individuals suffering from infectious/acute diseases. However, even though enough investigative and healing services are conveniently accessible to the individuals, the occurrence rate of life-threatening communicable infections is gradually rising, which causes more medical burden worldwide [1–3].

The contagious infections caused by viruses/bacteria commonly infect a sizeable human group, and early recognition and management is the solitary remedy to manage its increase. Recently, contagious infection named COVID-19 infected many individuals globally and are the prime reason for increased death rates in the years 2020 and 2021. Due to its severity and spreading speed, the World Health Organization (WHO) confirmed it as a pandemic in early 2020 [4, 5]. COVID-19 is caused by the SARS-CoV-2 virus, which creates mild to harsh pneumonia in individuals based on their immunity intensity. Even though the patient is vaccinated and following the COVID-19 protocol suggested by WHO, its infection rapidity is still unmanageable due to the speedy alteration in the virus.

Self/doctor can diagnose the symptom of COVID-19, and the clinical level screening of this disease involves

(i) collection of samples from individuals and execution of the reverse transcription-polymerase chain reaction (RT-PCR) using permitted clinical practice and (ii) radiological image-based lung screening. First, radiology-supported lung imaging is performed in a controlled environment, in which the infection in the lung is diagnosed with chest X-ray or computed tomography (CT) images. Then, the pulmonologist examines the collected X-ray/CT images to detect the severity of the infection, decision making, and treatment execution to cure the disease.

In hospitals, the usages of the X-ray/CT are prevalent to examine lung infection and compared to the CT, the implementation of the X-ray is simple and cost-effective. Hence, most of the initial level lung screening considers X-ray images. The disease and its harshness can be easily detected when the radiologist/pulmonologist examines it. Several computerized screening procedures for X-ray images are discussed in the literature, and these works confirm that X-ray-supported lung infection screening helps achieve a better diagnosis.

Several X-ray image examination methods are proposed and implemented in the literature using the machine learning (ML) schemes and deep learning (DL) methods. The existing works helped achieve better detection accuracy. However, the integration of the ML and DL approaches is minimal, and this scheme will help to achieve improved detection accuracy when the clinical-grade X-ray image is assessed. This research aims to develop a DL framework for automatic detection of COVID-19 in chest X-ray images. In order to achieve a better disease detection, this framework employs the following stages: (i) Collection and preprocessing of X-ray images; (ii) evaluating the performance of pretrained deep learning method (PDLM) and finding the appropriate practice to screen the X-ray database; (iii) mining of deep features from the X-ray; (iv) mining of handcrafted features (HF) using chosen procedures; (v) feature selection with Brownian Mayfly-Algorithm (BMA) and serial feature integration; and (vi) classification and validation of the performance of proposed COVID-19 screening framework.

This research primarily executes PDL scheme-supported X-ray evaluation and identifies the infection screening performance based on the attained metrics. The initial study confirms that the COVID-19 detection accuracy achieved by VGG16 is better (>95%) than other PDL schemes. Hence, the VGG16 supported framework is considered, and then its performance is enhanced by serially integrating the HF, such as local binary pattern (LBP) and PHOG. In order to avoid the overfitting problem, these features are then optimized by the BMA. Then, the necessary hybrid feature vector (Deep + HF) is generated, and it is then considered to train and validate the binary classifiers with 10-fold cross-validation. This study considers 4800 (2400 normal and 2400 COVID-19) X-ray images for the evaluation, in which 90% images are considered for training and 10% are considered for the validation. The experimental outcome of this study confirms that the proposed technique helps get a classification accuracy of 99.17% with the K-nearest neighbor (KNN) classifier.

The novelty and the merits of this research include

(i) Implementation of Brownian Mayfly-Algorithm (BMA) based deep and handcrafted feature optimization on improving the detection accuracy without the overfitting

(ii) Precise COVID-19 detection in X-ray images using hybrid features with 10-fold cross-validation

The upcoming sections of this work are demonstrated as follows: Section 2 shows the literature review; Section 3 presents the methodology; and Sections 4 and 5 demonstrate achieved results and conclusion of the presented work, respectively.

2. Related Research

Chest X-ray supported lung infection detection is a clinically accepted methodology in which the combined report of the radiologist and pulmonologist are considered to evaluate the disease in the lung to plan and implement the necessary treatment to cure the patient. The computer algorithm-supported X-ray examination is one of the widely accepted procedures. Hence, several PDL schemes have been implemented to examine the harshness of COVID-19 infection in patients. The employed PDL schemes help categorize the available X-ray images into normal and disease classes with better accuracy. This procedure is an essential process when a mass screening procedure is implemented, and this considerably reduces the COVID-19 detection burden when more patients are to be screened. Table 1 depicts some chosen deep-learning assisted COVID-19 infection procedures found in the recent literature.

The earlier works in the literature authenticate that the combination of deep and HF assists to acquire better disease detection accuracy [19]. The above table confirms that the maximum detection accuracy presented in the earlier work is 99.02% [13]. This work considered the hybrid feature-based X-ray classification to improve the detection accuracy. Hence, in this work, the classification of X-rays into normal/COVID-19 is implemented using the BMF algorithm optimized VGG16's features and the optimally selected LBP and PHOG features. The experimental outcome of this research confirms that the presented work helps to get better detection accuracy than the works considered in Table 1.

3. Methodology

This section represents the developed structure to examine the selected X-ray database. Also, it outlines the different procedures implemented to distinguish normal/COVID-19 class X-rays.

3.1. Framework. Figure 1 depicts the proposed framework developed to sense the COVID-19 in chosen test X-ray pictures. In this effort, the necessary images are primarily collected and resized into $224 \times 224 \times 1$ pixels, and these imagery are afterward considered to extract the deep features (DF) and HF. The DF mining is initially achieved using the pretraining schemes, and every scheme helps to get a one-dimensional (1D) feature vector of size $1 \times 1 \times 1000$. This feature vector is

Table 1: Summary of chosen methodologies employed to detect COVID-19 from X-ray images.

Reference	Methodology employed	Performance metrics (%)		
		Accuracy	Sensitivity:	Specificity:
Narin et al. [6]	The performance of pretrained deep-learning scheme supported COVID-19 detection is demonstrated using X-ray images	98.00	—	—
Apostolopoulos and Mpesiana [7]	Convolutional-neural-network (CNN) with transfer learning is employed to examine X-ray to detect COVID-19	93.48	92.85	98.75
Chouhan et al. [8]	Transfer learning based deep-learning scheme is employed to recognize pneumonia in X-rays	96.39	—	—
Stephen et al. [9]	Automatic detection of pneumonia in X-ray is performed using transfer learning	95.00	—	—
Liang and Zheng [10]	Classification of paediatric pneumonia in X-ray is achieved using pretrained CNN	90.00	—	—
Nour et al. [11]	Detection of COVID-19 in X-ray is discussed with deep features and Bayesian optimization	98.97	89.39	99.75
Brunese et al. [12]	Implementation of explainable deep-learning scheme to detect pulmonary abnormality and COVID-19 is presented with X-ray	96.00	96.00	98.00
Ardakani et al. [13]	A detailed analysis of ten widely adopted deep learning methods is discussed and their performance in detection the COVID-19 is demonstrated	99.02	98.04	100
Jaiswal et al. [14]	DenseNet201 supported detection of COVID-19 in X-ray is demonstrated with transfer learning technique	96.25	96.29	96.21
Ucar and Korkmaz [15]	This work proposed deep Bayes-SqueezeNet to detect COVID-19 in X-rays	98.26	99.13	—
Saiz and Barandiaran [16]	This work presented pretrained VGG16 based COVID-19 in X-rays	94.92	94.92	92.00
Panwar et al. [17]	This work demonstrated nCOVnet based identification of COVID-19 in X-rays	88.10	97.62	78.57
Waheed et al. [18]	This work developed a novel deep-learning scheme CovidGAN to detect COVID-19 in X-ray pictures	95.00	90.00	97.00
Kannan et al. [19]	This work demonstrated a study with various pretrained scheme supported COVID-19 classification, and the result of this study confirms that the VGG16 along with K-nearest neighbor (KNN) helped to get better accuracy with Deep + HF	96.48	95.56	95.37

adopted to confirm the SoftMax (SM) classifier's merit on the chosen test images. The initial experiment proves that the COVID-19 detection accuracy of VGG16 with SM classifier is better than AlexNet, VGG19, ResNet18, ResNet50, and ResNet101 schemes. Furthermore, the initial study confirmed that the VGG16 provides better results on the chosen data than other methods.

Later, the HF, such as LBP and PHOG, is extorted from the test imagery. The collected DF and HF are then reduced by the BMF algorithm, the selected features are then serially integrated, and the classification study is repeated. This experimental outcome authenticates that the proposed framework helped to accomplish a categorization accuracy of 99.17% for the chosen X-ray database. The various stages of this framework are clearly depicted in Figure 1, and the outcome of the framework is more significant compared to other results presented in Table 1.

3.2. X-Ray Database. The merit of the planned COVID-19 detection framework is then tested and validated with benchmark X-ray images found in the literature. In this scheme, the necessary test images for this study are collected from the following locations [20, 21]. During this study, 4800 test images were considered for the assessment. Table 2 presents the information about the images (total, training, and validation), and Figure 2 presents the sample test images of the chosen database. Finally, all the considered PDLM are tested with the considered database and the results are analyzed.

3.3. Deep-Features Mining. The concert of the planned framework relies mainly on the deep features obtained from the chosen PDLM [22]. In this work, the well-known PDLM, such as AlexNet, VGG16, VGG19, ResNet18, ResNet50, and ResNet101, are considered for the evaluation. During this task, the following parameter setting is implemented on all the chosen PDLM: initial-weights = imageNet features, total epochs = 100, optimizer = Adam, pooling = max/average (AVG), activation for hidden-layer = ReLu, classifier-activation = sigmoid, training images = 2160, validation images = 240, and classifier validation = 10-fold.

Before employing the chosen PDLM to assess the images, an image augmentation is employed to increase the number of images for training the PDLM scheme. The augmentation of images is achieved with horizontal flip, vertical flip, rotation = 25^o, zoom = 0.4, width shift = 0.4, height shift = 0.4, and shear range = 0.3, and this method assists the PDLM in distinguishing the image information correctly.

This scheme helps to extract $1 \times 1 \times 1000$ deep features from every PDLM, and this value is mathematically depicted in the following formula:

$$\begin{aligned} &\text{Deep} - \text{feature}_{(1\times1\times1000)} \\ &\quad = \text{Deep}_{(1,1)}, \text{Deep}_{(1,2)}, \ldots, \text{Deep}_{(1,1000)}. \end{aligned} \tag{1}$$

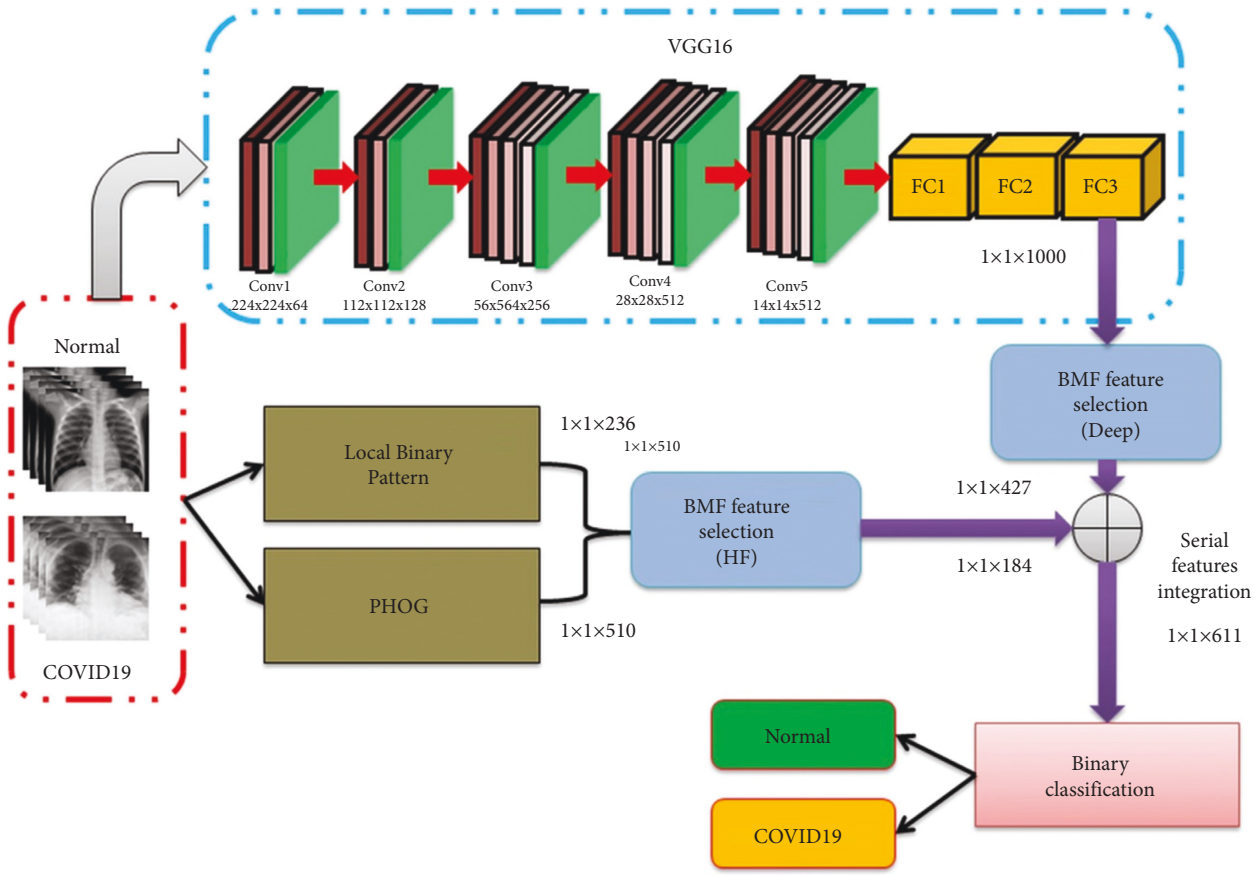


Figure 1: Developed framework to detect normal/COVID-19 X-ray using hybrid features.

Table 2: Dataset considered in this framework to test the performance of proposed scheme.

Class	Dimension	Total image	Training image	Testing image
Normal	$224 \times 224 \times 3$	2400	2160	240
COVID-19	$224 \times 224 \times 3$	2400	2160	240

3.4. Handcrafted-Feature Mining. In this work, the HF is mined using the LBP [23, 24] with varied weights and the PHOG [24, 25] with various bins, and the discussion about these procedures can be found in earlier research works. The essential HF is then mined using local binary patterns (LBP) with different weights ($W = 1$ to 4) and PHOG with various bins (Bin = 1 to 3). The outcome attained with LBP is depicted in Figure 3, in which Figures 3(a)–3(d) present the outcomes with various values of weights ($W = 1$ to 4) on a chosen test X-ray.

A similar practice is then implemented with the PHOG, and the achieved features for bin1, bin2, and bin3 are presented in Figure 4.

$$\mathrm{LBP}_{w1\,(1\times1\times59)} = W1_{(1,1)}, W1_{(1,2)}, \ldots, W1_{(1,59)}, \tag{2}$$

$$\mathrm{LBP}_{w2\,(1\times1\times59)} = W2_{(1,1)}, W2_{(1,2)}, \ldots, W2_{(1,59)}, \tag{3}$$

$$LBP_{w3\,(1\times1\times59)} = W3_{(1,1)}, W3_{(1,2)}, \ldots, W3_{(1,59)}, \tag{4}$$

$$\mathrm{LBP}_{w4\,(1\times1\times59)} = W4_{(1,1)}, W4_{(1,2)}, \ldots, W4_{(1,59)}, \tag{5}$$

$$\mathrm{LBP}_{\mathrm{total}\,(1\times1\times236)} = \mathrm{LBP}_{w1\,(59)} + \mathrm{LBP}_{w2\,(59)} + \mathrm{LBP}_{w3\,(59)} + \mathrm{LBP}_{w4\,(59)}, \tag{6}$$

$$\mathrm{PHOG}_{\mathrm{BIN1}\,(1\times1\times85)} = \mathrm{BIN1}_{(1,1)}, \mathrm{BIN2}_{(1,2)}, \ldots, \mathrm{BIN3}_{(1,85)}, \tag{7}$$

$$\mathrm{PHOG}_{\mathrm{BIN1}\,(1\times1\times170)} = \mathrm{BIN1}_{(1,1)}, \mathrm{BIN2}_{(1,2)}, \ldots, \mathrm{BIN3}_{(1,170)}, \tag{8}$$

$$\mathrm{PHOG}_{BIN1\,(1\times1\times255)} = \mathrm{BIN1}_{(1,1)}, \mathrm{BIN2}_{(1,2)}, \ldots, \mathrm{BIN3}_{(1,255)}, \tag{9}$$

$$\mathrm{PHOG}_{\mathrm{total}\,(1\times1\times510)} = \mathrm{PHOG}_{BIN1\,(85)}, +\mathrm{PHOG}_{BIN2\,(1,170)} + \mathrm{PHOG}_{BIN3\,(1,255)}, \tag{10}$$

$$\mathrm{HF}_{(1\times1\times746)} = \mathrm{LBP}_{\mathrm{total}\,(1\times1\times236)} + \mathrm{PHOG}_{\mathrm{total}\,(1\times1\times510)}. \tag{11}$$

In this research, equation (11) is considered as the HF, and the optimized HF is then combined with the deep-

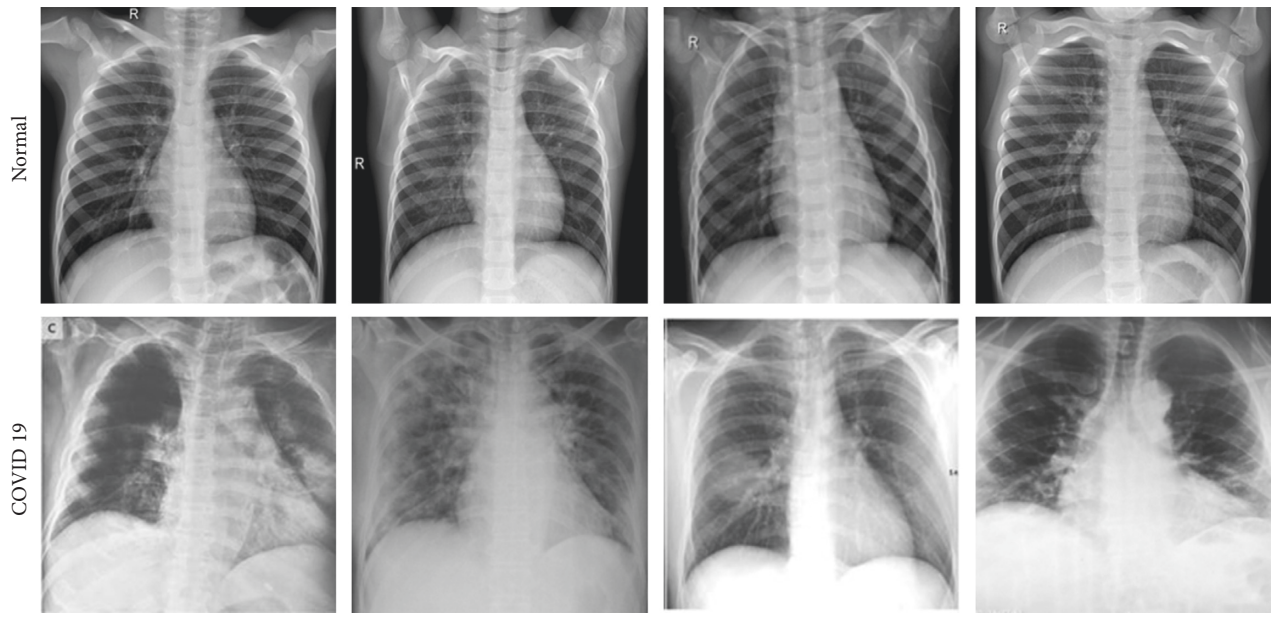


Figure 2: Sample test images of normal/COVID-19 class in the chosen database.

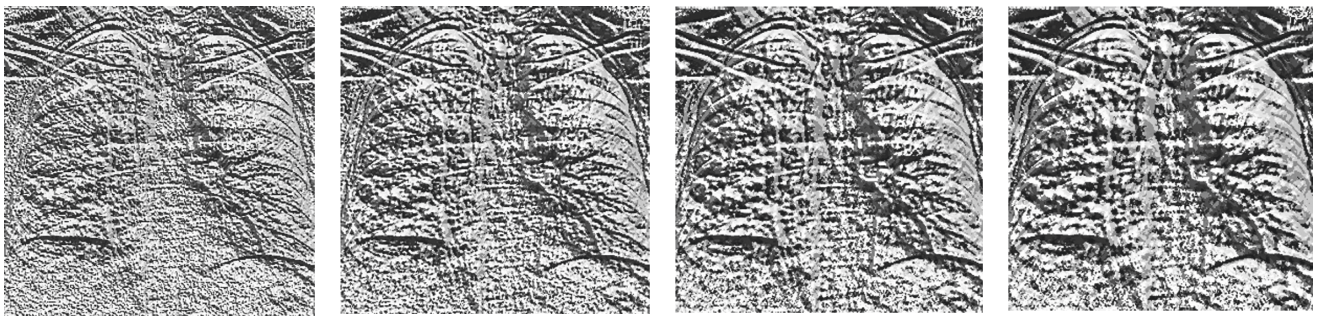


Figure 3: Sample LBP pattern generated for COVID-19 image: (a) $W = 1$, (b) $W = 2$, (c) $W = 3$, and (d) $W = 4$.

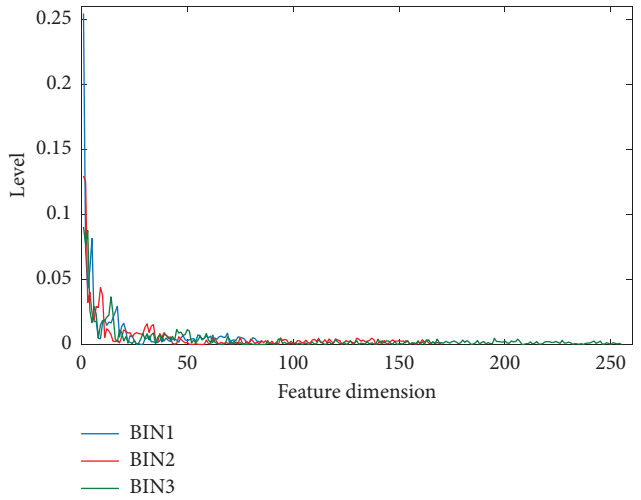


Figure 4: PHOG feature obtained for a sample COVID-19 image. The features obtained with these methods are presented in equations (2)–(11).

feature to get the deep + HF, which helps to classify the X-ray with better accuracy.

3.5. Feature Selection Using Brownian Mayfly-Algorithm. Feature selection is a prime task in this work, and as discussed in the earlier work, the deep and HF are optimized using the Mayfly Algorithm (MA) [26]. The MA is a nature-inspired algorithm invented by combining Firefly, particle swarm, and genetic algorithm. A levy-flight search operator guides the traditional MA, and in this work, the proposed MA is driven by a Brownian operator. The search process found in Brownian Mayfly-Algorithm (BMA) is smoothly compared to the traditional approach [27, 28]. Figure 5 depicts the working of the proposed BMA, in which Figure 5(a) illustrates the Brownian walk search process for a single Mayfly. The various stages (Stages 1 to 3) are depicted in Figures 5(b)–5(d).

The description of the MA is as follows.

Let, MA includes identical male (M) and female (F) flies, which are randomly distributed in search space. Let these flies are demoted as $i = 1, 2, ..., N$. During the examination task, each fly is authorized to fuse close to the optimum location (G_{best}). After reaching G_{best}, male-fly (M) is permitted to stay in G_{best}. This process is depicted in Figure 5(b).

This process is shown in equations (12) and (13).

$$P_i^{t+1} = P_i^t + V_i^{t+1}, \tag{12}$$

$$V_{i,j}^{t+1} = V_{i,j}^t + C_1 * e^{-\beta D_p 2\left(p_{\text{best}_{i,j}} - P_{i,j}^t\right)} + C_2 * e^{-\beta D_g 2\left(G_{\text{best}_{i,j}} - P_{i,j}^t\right)}, \tag{13}$$

where P_i^t and P_i^{t+1} are initial and ending spots and V_i^{t+1} and $V_{i,j}^{t+1}$ are initial and ending velocities. $C_1 = 1$ and $C_2 = 1.5$ denote local and global learning constraints. Other parameters are assigned as follows. $\beta = 2$, D_p and D_g are the Cartesian distance among flies.

During the relocation, every M will achieve G_{best} and executes a velocity update to attract female-fly (F) with the help of nuptial-dance.

The velocity update at this condition is shown in the following equation:

$$V_{i,j}^{t+1} = V_{i,j}^t + d * R, \tag{14}$$

where nuptial-dance (d) = 5 and R = random numeral [−1,1].

When the search by M is finished, every F is permitted to find M, which reached G_{best} and this process is depicted in Figure 5(c).

The expression for female-fly update given in equations (15) and (16).

$$P_i^{'t+1} = P_i^{'t} + V_i^{'t+1}, \tag{15}$$

$$F_{i,j}^{'t+1} = \left\{ \begin{array}{c} F_{i,j}^{'t} + C_2 e^{-\beta D_{mf}^2}\left(M_{i,j}^t - Y_{i,j}^t\right) \text{ if } \mathrm{O_{max}}(\mathrm{F_i}) > \mathrm{O_{max}}(\mathrm{M_i}) \\ F_{i,j}^{'t} + W * r \text{ if } \mathrm{O_{max}}(\mathrm{F_i}) \le \mathrm{O_{max}}(\mathrm{M_i}) \end{array} \right\}, \tag{16}$$

where $\mathrm{O_{max}}$ is the objective-value.

When the search process continues, each F will find the finest, the offspring generation happens, and other information on MA can be found in literature, and this procedure is depicted in Figure 5(d).

Figure 6 presents the feature optimization process. During the feature reduction process, the BMA is permitted to explore the deep/HF to reduce the value based on the Cartesian distance (CD). This process compares the features of the normal/COVID-19 class images and helps to find the features whose CD is large. The features with lesser CD are discarded, and this procedure is depicted graphically in Figure 4. This procedure helps to find the optimal features (deep-features = $1 \times 1 \times 417$ and HF = $1 \times 1 \times 241$) and the selected features are then combined to get a hybrid feature vector ($1 \times 1 \times 658$), which is considered to train and validate the classifiers.

3.6. Classification and Validation. In the proposed research, initially, the SoftMax classifier is employed to recognize the classifier performance with the selected PDLM, and after achieving the results, the performance of other binary classifiers, like decision tree (DT), random forest (RF), aïve Bayes (NB), K-nearest neighbor (KNN), and support vector machine (SVM) with linear kernel is considered and the attained results are measured. The merit of planned practice is measured using the essential measures, like true ositive (TP), false negative (FN), true negative (TN), and false positive (FP), accuracy (AC), precision (PR), sensitivity (SE), specificity (SP), F1-score (F1S), and negative predictive value (NPV) are obtained from these values.

The mathematical expression for these measures is presented in equations (17) to (22) [29–32].

$$\mathrm{AC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{17}$$

$$\mathrm{PR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{18}$$

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{19}$$

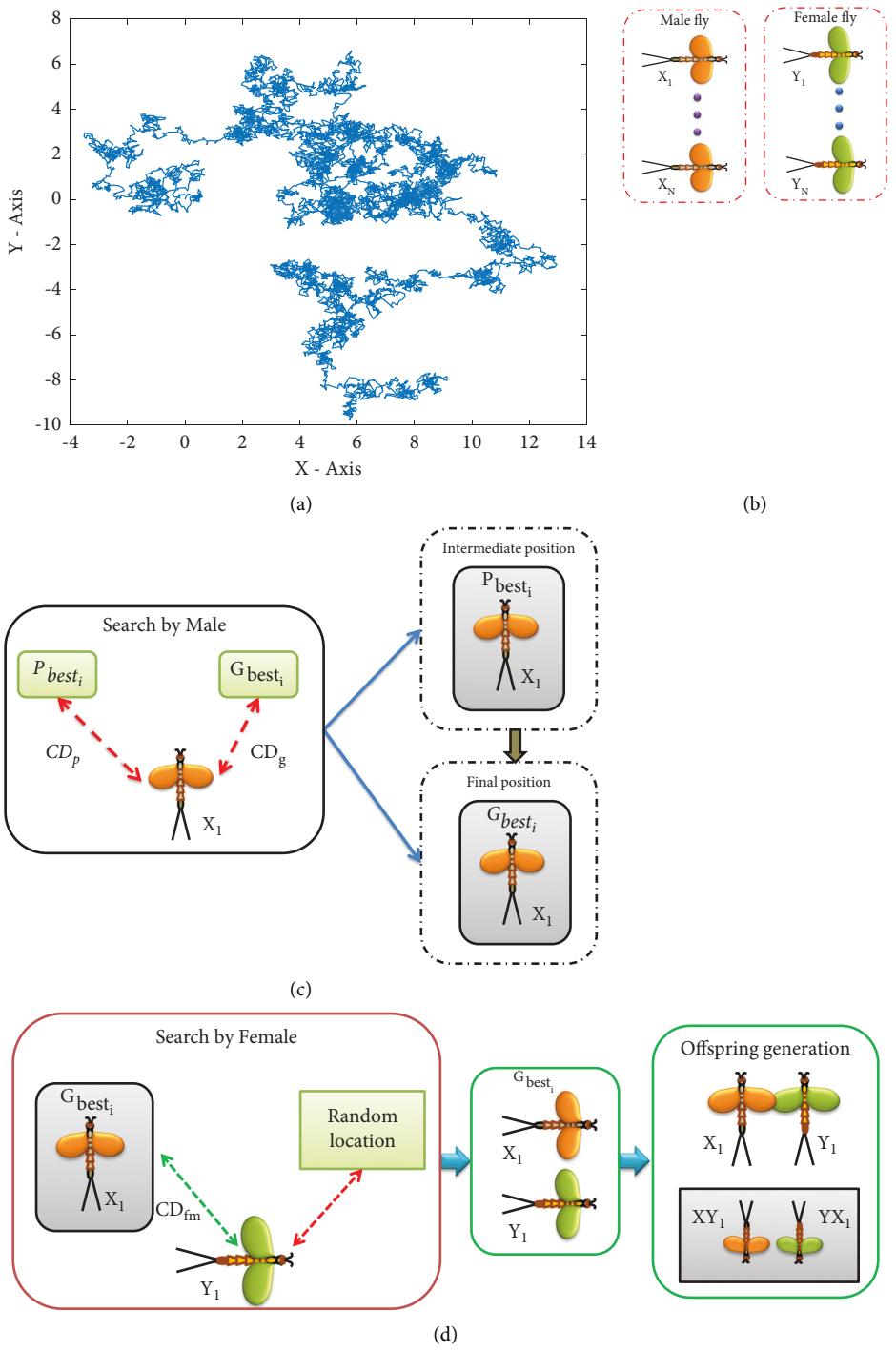


Figure 5: Demonstration of the Brownian search and Mayfly algorithm: (a) Brownian walk, (b) Stage1, (c) Stage2, and (d) Stage3.

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{20}$$

$$\mathrm{F1S} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}, \tag{21}$$

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}}. \tag{22}$$

4. Results and Discussion

This part of the work present the investigational results obtained with an Intel i5 2.6 GHz CPU, with 18 GB RAM and 4 GB VRAM, and equipped with Python®. In this work, 4800 images (2400 normal and 2400 COVID-19) are considered for evaluating the merit of the PDLM on the assigned task. In this work, the performance of the proposed scheme is verified with max-pooling (MP) and average pooling (AP)

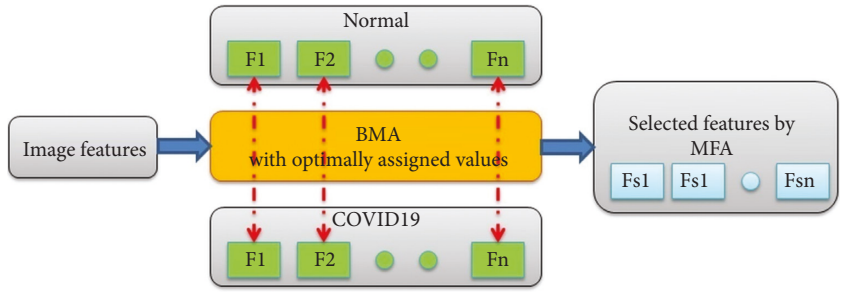


Figure 6: Feature optimization with BMA.

Table 3: Initial results achieved with pretrained deep learning methods.

Method	Pooling	TP	FN	TN	FP	AC	PR	SE	SP	F1S	NPV
AlexNet	MP	221	17	228	14	0.9354	0.9404	0.9286	0.9421	0.9345	0.9306
	AP	223	12	230	15	0.9437	0.9370	0.9489	0.9388	0.9429	0.9504
VGG16	MP	223	16	231	10	0.9458	0.9571	0.9331	0.9585	0.9449	0.9352
	AP	224	14	233	9	0.9521	0.9614	0.9412	0.9628	0.9512	0.9433
VGG19	MP	222	12	227	19	0.9354	0.9212	0.9487	0.9228	0.9347	0.9498
	AP	228	19	219	14	0.9313	0.9421	0.9231	0.9399	0.9325	0.9202
ResNet18	MP	230	16	218	16	0.9333	0.9350	0.9350	0.9316	0.9350	0.9316
	AP	231	13	222	14	0.9437	0.9429	0.9467	0.9407	0.9448	0.9447
ResNet50	MP	216	19	228	17	0.9250	0.9270	0.9191	0.9306	0.9231	0.9231
	AP	227	16	219	18	0.9292	0.9265	0.9342	0.9241	0.9303	0.9319
ResNet101	MP	229	13	220	18	0.9354	0.9271	0.9463	0.9244	0.9366	0.9442
	AP	219	15	230	16	0.9354	0.9319	0.9359	0.9350	0.9339	0.9388

Table 4: Experimental outcome of VGG16 with SoftMax classifier through 10-fold validation.

Cross validation	TP	FN	TN	FP	AC	PR	SE	SP	F1S	NPV
Fold1	217	23	223	17	0.9167	0.9274	0.9042	0.9292	0.9156	0.9065
Fold2	222	19	219	20	0.9187	0.9174	0.9212	0.9163	0.9193	0.9202
Fold3	228	18	219	15	0.9313	0.9383	0.9268	0.9359	0.9325	0.9241
Fold4	221	14	227	18	0.9333	0.9247	0.9404	0.9265	0.9325	0.9419
Fold5	231	11	221	17	0.9417	0.9315	0.9545	0.9286	0.9429	0.9526
Fold6	224	14	233	9	0.9521	0.9614	0.9412	0.9628	0.9512	0.9433
Fold7	226	15	221	18	0.9313	0.9262	0.9378	0.9247	0.9320	0.9364
Fold8	226	14	221	19	0.9313	0.9224	0.9417	0.9208	0.9320	0.9404
Fold9	228	19	216	17	0.9250	0.9306	0.9231	0.9270	0.9268	0.9191
Fold10	220	19	219	22	0.9146	0.9091	0.9205	0.9087	0.9148	0.9202

approaches, and the merit of the scheme is confirmed based on the achieved metrics.

Initially, the performance of the PDLM is tested on the considered images with the SoftMax classifier, and the achieved results are shown in Table 3. This table proves that the VGG16 scheme with AP helps to get superior categorization accuracy (95.21%) contrast to other methods. Table 4 confirms that fold6 presents a better result compared to other folds, and its graphical verification is presented in Figure 7.

The experimental outcome shown in Table 3 proves that the outcome achieved with VGG16 is superior contrast to other PDLM of this study. Also, this study verifies that the outcome of the average pooling is superior to max-pooling. Hence, the VGG16 with average pooling is then considered to verify the performance of the classifiers, such as DT, RF, NB, KNN, and SVM, and the results are presented in Table 5.

After verifying the performance of the VGG16 with deep-features, its performance is then confirmed using the BMA optimized serially integrated deep and HF. During this task, the BMF-based feature selection is then employed to find the optimal deep (equation (1)) and HF (equation (11)) features. The BMF algorithm based feature selection helps to get a deep-feature of size $1 \times 1 \times 417$, HF of size $1 \times 1 \times 241$, and the integrated feature of size $1 \times 1 \times 658$. This hybrid features are then used to verify the merit of VGG16 in detecting normal/COVID-19 X-ray images using the different classifiers using 10-fold validation, and the attained result is depicted in Table 5. This table validates that the KNN is better (accuracy = 99.17%) compared to other methods.

The various convolutional-layer (CL) outcome of the VGG16 achieved for a sample test image is presented in Figure 8. Figure 8(a) depicts the sample test image and

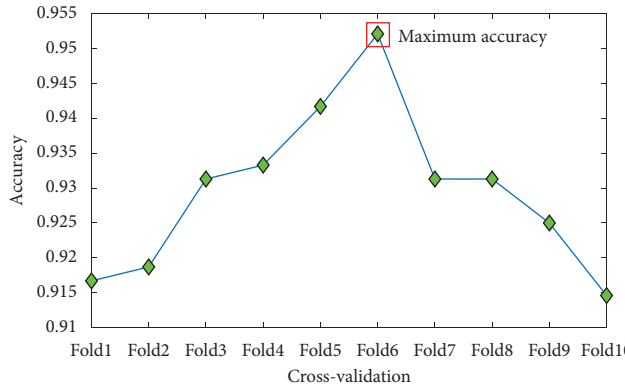


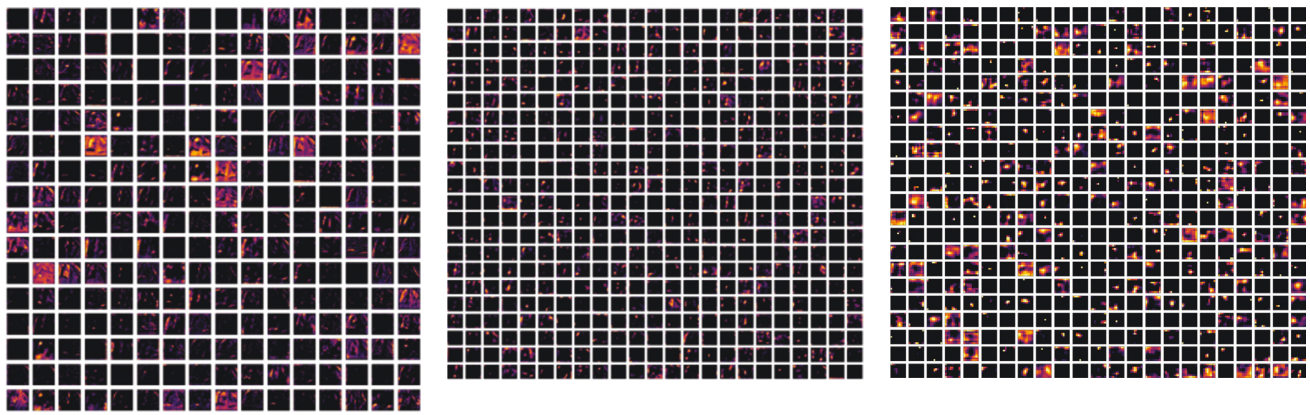
Figure 7: 10-fold cross validation outcome for VGG16 with SoftMax.

Table 5: Investigational outcome of VGG16 with deep and hybrid features.

Feature	Classifier	TP	FN	TN	FP	AC	PR	SE	SP	F1S	NPV
Deep features	SoftMax	224	14	233	9	0.9521	0.9614	0.9412	0.9628	0.9512	0.9433
	DT	227	8	231	14	0.9542	0.9419	0.9660	0.9429	0.9538	0.9665
	RF	228	15	226	11	0.9458	0.9540	0.9383	0.9536	0.9461	0.9378
	NB	230	11	224	15	0.9458	0.9388	0.9544	0.9372	0.9465	0.9532
	KNN	229	7	225	19	0.9458	0.9234	0.9703	0.9221	0.9463	0.9698
	SVM	226	10	232	12	0.9542	0.9496	0.9576	0.9508	0.9536	0.9587
Optimal Deep + HF	SoftMax	236	3	234	7	0.9792	0.9712	0.9874	0.9710	0.9793	0.9873
	DT	238	4	232	6	0.9792	0.9754	0.9835	0.9748	0.9794	0.9831
	RF	240	6	231	3	0.9812	0.9877	0.9756	0.9872	0.9816	0.9747
	NB	239	5	232	4	0.9812	0.9835	0.9795	0.9831	0.9815	0.9789
	KNN	244	0	232	4	0.9917	0.9839	1.0000	0.9831	0.9919	1.0000
	SVM	238	6	231	5	0.9771	0.9794	0.9754	0.9788	0.9774	0.9747

(a) (b) (c)

Figure 8: Continued.



(d) (e) (f)

Figure 8: Various convolutional-layer outcome for VGG16 scheme: (a) test image, (b) CL1, (c) CL2, (d) CL3, (e) CL4, and (f) CL5.

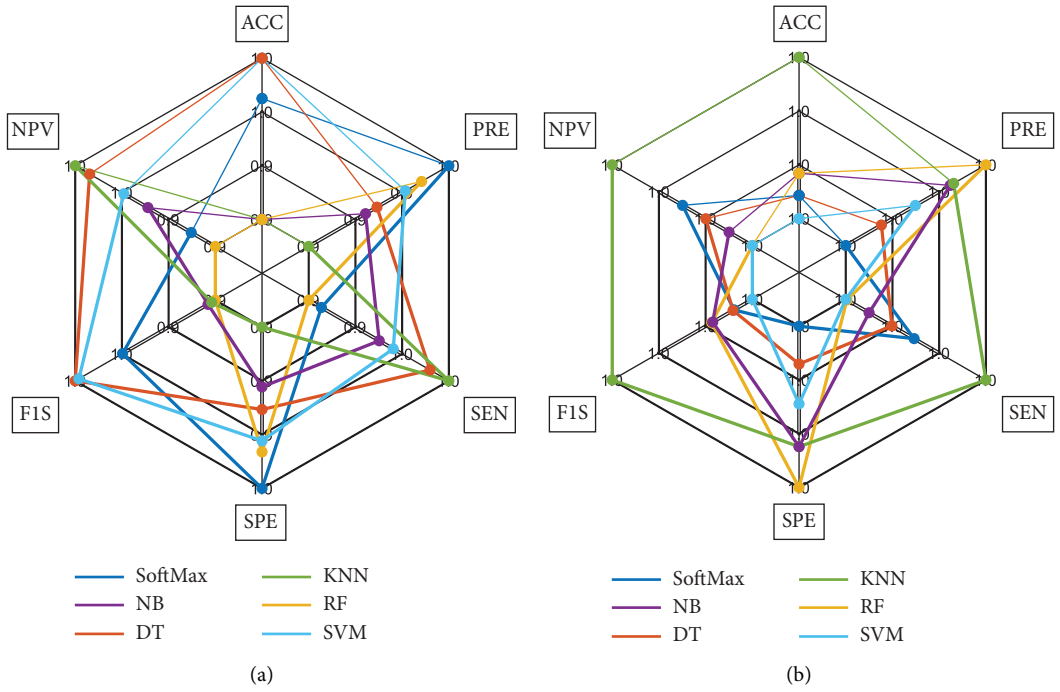


Figure 9: Spider-plot to display the overall merit of VGG16 with various features with chosen classifiers: (a) spider-plot1 and (b) spider-plot2.

Figures 8(b)–8(f) show the outcome of CL1 to 5, respectively. The overall performance of the binary classifiers is verified using the metrics in Table 5, analyzed with a spider plot, and is depicted in Figure 9. Figure 9(a) presents the plot to confirm the merit of VGG16 with the traditional deep-feature of dimension $1 \times 1 \times 1000$ and Figure 9(b) shows the result for deep + HF of size $1 \times 1 \times 658$. The spider plot, which creates a major pattern, is considered to be superior, and this plot confirms that the result of DT (with deep-feature) and KNN (with hybrid features) is better. The achieved experimental results with VGG16 and KNN for deep + HF are presented in Figure 10. Figures 10(a) and 10(b) demonstrate the validation/validation accuracy and loss function for 100 epochs. Figures 10(c) and 10(d) show the confusion matrix and ROC curves, respectively. From this result, it can be verified that the outcome of this experiment confirms that the proposed scheme helps to achieve a better classification metric during the assessment of the considered image database.

The performance of this practice is further demonstrated with the experimental outcome (classification accuracy) of other methods discussed in Table 1, and its value is graphically depicted in Figure 11. This comparison validates that the accurateness realized with the proposed scheme is improved compared to earlier works. This ensures that this proposal is clinically noteworthy, and the proposed technique can be considered to inspect the clinical-grade X-ray imagery, in future. In the future,

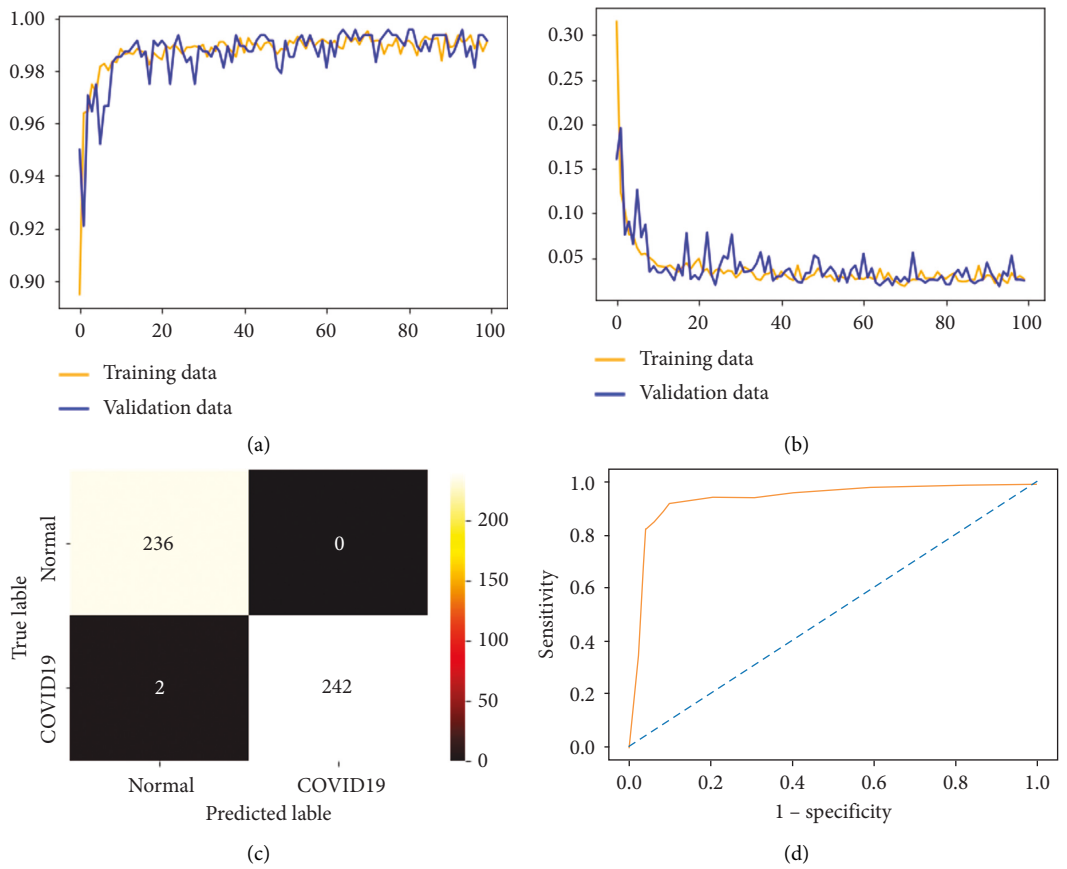


Figure 10: Experimental outcome for VGG16 with hybrid features and KNN classifier: (a) accuracy; (b) loss; (c) confusion matrix; (d) RoC curve.

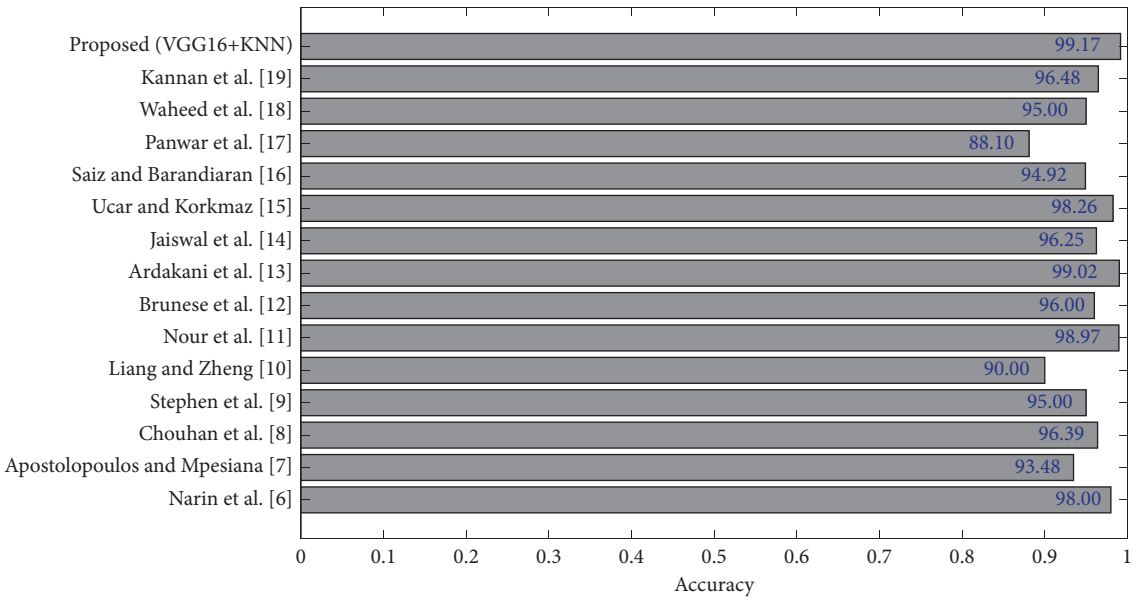


Figure 11: Comparison of classification accuracy of proposed scheme with existing methods.

the proposed methodology's performance can be enhanced by considering other handcrafted methods accessible in the literature.

5. Conclusion

COVID-19 is a pandemic disease that causes pneumonia in humankind, and the unrecognized infection will lead to death. X-ray-supported lung infection detection is an extensively implemented medical procedure, and radiologists and pulmonologists typically assess the recorded X-ray to recognize the disease. This research developed a PDLM-based COVID-19 recognition from X-ray, and this scheme executes different features assisted detection of COVID-19. This research considers the serially combined features of VGG16 and HF to classify the X-ray images into normal/COVID-19. Furthermore, this work employed the BMA to optimize the deep features and HF to reduce overfitting. The investigation is implemented using a binary classifier with 10-fold cross-validation. This study confirms that the BMA optimized Deep+HF helps get an improved accuracy (99.17%) with the KNN classifier. This accuracy is compared with other results existing in the literature, and this study confirms that the proposed scheme is better. This scheme can be considered to evaluate the clinically collected X-ray images in the future.

Data Availability

The Experimental data can be accessed from the following links: (1) https://www.kaggle.com/tawsifurrahman/COVID-19-radiography-database and (2) https://ieee-dataport.org/open-access/covid-19-and-normal-chest-x-ray.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. L. Fernandes, U. J. Tanik, and K. A. Karthik, "A reliable framework for accurate brain image examination and treatment planning based on early diagnosis support for clinicians," *Neural Computing & Applications*, vol. 32, no. 20, pp. 15897–15908, 2020.

[2] V. Kadry and S. Kadry, "Development of a framework for preserving the disease-evidence-information to support efficient disease diagnosis," *International Journal of Data Warehousing and Mining*, vol. 17, no. 2, pp. 63–84, 2021.

[3] M. A. Azam, K. B. Khan, S. Salahuddin et al., "A review on multimodal medical image fusion: compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics," *Computers in Biology and Medicine*, vol. 144, Article ID 105253, 2022.

[4] "Coronavirus disease (COVID-19) pandemic," 2019, https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[5] M. A. Khan, M. Alhaisoni, U. Tariq et al., "COVID-19 case recognition from chest CT images by deep learning, entropy-controlled firefly optimization, and parallel feature fusion," *Sensors*, vol. 21, no. 21, p. 7286, 2021.

[6] A. Narin, C. Pamuk, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *Pattern Analysis & Applications*, vol. 24, no. 3, pp. 1207–1220, 2021.

[7] I. D. Mpesiana and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 635–640, 2020.

[8] V. Chouhan, S. K. Khamparia, and A. Tiwari, "A novel transfer learning based approach for pneumonia detection in chest X-ray images," *Journal of Applied Sciences*, vol. 10, p. 559, 2020.

[9] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *J. Healthc. Eng*, vol. 2019, Article ID 4180949, 2019.

[10] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Computer Methods and Programs in Biomedicine*, vol. 187, Article ID 104964, 2020.

[11] M. Nour, Z. Cömert, and K. Polat, "A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization," *J. Appl. Soft. Comput*, vol. 97, Article ID 106580, 2020.

[12] L. Brunese, F. Mercaldo, A. Reginelli, and S. Antonella, "Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays," *Computer Methods and Programs in Biomedicine*, vol. 196, Article ID 105608, 2020.

[13] A. A. Ardakani, A. R. Kanafi, and U. R. Acharya, K. Nazanin and M. Afshin, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks," *Comput. Biol. Med*, vol. 121, Article ID 103795, 2020.

[14] A. Jaiswal, N. Gianchandani, D. Singh, K. Vijay, and K. Manjit, "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning," *Journal of Biomolecular Structure & Dynamics*, pp. 5682–5689, 2020.

[15] F. Ucar and D. Korkmaz, "COVIDiagnosis-Net: deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images," *J. Med. Hypotheses*, vol. 140, Article ID 109761, 2020.

[16] F. A. Saiz and I. Barandiaran, "COVID-19 detection in chest X-ray images using a deep learning approach," *I. J. Interactive Multi. Artifi. Inte*, vol. 6, no. 1, 2020.

[17] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. MoralesMenendez, and V. Singh, "Application of deep learning for fast detection of COVID-19 in X-rays using nCOVnet," *Chaos, Solitons & Fractals*, vol. 138, Article ID 109944, 2020.

[18] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "CovidGAN: data augmentation using auxiliary classifier GAN for improved COVID-19 detection," *IEEE Access*, vol. 8, pp. 91916–91923, 2020.

[19] S. R. Kannan, J. Sivakumar, and P. Ezhilarasi, "Automatic detection of COVID-19 in chest radiographs using serially concatenated deep and handcrafted features," *Journal of X-Ray Science and Technology*, vol. 30, no. 2, 2021.

[20] Kaggle, "COVID-19 Radiography Database," 2022, https://www.kaggle.com/tawsifurrahman/COVID-19-radiography-database.

[21] HappyMonday, "COVID-19 AND NORMAL CHEST X-RAY," 2021, https://ieee-dataport.org/open-access/covid-19-and-normal-chest-x-ray.

[22] T. Akram, M. Attique, S. Gul et al., "A novel framework for rapid diagnosis of COVID-19 on computed tomography scans," *Pattern Analysis & Applications*, vol. 24, no. 3, pp. 951–964, 2021.

[23] A. Gudigar, U. Raghavendra, T. Devasia et al., "Global weighted LBP based entropy features for the assessment of pulmonary hypertension," *Pattern Recognition Letters*, vol. 125, pp. 35–41, 2019.

[24] S. Kadry, V. Rajinikanth, R. González Crespo, and E. Verdú, "Automated detection of age-related macular degeneration using a pre-trained deep-learning scheme," *The Journal of Supercomputing*, pp. 1–20, 2021.

[25] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion Recognition Using PHOG and LPQ Features," in *Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp. 878–883, Santa Barbara, CA, USA, March 2011.

[26] K. Zervoudakis and S. safarakis, "A mayfly optimization algorithm," *Computers & Industrial Engineering*, vol. 145, Article ID 106559, 2020.

[27] S. Kadry, V. Rajinikanth, J. Koo, and B. G. Kang, "Image multi-level-thresholding with Mayfly optimization," *Iranian Journal of Electrical and Computer Engineering*, vol. 11, no. 6, pp. 2088–8708, 2021.

[28] N. S. M. Raja, K. S. Manic, and V. Rajinikanth, "Firefly algorithm with various randomization parameters: an analysis," in *Proceedings of the International Conference on Swarm Evolutionaryand Memetic Computing*, vol. 8297, pp. 110–121, Springer, Cham Germany, September 2013.

[29] Y. Xu, G. Yang, J. Luo, and J. He, "An Electronic Component Recognition Algorithm Based on Deep Learning with a Faster SqueezeNet," *Mathematical Problems in Engineering*, vol. 2020, Article ID 2940286, 2020.

[30] D. Sun, J. Wu, H. Huang, R. Wang, F. Liang, and H. Xinhua, "Prediction of short-time rainfall based on deep learning," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6664413, 2021.

[31] M. D. Hasan, S. Ahmed, Z. M. Abdullah et al., "Deep learning approaches for detecting pneumonia in COVID-19 patients by analyzing chest X-ray images," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9929274, 2021.

[32] A. M. Karim, M. S. Güzel, M. R. Tolun, H. Kaya, and F. V. Çelebi, "A new generalized deep learning framework combining sparse autoencoder and Taguchi method for novel data classification and processing," *Mathematical Problems in Engineering*, vol. 2018, Article ID 3145947, 2018.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 2876481, 10 pages
https://doi.org/10.1155/2022/2876481

Research Article

Personalized Recommendation Algorithm for Interactive Medical Image Using Deep Learning

Feng Liu[1,2] **and Weiwei Guo**[1]

[1]*School of Electrical and Information Engineering, Heilongjiang University of Technology, Jixi 158100, China*
[2]*Faculty of Communication, Visual Art and Computing, Universiti Selangor, Shah Alam 40000, Malaysia*

Correspondence should be addressed to Weiwei Guo; gwwguoweiwei@163.com

Received 21 April 2022; Revised 24 May 2022; Accepted 14 June 2022; Published 27 June 2022

Academic Editor: Xiaofeng Li

Personalized interactive image recommendation has several issues, such as being slow or having poor recommendation quality. Therefore, we propose an image personalized recommendation algorithm (IPRA) using deep learning to improve the time and quality of personalized interactive image recommendations. First, the feature subimage is obtained and converted into a one-dimensional vector using the convolution neural network model. Single input and single output functional and dual input and single output generalized functional network model are integrated into the model to improve the learning ability of nonlinear mapping and avoid overfitting during the training process; second, a one-dimensional vector is clustered using the fuzzy k-means approach and then translated into hyperbolic space; Finally, the Poincare map model is used to map the updated vector, the transformed vector is mapped using the PM model, and the image information is fed back to the two-dimensional plane, and the image recommendation set is formed based on the ranking of similarity, and the visual recommendation is presented to the user. The results show that the size of the convolution kernel is 2×2, and the image one-dimensional vector clustering can be better completed. The optimal value of $F1$ is 0.92, and the optimal value of average time is 11 s. The image recommendation quality is better, and the image recommendation can be formed according to the photographic similarity, which has good application value.

1. Introduction

At the moment, with the rapid development of Internet technology and multimedia [1], image interactive personalized recommendation is an important research in the field of image application to obtain the required images accurately and quickly and transmit them to users in time [2]. The personalized recommendation is a clear feature of content distribution, which has a high performance, high availability level, which can recommend it related content to the user in a short time, recommended content reliability, and high accuracy. Personalization recommendation avoids the filling information push of traditional recommendation services, which can set the push time and the number of pushes and can push the relevant information to the user according to the user's interest and preferences [3, 4]. The main contributions of this article are as follows: (1) image features are obtained through the training of the image recommendation process model to visualize the image recommendation results. (2) To realize interactive image personalized recommendation, hyperbolic space and the Poincare map (PM) are used to complete the visualization of images.

For the personalized recommendation problem of the image, many scholars put forward related learning methods. Zhou W et al. [5] utilized image structures for the recommendation, and image-based recommendation methods focus on capturing the user's preferences and using the image model to exploit the relationships between different entities in the image, and a new-based ranking recommended algorithm is proposed, which utilizes the user's explicit and implicit feedback. Yin P et al. [6] proposed a recommended algorithm based on deep learning, the algorithm utilizes a user preference three-dimensional model and improved resource

allocation processes, matching target users with similar preferences, and performing personalized recommendations. The principle of the additional preference layer is to capture the user's pair preference, providing detailed information for the user for further recommendation. The results show that this algorithm has better performance than other image based and ranking oriented benchmark algorithms. Chen S et al. [7] calculated the semantic similarity between learners and all learning resources based on the similarity measurement based on meta path, combined with knowledge transformation probability and learning feedback information. According to the similarity ranking, the learning resources ranked Top-k are recommended to learners relevant experiments show that it effectively realizes the accurate recommendation of learning resources in adaptive learning. Ye Junmin et al. [8] proposed a research on Learning Resource Recommendation Algorithm Based on HIN. On the basis of similarity measurement based on meta path, combined with knowledge transformation probability and learning feedback information, semantic similarity is calculated, learners are sorted according to the similarity, and learning resources ranked Top-k are recommended to teachers and students. Yang X et al. [9] proposed a novel Translation-based Neural Fashion Compatibility Modeling (TransNFCM) framework, which jointly optimizes fashion item embeddings and category-specific complementary relations in a unified space via an end-to-end learning manner. Extensive experiments demonstrate the effectiveness of TransNFCM over the state-of-the-art on two real-world datasets. Jian M et al. [10] proposed a semantic manifold modularization-based ranking (MMR) for image recommendation. Experimental results demonstrate that user-consumed visual correlations play actively to capture users' interests, and the proposed MMR can infer user-image correlations via visual manifold propagation for image recommendation. Qiu Ningjia et al. [11] proposed the research on recommendation algorithm based on user preference optimization model and constructed user preference matrix for project type by using user project scoring matrix and project type information; Then, the linear regression model is used to calculate the user's weight for each type; finally, the prediction score is combined with slope one algorithm to improve the quality of user preferred recommendation algorithm. Although the above-given methods have made some progress, in the process of recommendation, the recommendation of the image set cannot be completed according to the similarity of the images. Therefore, this paper proposes an image personalized recommendation algorithm (IPRA) under deep learning. The IPRA cannot only complete image recommendations, but also realize image visualization and image interaction.

2. Methodology

2.1. Method Framework. In this paper, in order to realize the personalized recommendation of the interactive image, the IPRA based on a convolution neural network (CNN) is proposed. The method includes two parts: the model training process and the image recommendation process. The method structure is shown in Figure 1.

As can be seen from Figure 1, the method framework is mainly completed by three steps, and steps 1 and 2 belong to the model training process, and Step 3 is the recommended process and visualization. In step 1, it is necessary to determine the input of the model. In this paper, we used the features of images as input to build an association model between users and images, determine whether the images meet the needs of users, and judge whether the images are recommended. Step 2 is to construct a CNN model according to the association between the user and the image, and complete the model training. The model expression is

$$F(X_i) = \sum_{i=1}^{k} \varphi_i \times \beta_{i.t}, \tag{1}$$

where φ_i represents the weight of Gaussian distribution i. $\beta_{i.t}$ represents the mean value of Gaussian distribution i when the time is t. Equation (1) shows that the probability density function of the convolution neural network model is established by using the three-dimensional Gaussian function with the quantity of k.

Step 3 is to input the training data into the training completed model, complete the personalized recommendation, and complete the visualization of the recommendation results. The training process of the model is to realize the algorithm design as well as to obtain the image features; the image personalized recommendation process is to realize the visualization of image recommendation results based on the training process [12, 13].

2.2. Interactive Image Feature Extraction from the Perspective of Deep Learning. The structure of hierarchical generalized networks in interactive image features from a deep learning perspective is basically the same as that of artificial neural networks, and not all hierarchical generalized networks can be described by a universal structure, nor can all hierarchical generalized networks be represented using a unified generalized equation [14, 15]. Based on these features, interactive image features are classified into single input and single output type and dual input and single output type, and the two models are used as basic components in extracting interactive image features of hierarchical generalized networks [16].

2.2.1. Construction of Single Input and Single Output Generalized Function Network Model. Figure 2 shows the specific structure of the single input and single output type generalized network model.

Figure 2 shows that the structure of the single input and single output functional network model shows that it can continuously map between input and output, thus overcoming the problems of network continuity and approximation and can be well applied to various complex problems.

The output expression for a single input and single output generalized function network is

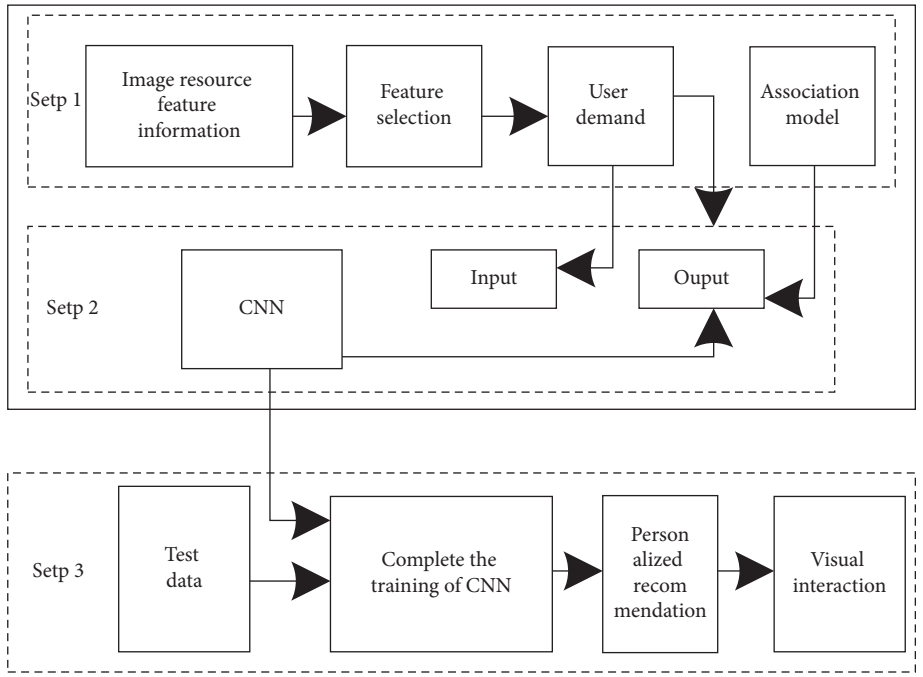


Figure 1: Method framework.

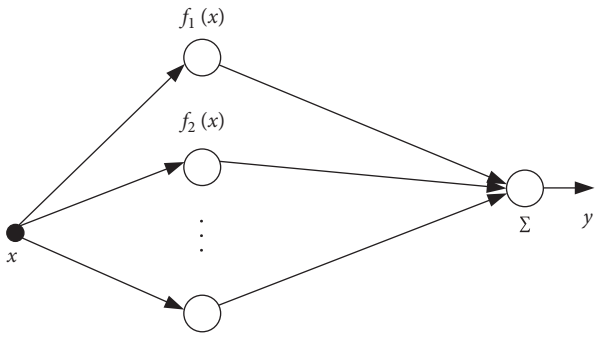


Figure 2: Structure of single input and single output generalized network model.

$$y = \sum_{i=1}^{n} f_i(x). \tag{2}$$

In hierarchical generalized networks, the expression uniqueness problem is an important issue to be solved.

2.2.2. Two-Input Single-Output Generalized Network Model Construction. The structure of the two-input single-output flooding network model is shown in Figure 3.

It can be seen from Figure 3 that, it can perform a continuous mapping between input and output. Its form is different from the single input and single output functional network model, and the selection of functional neuron function in the functional network is not fixed, which makes the model more adaptable.

Assuming that $\{x, y\}$ and $\{z\}$ represent the input and output vectors of a two-input single-output generalized network, respectively [17, 18], then the output of the generalized network can be expressed as follows:

$$z = G(x, y) = \sum_{j=1}^{m} g_j(x, y). \tag{3}$$

When Equation (3) holds the generalized network can be defined as a separable generalized network.

$$g_j(x, y) = p_j(x) q_j(y). \tag{4}$$

Based on the above-given single input and single output generalized network model and dual input and single output generalized network model, the specific steps for extracting interactive image features from a deep learning perspective are [19, 20].

Step 1. In the first layer of the hierarchical generalized function network, assume that there exist n_1 input variables $x_1, x_2, \ldots, x_n$. When $i = 1$, the output of the first layer of the hierarchical generalized network is

$$y_1 = f\left(\overrightarrow{x_1}\right), \tag{5}$$

where the expression for $\overrightarrow{x_1}$ is

$$\overrightarrow{x_1} = \left(x_1, x_2, \ldots, x_{n_1}\right). \tag{6}$$

Step 2. Let $i = i + 1$, then there are $n_i + 1$ input variables in the i th basic level generalized network, then the output of the basic level generalized network is

$$f_i\left(x_{N_i+1}, x_{N_i+2}, \ldots, x_{N_i+n}, y_{i-1}\right), \tag{7}$$

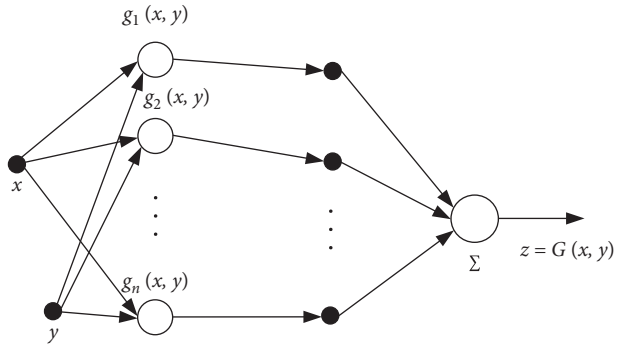


Figure 3: Dual input and single output generalized network model structure.

where the expression for y_{i-1} is

$$y_{i-1} = f_{i-1}(\overrightarrow{x_{i-1}}),$$
$$N_i = \sum_{j=1}^{i-1} n_j. \tag{8}$$

Step 3. If the relation $\sum_{j=1}^{n} n_j < n$ exists, return to Step 2, and vice versa build the hierarchical generalized function network [21, 22].

According to the above-given steps, interactive image feature extraction from a deep learning perspective is completed.

2.3. Proposed Algorithm. To achieve image personalized recommendation, clustering is used to cluster the one-dimensional vectors of mapped images, and it is done based on the degree of similarity of visual content, which is done in this paper using the fuzzy *K*-means method. The interactive image personalization recommendation algorithm is described as follows:

Input: introduction of the affiliation function $P_i(x_j)$. The function Equation of the method is

$$J_f = \sum_{i=1}^{k} \sum_{j=1}^{k} \left[P_i\left(x_j\right)\right]^b \left\|x_j - m_i\right\|^2, \tag{9}$$

where j represents the j-th sample; b is the constant; m_i represents the cluster center.

Output: the result of interactive image personalized recommendation.

In order to realize interactive image personalized recommendation, hyperbolic space and PM are used to complete the visualization of images. The specific steps are as follows:

(1) The one-dimensional feature vector after clustering is transformed into hyperbolic space; The expression is

$$P = \frac{H_2 - H_1}{J_f}, \tag{10}$$

where H_1 is the hyperbolic space coefficient. H_2 represents hyperbolic space conversion coefficient.

(2) PM model is used to map the transformed feature vector, and the image information is fed back to the two-dimensional plane to realize image interaction. The expression of the PM model is:

$$X_{abc} = Q_a \times W_b \times R_c, \tag{11}$$

where Q_a is the offset function, W_b is the excitation function. R_c indicates the output characteristic.

(3) Hyperbolic space can maximize the retention of image feature similarity information and display the nonlinear growth trend of an image, which belongs to feature similarity. The expression is

$$D_e = \frac{P_{\max}}{X_{abc}}, \tag{12}$$

where $P_{\max}$ represents the maximum load of image nonlinear growth trend.

(4) Since the space cannot present the image on the two-dimensional plane, Poincare disk mapping is used to map the coordinate points of the space and present them on the two-dimensional plane to realize the presentation of the image on the two-dimensional plane.

(5) The PM model is also known as a codisc model, which consists of a hyperplane geometric model, and the dimension is *n*. During mapping, if $[t, x_1, \ldots, x_p]$ is a point on the hyperboloid and located in this space, the definition of a point in the hyperboloid model can be completed, the point can be connected with $[-1, 0, \ldots, 0]$, and the connecting line between them can be mapped to hypersurface $t = 0$, so as to obtain the corresponding point in the PM model.

2.4. Experimental Analysis and Results. In order to verify the effectiveness and validity of the IPRA from the deep learning perspective, the experimental environment is built in MATLAB simulation software, and the operating system required for the experiment is Windows 10. The algorithm of Zhou W et al. [5], the algorithm of Yin P et al. [6], the algorithm of Chen S et al. [7], and the algorithm of Ye Junmin et al. [8] are used as experiments to compare the IPRA. The LIDC-IDRI data set and the LUNA16 data set are used as experimental objects and the images in two images are named L image and C image, respectively, and real-time tracking of images in two datasets using the IPRA to verify the effectiveness of the recommendations of the IPRA. The LIDC-IDRI data set was collected at the initiative of the National Cancer Institute (NCI) to study the early detection of pulmonary nodules in high-risk populations. In this data set, a total of 1018 study instances are included. For each instance, the images were diagnostically annotated in two stages by four experienced chest radiologists. The data set consists of chest medical image files (e.g., CT and

radiographs) and corresponding diagnostic outcome lesion annotations. The LUNA16 data set contains 888 CT images with 1084 tumors, with a desirable range of image quality and tumor size. The data were divided into 10 subsets, and the subset contained 89/88 CT scans. The CT images of LUNA16 were taken from the LIDC/IDRI data set, and the annotation with more than three radiologists' agreement was selected, and the tumors smaller than 3 mm were removed, so the data set does not contain tumors smaller than 3 mm, which is convenient for training. 500 images are randomly selected from each image data set for experimental analysis. 70% of the images will be selected as the training set, and the other images will be used as the test set.

The evaluation indexes are as follows:

(1) *Convolution Kernel Size.* In the training process of IPRA, the size of the convolution kernel has a direct impact on the training results. Therefore, it is necessary to determine the optimal size of the convolution kernel and take the value of the loss function as the measurement standard.

(2) *Image Feature Clustering Effect.* Taking adjusted Rand index (ARI), Macro-$F1$ measure ($F1$), and average time as evaluation criteria, the calculation equations of the first two are expressed by Equations (13) and (14), and the larger the values of the two, the better the performance of the IPRA.

$$K_{\mathrm{API}} = \frac{\sum_{ij}\left(n_{ij}/2\right) - \left[\sum_i (a_i/2)\sum_j \left(b_j/2\right)\right]/(n/2)}{1/2\left[\sum_i (a_i/2) + \sum_j \left(b_j/2\right)\right] - \left[\sum_i (a_i/2)\sum_j \left(b_j/2\right)\right]/(n/2)}. \tag{13}$$

$$K_{\mathrm{Macro}-F1} = \frac{2K_{\mathrm{Macro}-P} * K_{\mathrm{Macro}-P}}{K_{\mathrm{Macro}-P} + K_{\mathrm{Macro}-P}}, \tag{14}$$

where a_i represents the average distance, $K_{\mathrm{Macro}-P}$ and $K_{\mathrm{Macro}-R}$ represent macroprecision and macrorecall, respectively.

(3) *Personalized Recommendation Effect.* the Normalize Discounted Cumulative Gain (nDCG) is adopted as the evaluation index, which is standardized, and its calculation equation is

$$n\mathrm{DCG} = \sum_{i=1}^{k} \frac{2^{rel(i)} - 1}{\log_2 (i+1)}, \tag{15}$$

where k represents the k-th image. $2^{rel(i)} - 1$ and $\log_2 (i+1)$ represent the quality and weight of each image recommendation result, respectively. The larger the NDCG value, the better the quality of the recommended image.

(4) In order to intuitively measure the image recommendation effect of the five algorithms, the images in the verification set of the five algorithms are used for recommendation, and one image is randomly extracted from the Y image as the target image to obtain the image recommendation results of the five algorithms.

(5) *Expect Average Overlaprate (EAO).* The algorithm from Zhou W et al. [5], Yin P et al. [6], Chen S et al. [7], and Ye Junmin et al. [8] is selected as the comparison algorithm of the IPRA, and the images within the recommended dataset using the five algorithms are analyzed for the EAO in the process of recommending 10 images with different attributes. EAO belongs to the comprehensive evaluation index of tracking accuracy and robustness, and the value of EAO is proportional to the tracking effect.

3. Results and Discussion

The function value results are tested under different convolution kernel sizes are shown in Figure 4.

According to Figure 4, with the increase of the number of iterations, the value of the loss function of the convolution kernel changes significantly when the size of the convolution kernel is 2×2, the loss function value shows a slow downward trend with small fluctuation. The other two convolution kernels are different in size and the loss function value is higher. Therefore, the convolution kernel size is determined to be 2×2.

In order to analyze the advantages of IPRA, Zhou W et al. [5], Yin P et al. [6], Chen S et al. [7], and Ye Junmin et al. [8] algorithms are used as the comparison algorithms of IPRA. The test results of the five algorithms on the training set are obtained according to Equations (13) and (14), as shown in Table 1.

According to the test results of Table 1, among the test results of the five algorithms with the different number of images and three evaluation indexes, the results of each index of IPRA are better than the other four algorithms, and the best value of ARI is 0.62, the best value of $F1$ is 0.92, and the best value of average time is 11 s, which are significantly better than the four comparison algorithms. Therefore, the clustering effect of IPRA is good.

According to equation (6), the recommended nDCG results for the verification set are obtained, as shown in Figure 5.

According to Figure 5, in the test set, with the gradual increase of the number of images, the nDCG values of the five algorithms fluctuate to a certain extent, and rise slowly and slightly; in which, the nDCG values of the IPRA are above 0.8, and the nDCG values of the other four

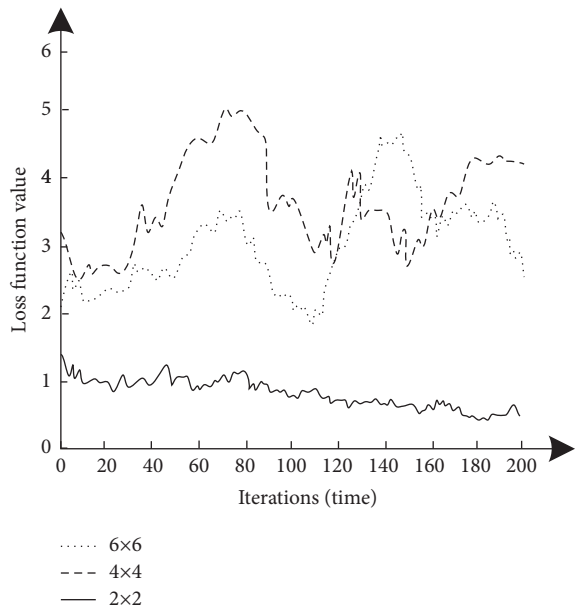


Figure 4: Convolution nuclear size test results.

Table 1: Comparison of five algorithms.

Number of images	IPRA			Zhou W et al. [5]			Yin P et al. [6]		
	ARI	*F*1	Average time (s)	Ari	*F*1	Average time (s)	ARI	*F*1	Average time (s)
25	0.58	0.92	12	0.13	0.54	34	0.18	0.59	29
50	0.61	0.88	14	0.14	0.58	31	0.17	0.58	32
75	0.59	0.91	Mean	0.12	0.56	30	0.19	0.59	31
100	0.62	0.83	16	0.13	0.57	32	0.18	0.62	28
125	0.58	0.9	14	0.12	0.63	29	0.19	0.61	30
150	0.61	0.83	Mean	0.13	0.57	33	0.18	0.59	33
175	0.62	0.84	11	0.16	0.53	35	0.19	0.63	29
200	0.59	0.91	16	0.14	0.64	34	0.18	0.6	31
225	0.61	0.89	15	0.13	0.55	32	0.19	0.59	30

Number of images	Chen S et al. [7]			Ye Junmin et al. [8]		
	ARI	*F*1	Average time (s)	ARI	*F*1	Average time (s)
25	0.12	0.53	33	0.17	0.58	28
50	0.13	0.57	30	0.16	0.57	31
75	0.11	0.55	29	0.18	0.58	30
100	0.12	0.56	31	0.17	0.61	27
125	0.11	0.62	28	0.18	0.60	29
150	0.12	0.56	32	0.17	0.58	32
175	0.15	0.52	34	0.18	0.62	28
200	0.13	0.63	33	0.17	0.59	30
225	0.12	0.54	31	0.18	0.58	29

comparison algorithms are between 0.5 and 0.8, which is significantly lower than the IPRA. Therefore, the image recommendation quality of the IPRA is better.

The images in the validation set of the five algorithms are used for recommendation, and one image is randomly extracted from the Y image as the target image, and the image recommendation results of the five algorithms are obtained as shown in Figures 6 and 7.

According to the test results of Figure 7, the five algorithms can complete the personalized recommendation of the target image. However, the IPRA can form an image recommendation set according to the photographic similarity. Users can click any recommendation set to view all the images in it; the other four algorithms cannot form the image recommendation set, cannot complete the formation of the image recommendation set according to the photographic similarity and can only present the recommendation results of relevant images. Therefore, the recommendation effect of IPRA is better than the four comparison algorithms.

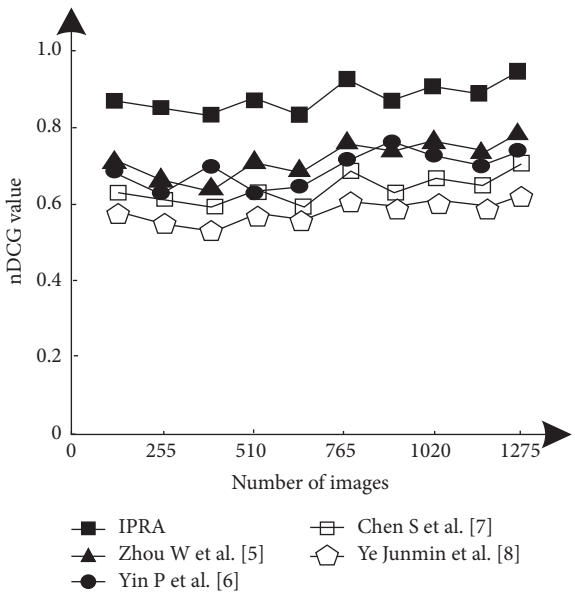


Figure 5: nDCG test results of the five algorithms.

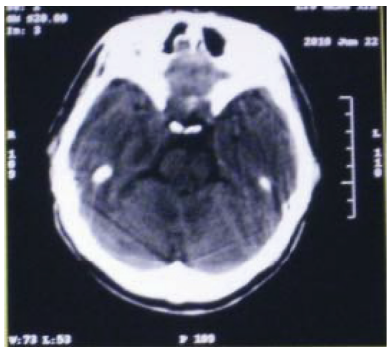


Figure 6: Target image.

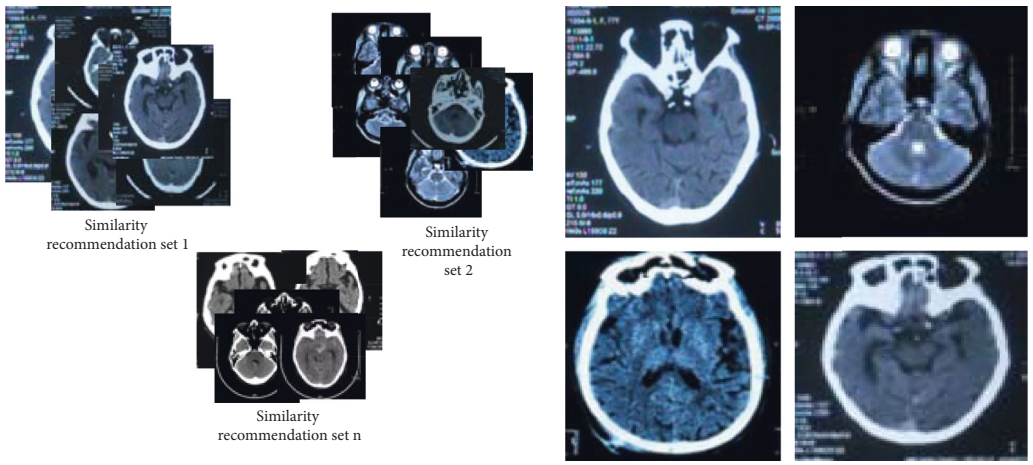


Figure 7: Continued.

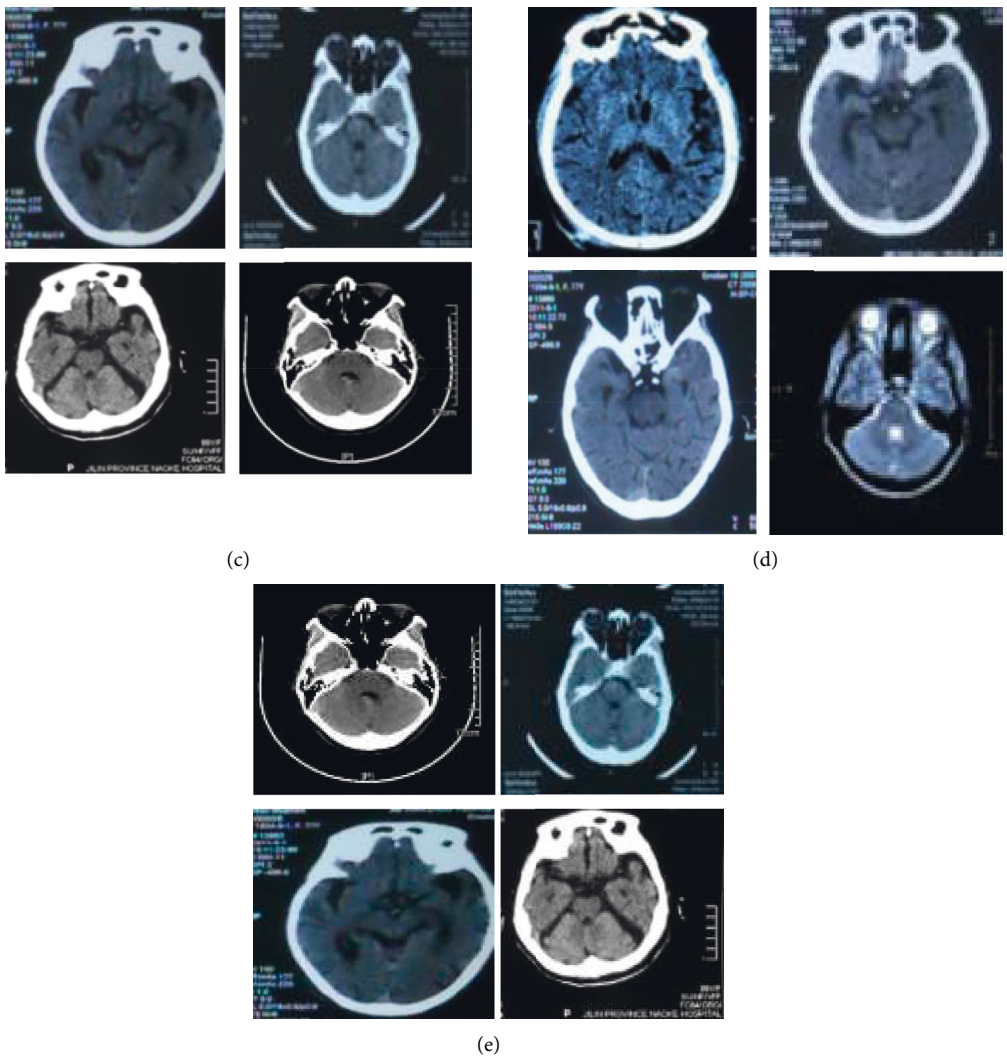


Figure 7: Recommended results of the five algorithms. (a) Recommendation results of the IPRA. (b) Recommended results of Zhou W et al. [5]. (c) Recommended results of Yin P et al. [6]. (d) Recommended results of Chen S et al. [7]. (e) Recommended results of Ye Junmin et al. [8].

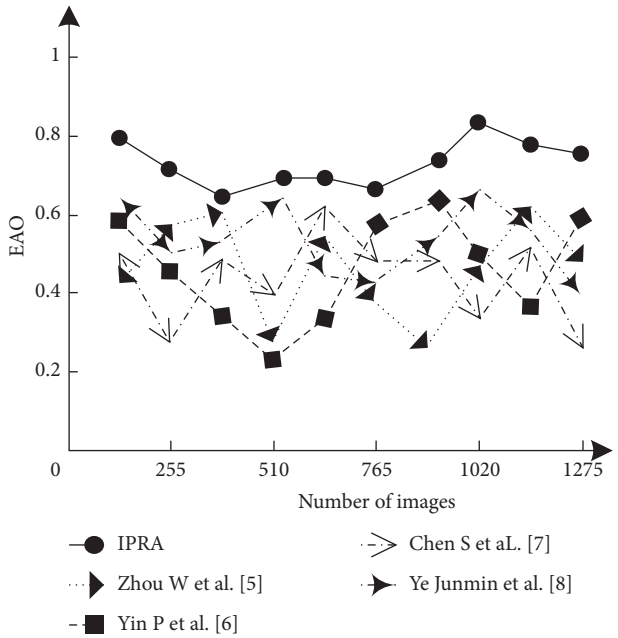


Figure 8: EAO analysis results for five algorithms.

The algorithm from Zhou W et al. [5], Yin P et al. [6], Chen S et al. [7], and Ye Junmin et al. [8] was chosen as the comparison algorithm for the IPRA. Five algorithms are used to recommend interactive images in the data set. The EAO analysis results are shown in Figure 8.

According to Figure 8, when the five algorithms recommend interactive images, the EAO values of this algorithm are significantly higher than those of the other four algorithms, and the change range of EAO values of IPRA in image personalized recommendation is small, and the fluctuation range of EAO values of the other five algorithms is large. Experiments show when recommending interactive images, the EAO values of IPRA are high, which has a better personalized recommendation effect.

4. Conclusions

In order to realize interactive image personalized recommendation and provide users with more reliable and higher quality images, this paper proposes an IPRA from the perspective of deep learning. The conclusions are as follows: (1) the IPRA combines the CNN model, hyperbolic spatialization, PM, and fuzzy k-means algorithm to realize image feature extraction, processing, transformation, and clustering. (2) The IPRA realizes image recommendation and visualization. (3) With the gradual increase of the number of images, the nDCG values of this algorithm fluctuate to a certain extent, and rise slowly and slightly. The nDCG values are above 0.8, and the image recommendation quality is better. (4) The IPRA has good image feature clustering performance, and the recommended image quality is high. At the same time, it can form the recommended set of images required by users according to the similarity, and present the image recommendation results. (5) The EAO value of IPRA is high and has a better personalized recommendation effect.

From the perspective of deep learning, IPRA has shortcomings. In future works, it is necessary to continuously optimize the algorithm according to the development of the image, so as to truly and accurately provide the basis for the image algorithm. The security and omnipotence of the personalized recommendation process are studied to further optimize the performance of interactive image personalized recommendation. (1) The recommended algorithm is fully considering different features, and the new metric is performed according to the actual development. (2) Improving the performance of recommendation capability prediction in the recommendation process.

Data Availability

The data used to support the findings of this study are included within the article. Readers can access the data supporting the conclusions of the study from LIDC-IDRI data set and the LUNA16 data set.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of Heilongjiang Province of China under Grant no. LH2021F049.

References

[1] J. Chen, Y. Wu, and L. Xiang, "Double layered recommendation algorithm based on fast density clustering with graph-based filtering & Applications," *Control theory and application*, vol. 36, no. 04, pp. 542–552, 2019.

[2] X. Zhang, Y. Zhao, L. Zhang, and X. Song, "CEST magnetic resonance image analysis and display system with intelligent interactive region of interest selection," *Journal of Northwest University*, vol. 50, no. 04, pp. 582–588, 2020.

[3] G. Luo, Z. Liu, L. Zhang, J. Zhang, Z. He, and X. Zhang, "Research on personalized video recommendation algorithm based on context awareness in mobile environment," *Application Research of Computers*, vol. 37, no. 05, pp. 1306–1310, 2020.

[4] P. B. Mallikarjuna, M. Sreenatha, S. Manjunath, and N. C. Kundur, "Aircraft gearbox fault diagnosis system: an approach based on deep learning techniques," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 258–272, 2020.

[5] W. Zhou and W. Han, "Personalized recommendation via user preference matching," *Information Processing & Management*, vol. 56, no. 3, pp. 955–968, 2019.

[6] P. Yin and L. Zhang, "Image recommendation algorithm based on deep learning," *IEEE Access*, vol. 8, no. 7, pp. 132799–132807, 2020.

[7] S. Chen, L. Huang, Z. Lei, and S. Wang, "Research on personalized recommendation hybrid algorithm for interactive experience equipment," *Computational Intelligence*, vol. 36, no. 3, pp. 1348–1373, 2020.

[8] J. Ye, P. Huang, D. Luo, Z. H. Wang, and S. Chen, "Research on a learning resource recommendation algorithm based on HIN," *Microcomputer System*, vol. 040, no. 004, pp. 726–732, 2019.

[9] X. Yang, Y. Ma, L. Liao, M. Wang, and T. S. Chua, "TransNFCM: translation-based neural fashion compatibility modeling," *AAAI*, vol. 33, pp. 403–410, 2019.

[10] M. Jian, J. Guo, C. Zhang et al., "Semantic manifold modularization-based ranking for image recommendation," *Pattern Recognition*, vol. 120, no. 3, Article ID 108100, 2021.

[11] N. Qiu, Z. He, P. Wang, and Y. Li, "Research on recommendation algorithm based on user preference optimization model," *Application Research of Computers*, vol. 036, no. 012, pp. 3579–3582, 2019.

[12] F. Feng, S. Wang, C. Wang, and J. Zhang, "Learning deep hierarchical spatial-spectral features for hyperspectral image classification based on residual 3D-2D CNN," *Sensors*, vol. 19, no. 23, p. 5276, 2019.

[13] F. Li, M. Liu, Y. Zhao et al., "Feature extraction and classification of heart sound using 1D convolutional neural networks," *EURASIP Journal on Applied Signal Processing*, vol. 59, no. 1, pp. 1–11, 2019.

[14] W. Liu, C. Qin, K. Gao et al., "Research on medical data feature extraction and intelligent recognition technology based on convolutional neural network," *IEEE Access*, vol. 7, no. 9, pp. 150157–150167, 2019.

[15] W. Wang, F. Bu, Z. Lin, and S. Zhai, "Learning methods of convolutional neural network combined with image feature

extraction in brain tumor detection," *IEEE Access*, vol. 8, no. 8, pp. 152659–152668, 2020.

[16] W. Li, B. Li, C. Yuan et al., "Anisotropic convolution for image classification," *IEEE Transactions on Image Processing*, vol. 29, no. 4, pp. 5584–5595, 2020.

[17] W. H. Tu, "Resting-state functional network models for posttraumatic stress disorder," *Journal of Neurophysiology*, vol. 125, no. 3, pp. 1–12, 2021.

[18] W. Wang, Y. Fu, F. Dong, and F. Li, "Semantic segmentation of remote sensing ship image via a convolutional neural networks model," *IET Image Processing*, vol. 13, no. 6, pp. 1016–1022, 2019.

[19] Y. Guan, Q. Wei, and G. Chen, "Deep learning based personalized recommendation with multi-view information integration," *Decision Support Systems*, vol. 118, no. 5, pp. 58–69, 2019.

[20] Y. Zhang, T. S. Lee, M. Li, F. Liu, and S. Tang, "Convolutional neural network models of V1 responses to complex patterns," *Journal of Computational Neuroscience*, vol. 46, no. 1, pp. 33–54, 2019.

[21] H. Ren and T. Hu, "An adaptive feature selection algorithm for fuzzy clustering image segmentation based on embedded neighbourhood information constraints," *Sensors*, vol. 20, no. 13, p. 3722, 2020.

[22] Q. Zhang, X. Wang, S. Wang, Z. Sun, S. Kwong, and J. Jiang, "Learning to explore saliency for stereoscopic videos via component-based interaction," *IEEE Transactions on Image Processing*, vol. 29, no. 99, pp. 5722–5736, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 8033806, 16 pages
https://doi.org/10.1155/2022/8033806

Research Article

A Copula Type-Model for Examining the Role of Microbiome as a Potential Tool in Diagnosis

Enrique Calderín–Ojeda,[1] **Guillermo López–Campos**,[2] **and Emilio Gómez–Déniz**[3]

[1]*Centre for Actuarial Studies, Department of Economics, University of Melbourne, Melbourne, Australia*
[2]*Wellcome-Wolfson Institute for Experimental Medicine, Medical School, Queen's University Belfast, Belfast, UK*
[3]*Department of Quantitative Methods in Economics and TiDES Institute, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain*

Correspondence should be addressed to Enrique Calderín–Ojeda; ecalderin@unimelb.edu.au

Received 24 January 2022; Revised 14 March 2022; Accepted 2 April 2022; Published 6 June 2022

Academic Editor: Xiaofeng Li

Continuous advancements in biotechnology are generating new knowledge and data sources that might be of interest for the insurance industry. A paradigmatic example of these advancements is genetic information which can reliably notify about future appearance of certain diseases making it an element of great interest for insurers. However, this information is considered by regulators in the highest confidentiality level and protected from disclosure. Recent investigations have shown that the microbiome can be correlated with several health conditions. In this paper, we examine the potential use of microbiome information as a potential tool for cardiovascular diagnosis. By using a recent dataset, we analyze the relation of some variables associated to coronary illnesses and several components of the microbiome in the organism by using a new copula-based multivariate regression model for compositional data in the predictor. Our findings show that the coabundance group associated to *Ruminococcaceae-Bifidobacteriaceae* has a negative impact on the age for nonsedentary individuals. However, one should be cautious with this conclusion since environmental conditions also influence the baseline microbiome.

1. Introduction

In recent years, the advances in biomedical sciences and biotechnology have enabled an unprecedented leap forward in the amounts and variety of data and information available for research and other purposes. This has translated in new applications and the development of new approaches such as a personalised or precision medicine where the aim is to use these new data and information sources (mostly related with genetic/genomic information) for diagnostic and therapeutic purposes and tailoring them to individuals or groups of patients (see Ginsburg and Phillips [1]). However, genomic information is considered in the highest level of confidentiality and protected from disclosure. For these reasons, the use of genomic data sparked a debate around the ethics and limits associated with the use of this knowledge and information in different sectors how it could be eventually regulated. In an attempt to overcome this potential limitation, in this paper, we propose to explore new avenues and information sources and the use of microbiome as a potential tool for disease diagnosis.

The microbiome is defined as the set of microorganisms that live inside or on the organism and its analysis has attracted a great interest in the biomedical domain, particularly since the development of new technologies that have facilitated and reduced the costs of accessing this information. Microbiome is an extremely dynamic element, and it changes with time and environmental conditions and other external factors, such as diet, geographical location, or physical activity and even with interaction between microbes and microbes and the host. The advances in biotechnology allow researchers to measure dynamic behaviors of the microbiota at a large scale (see [2]). Recent studies have shown that differences in the microbiome composition are

correlated with an increasing number of conditions ranging from cardiovascular diseases, autoimmune diseases, metabolic diseases, or neurological disorders and mental health aspects [3–7]. Another interesting reference linking a well-established cohort in the biomedical domain (the Framingham Study) with changes in the microbiome for multiple relevant health parameters such as cardiovascular risk, metabolic syndrome, and diabetes is Walker et al. [8]. BMI and physical activity have been also studied in the context of the microbiome finding relationships with different microbiome compositions [9–12]. Quite often, these relationships have been studied under the umbrella of either different ages or other health conditions. For a recent review highlighting that physical activity has an impact in gut microbiota and that physical exercise could be used to control obesity and health (see [13]). Both BMI and physical activity are important factors considered in insurance underwriting. For example, a higher BMI was predominantly related to blood pressure and lipids, which is consistent with results found in the literature (see [14] or [15]). There also exists an association between obesity and higher BMI with all-cause mortality (see [16]). Besides, BMI is connected to increased cancer risk as was recently described by Bhaskaran et al. [17] in a recent paper. On the other hand, microbiome data present a singular challenge due to its inherently high-dimensional and sparse structure. To handle the high dimensionality and compositional nature of the data, Wang et al. [18] proposed a sparse microbial causal mediation model specifically; also, Zhang et al. [19] used an isometric log-ratio transformation of the relative abundances as the mediator variables between treatment and outcome. A statistical approach that enables the inclusion of all daily activity behaviors, based on the principles of compositional data analysis was described by Dumuid et al. [20].

In our cross-sectional analysis, using a sample of eligible individuals with unique microorganisms across the Indian microbiome population, due to the large number of operational taxonomic units available in the gut microbiome across the sample, a clustering analysis to reduce the dimensionality of the dataset was initially carried out. Then, the resulting proportions of each of the groups of bacterial coabundance are combined in compositional data predictors that will be used to jointly explain the relationship of age, BMI, and level of physical activity by using a new bivariate regression model for compositional data in the predictor. In this paper, the margins are a beta regression and mixture of logistic regression models. As the age in years of the individuals is restricted to the interval 18–65, a beta regression model indexed by mean and dispersion parameters is considered. This regression family is useful in situations where the dependent variable is continuous and restricted to a bounded interval. On the other hand, regardless of the gender and physical activity level, the empirical distribution of BMI in humans is bimodal. Therefore, choosing suitable parametric models that can capture this feature is crucial; for that reason, a mixture of logistic regression model has been chosen due to its flexibility and simplicity. These margins are linked via a t-copula. This is an elliptical copula that is particularly well suited for this purpose as they not only allow for separate modeling of the univariate marginal distributions from the dependency structure but also for covariate adjustment in the margins and uncertainty quantification of their dependence estimates. In addition, they can specify different levels of correlation between the marginals. The compositional data included in the linear predictor are rewritten as logarithms of ratios. Then, we perform estimation via inference for margins method to explain the age, BMI, and level of physical activity using as margins a beta regression and mixture of regression models. Although copula models have been widely applied to model the joint distributions with mixed margins, copula models with the margins proposed in this work with compositional data in the predictor have not been extensively studied in the literature.

The rest of the paper is structured as follows: in Section 2, an examination of a human microbiome dataset is carried out. Here, a cluster analysis to classify the proportion of the most significant bacterial coabundance groups in the sample is completed. Furthermore, an approach to deal with the implementation of microbiome data as compositional data in the predictor is presented. The relationship of age, body mass index, and physical activity level with microbiome is analyzed in Section 3. Here, we firstly consider the marginal relation of age given a level of physical activity with microbiome through a beta regression model. Next, we examine the connection of BMI with the microbiome given the level of physical activeness by using a mixture of logistics regression model. Later, the joint relationship of these variables is examined by using a t-copula. Finally, discussion and extensions conclude the paper.

2. Analysis of a Human Microbiome Dataset

In our analyses we use a dataset available in Dubey et al.'s [21] *LogMPIE* study. This dataset is freely accessible, and it may be downloaded from the *European Nucleotide Archive* (ENA) portal of the *European Bioinformatics Institute* (https://www.ebi.ac.uk/ena/data/view/PRJEB25642). In this study, as it was portrayed in the original description of the dataset, they identify and map the Indian gut microbiome. It was carried out in fourteen geographical locations. Individuals were uniformly selected across geographical regions and some variables associated with changes in the structure of microbiome such as BMI, age in years of the individual, restricted to the interval 18–65 and level of physical activity (sedentary-nonsedentary) and gender (male-female) were also considered in their study design. In addition, a subject is classified as an obese if his/her BMI is greater than 30. This study recorded data from 1004 eligible individuals and reported 993 unique microorganisms across the Indian microbiome population. Unfortunately, in this dataset neither a longitudinal analysis across time of individuals nor changes in the composition of microbiome in old subjects are available.

2.1. Cluster Analysis. In general, microbiome empirical distribution includes a high proportion of zero observation

and a truncation point mass to account for high values that are too sparse to model; for that reason, models that gives an accurate estimates of the true proportion of zeros have been considered in the literature (see [22] and [23]). In addition, given the dynamic character of the microbiome other techniques such as functional response regression on correlated longitudinal microbiome sequencing data has been recently considered in the literature [24]. In this work, in order to facilitate further analyses and reduce the dimensionality of our data set, we started carrying out a cluster analysis. A main task of exploratory data mining, to group a set of bacterial coabundance collections in such a way that objects in the same group or cluster are more similar to each other than to those in other clusters. We performed our clustering based on the coabundance of genus-like groups at a taxonomic level of species within a sample of 1004 subjects. A total of 993 bacterial genera were identified. The core microbiota analysis was completed by using the *Hierarchical Ordered Partitioning and Collapsing Hybrid* (HOPACH) package in **R** that can be downloaded from the *Bioconductor* website http://www.bioconductor.org/. This package includes the HOPACH clustering algorithm that assembles a hierarchical tree of clusters by recursively portioning the whole dataset while ordering and collapsing clusters at each level. In our analysis, we have discarded redgenus that contain at least a minimum relative abundance of 30%, i.e., 70% of zeros in the sample of 1004 individuals. The algorithm uses the MSS (Mean/Median Split Silhouette) criteria to identify the level of the tree with maximally homogeneous clusters. The correlation distance (cor) was the metric selected for clustering the microbiome species by calculating dissimilarities between variables. We have also used a nonparametric bootstrap to estimate the probability that each species belongs to each cluster and to better understand the variability of each cluster. For that reason, we employed the "boothopach" function by taking 1000 bootstrap resample datasets to obtain a suitable balance between precision and speed. As a result of this, we were able to group the microbiome in five groups containing different numbers of genera (see supplementary tables in Table 1). The five different group of bacteria (classes) identified from the cluster analyses could be associated with different taxonomic groups according to the most abundant or representative genus for each of the identified clusters. Groups 1 and 4 are the two largest groups in terms of number of taxonomic elements. Also, as in Group 1, a majority of members comes from the *Bacteroidales-Bacteroidaceae* group that represents almost 2/3 of the species contained in this cluster (17 out of 27 members), it could be related to *Bacteroidales-Bacteroidaceae* cluster. Group 4 is associated with *Lachnospiraceae* which represent almost 1/3 of the total in this group (7 out of 23 members). The other three groups (2, 3, and 5) were assigned to the *Ruminococcaceae-Bifidobacteriaceae* group (5 out of 19 members), *Negativicutes* group (4 out of 19 members), and *Pasteurellaceae* group (3 out of 15 members), respectively. The results and relationships between the different elements on each of the clusters are presented in Figure 1. Here, species close to each other in the tree are shown in a similar way. The ordered distance matrix shows the clustering structure. Similar clusters appear as blocks on the diagonal of this heatmap. Darker colours represent small distances whereas the lighter colours represent large distances. The identified clusters have different sizes and compositions, with two large coabundance clusters, grouping the majority of the genus analyzed. It is important to note that we have combined under the name Group 0 all the discarded operational taxonomic units, that is, all the species with more that 70% of zeroes in the sample.

Table 2 displays the mean, median, and standard deviation for each one of the coabundance groups. It is noticeable that the proportion of bacteria that belongs to Group 1 is higher in average than the proportion in the other groups. The variability is also larger for the first coabundance group.

In Figure 2, some ternary plots for different combinations of the bacterial groups are displayed. In particular we have compared the coabundance Group 1, with Group 2 (top left), Group 3 (top right), Group 4 (bottom left), and Group 5 (bottom right). In order to ensure that the total sum is one, we have combined the coabundance proportion for the rest of the groups in each graph as *Others*. Group 1 is always located at the top of each triangle. The proportion of coabundance of Group 1 is measured in terms of the horizontal lines, i.e., 0% of coabundance is measured in terms of base of the triangle (farthest from the vertex Group 1). In the lower left apex of each triangle is represented the groups compared to Group 1. The right side of the triangle now becomes the baseline for the percentage of the groups located in this vertex. Finally, the combined groups are located at the lower right apex of the triangle.

The rate of coabundance for the combined groups is calculated from the left side of the triangle (0% abundance) to the lower right corner (100% abundance). It is observable that the data lie from a high amount of coabundance of Group 1 and Group 2 with a low coabundance of third, fourth, and fifth groups (top left graph). From the rest of the graphs, it can be inferred that when Group 1 is compared to the other groups, the coabundance of these groups is lower than in the former graph. Also, as Group 2 has been included in the lot *Others*, the corresponding coabundance of the combined group is higher than in the top left graph.

2.1.1. Compositional Data Predictor. Compositional data can be defined as arrays of strictly positive numbers for which ratios between them are important without any further requirement [25]. Microbiome data are compositional, that is, the distance between component values is only meaningful proportionally (see [26]). The elements of the composition are non-negative and sum to unity. An important issue in microbiome data is the large presence of zeros; however, the issue of zero values in some components is not addressed in most papers and especially in the task of regression. In general, in compositional research problems, most of the basic statistical analysis tools are incorrect unless the variables are rewritten in terms of logarithms of ratios as proposed in the log-ratio methodology for compositional data. After computing these log-ratios, standard regression methods can be used since the relative character of the

Table 1: Composition of the five clusters identified in the analysis of the Indian dataset (Dubey et al., [21]).

Group 1
Bacteroidetes-Bacteroidia-Bacteroidales-Rikenellaceae-Alistipes-onderdonkii
Bacteroidetes-Bacteroidia-Bacteroidales-Rikenellaceae-Alistipes-putredinis
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-coprocola
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-coprophilus
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-dorei
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-fragilis
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-intestinalis
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-plebeius
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-thetaiotaomicron
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-uniformis
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-vulgatus
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-xylanisolvens
Bacteroidetes-Bacteroidia-Bacteroidales-Porphyromonadaceae-Barnesiella-intestinihominis
Firmicutes-Clostridia-Clostridiales-Eubacteriaceae-Eubacterium-biforme
Firmicutes-Clostridia-Clostridiales-Eubacteriaceae-Eubacterium-ventriosum
Firmicutes-Clostridia-Clostridiales-Clostridiales-Flavonifractor-plautii
Proteobacteria-Gammaproteobacteria-Enterobacteriales-Enterobacteriaceae-Klebsiella-variicola
Firmicutes-Bacilli-Lactobacillales-Lactobacillaceae-Lactobacillus-rogosae
Firmicutes-Negativicutes-Veillonellales-Veillonellaceae-Megasphaera-sp.
Bacteroidetes-Bacteroidia-Bacteroidales-Porphyromonadaceae-Odoribacter-splanchnicus
Bacteroidetes-Bacteroidia-Bacteroidales-Porphyromonadaceae-Parabacteroides-distasonis
Bacteroidetes-Bacteroidia-Bacteroidales-Porphyromonadaceae-Parabacteroides-merdae
Bacteroidetes-Bacteroidia-Bacteroidales-Prevotellaceae-Prevotella-copri
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Roseburia-sp.
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Ruminococcus-bromii
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Ruminococcus-callidus
Proteobacteria-Betaproteobacteria-Burkholderiales-Sutterellaceae-Sutterella-wadsworthensis
Group 2
Actinobacteria-Actinomycetales-Actinomycineae-Actinomycetaceae-Actinomyces-odontolyticus
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-massiliensis
Actinobacteria-Actinobacteria-Bifidobacteriales-Bifidobacteriaceae-Bifidobacterium-adolescentis
Actinobacteria-Actinobacteria-Bifidobacteriales-Bifidobacteriaceae-Bifidobacterium-bifidum
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Blautia-luti
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Blautia-wexlerae
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Butyrivibrio-crossotus
Actinobacteria-Coriobacteriia-Coriobacteriales-Coriobacteriaceae-Collinsella-aerofaciens
Firmicutes-Clostridia-Clostridiales-Eubacteriaceae-Eubacterium-eligens
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Faecalibacterium-prausnitzii
Firmicutes-Clostridia-Clostridiales-Clostridiales-Howardella-ureilytica
Proteobacteria-Gammaproteobacteria-Enterobacteriales-Enterobacteriaceae-Klebsiella-pneumoniae
Firmicutes-Negativicutes-Veillonellales-Veillonellaceae-Megasphaera-micronuciformis
Proteobacteria-Betaproteobacteria-Burkholderiales-Sutterellaceae-Parasutterella-excrementihominis
Firmicutes-Clostridia-Clostridiales-Peptostreptococcaceae-Peptostreptococcus-stomatis
Firmicutes-Negativicutes-Acidaminococcales-Acidaminococcaceae-Phascolarctobacterium-faecium
Spirochaetes-Spirochaetes-Spirochaetales-Spirochaetaceae-Treponema-succinifaciens
Firmicutes-Erysipelotrichia-Erysipelotrichales-Erysipelotrichaceae-Turicibacter-sanguinis
Lentisphaerae-Lentisphaeria-Victivallales-Victivallaceae-Victivallis-vadensis
Group 3
Proteobacteria-Alphaproteobacteria-Rhodospirillales-Acetobacteraceae-Acidiphilium-sp.
Verrucomicrobia-Verrucomicrobiae-Verrucomicrobiales-Akkermansiaceae-Akkermansia-muciniphila
Proteobacteria-Gammaproteobacteria-Aeromonadales-Succinivibrionaceae-Anaerobiospirillum-succiniciproducens
Actinobacteria-Actinobacteria-Bifidobacteriales-Bifidobacteriaceae-Bifidobacterium-longum
Proteobacteria-Deltaproteobacteria-Desulfovibrionales-Desulfovibrionaceae-Bilophila-wadsworthia
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Blautia-obeum
Firmicutes-Erysipelotrichia-Erysipelotrichales-Erysipelotrichaceae-Bulleidia-p-1630-c5
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Coprococcus-catus
Proteobacteria-Deltaproteobacteria-Desulfovibrionales-Desulfovibrionaceae-Desulfovibrio-piger
Firmicutes-Negativicutes-Veillonellales-Veillonellaceae-Dialister-succinatiphilus
Firmicutes-Clostridia-Clostridiales-Eubacteriaceae-Eubacterium-siraeum
Firmicutes-Negativicutes-Veillonellales-Veillonellaceae-Megasphaera-elsdenii

Table 1: Continued.

Firmicutes-Negativicutes-Selenomonadales-Selenomonadaceae-Mitsuokella-jalaludinii
Firmicutes-Negativicutes-Selenomonadales-Selenomonadaceae-Mitsuokella-multacida
Bacteroidetes-Bacteroidia-Bacteroidales-Prevotellaceae-Prevotella-stercorea
Firmicutes-Clostridia-Clostridiales-Clostridiaceae-Romboutsia-ilealis
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Ruminiclostridium-siraeum
Actinobacteria-Coriobacteriia-Eggerthellales-Eggerthellaceae-Slackia-isoflavoniconvertens
Proteobacteria-Betaproteobacteria-Burkholderiales-Sutterellaceae-Sutterella-sp.
Group 4
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-caccae
Bacteroidetes-Bacteroidia-Bacteroidales-Bacteroidaceae-Bacteroides-ovatus
Firmicutes-Erysipelotrichia-Erysipelotrichales-Erysipelotrichaceae-Catenibacterium-mitsuokai
Firmicutes-Clostridia-Clostridiales-Clostridiaceae-Clostridium-bartlettii
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Coprococcus-comes
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Coprococcus-eutactus
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Dorea-formicigenerans
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Dorea-longicatena
Firmicutes-Clostridia-Clostridiales-Eubacteriaceae-Eubacterium-hadrum
Firmicutes-Clostridia-Clostridiales-Eubacteriaceae-Eubacterium-hallii
Firmicutes-Clostridia-Clostridiales-Eubacteriaceae-Eubacterium-ramulus
Proteobacteria-Alphaproteobacteria-Rhizobiales-Hyphomicrobiaceae-Gemmiger-formicilis
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Lachnoclostridium-clostridioforme
Firmicutes-Clostridia-Clostridiales-Oscillospiraceae-Oscillibacter-sp.
Firmicutes-Negativicutes-Acidaminococcales-Acidaminococcaceae-Phascolarctobacterium-succinatutens
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Roseburia-faecis
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Roseburia-inulinivorans
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Ruminococcus-faecis
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Ruminococcus-gnavus
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Ruminococcus-sp.
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Ruminococcus-torques
Firmicutes-Negativicutes-Selenomonadales-Selenomonadaceae-Selenomonas-bovis
Proteobacteria-Betaproteobacteria-Burkholderiales-Sutterellaceae-Sutterella-stercoricanis
Group 5
Proteobacteria-Gammaproteobacteria-Pasteurellales-Pasteurellaceae-Actinobacillus-minor
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Blautia-faecis
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Blautia-producta
Firmicutes-Clostridia-Clostridiales-Clostridiaceae-Clostridium-disporicum
Firmicutes-Clostridia-Clostridiales-Clostridiaceae-Clostridium-perfringens
Firmicutes-Clostridia-Clostridiales-Clostridiaceae-Clostridium-sp.
Proteobacteria-Deltaproteobacteria-Desulfovibrionales-Desulfovibrionaceae-Desulfovibrio-D168
Proteobacteria-Gammaproteobacteria-Enterobacteriales-Enterobacteriaceae-Escherichia-coli
Proteobacteria-Gammaproteobacteria-Pasteurellales-Pasteurellaceae-Haemophilus-parainfluenzae
Proteobacteria-Gammaproteobacteria-Pasteurellales-Pasteurellaceae-Haemophilus-pittmaniae
Firmicutes-Bacilli-Lactobacillales-Lactobacillaceae-Lactobacillus-ruminis
Proteobacteria-Gammaproteobacteria-Pseudomonadales-Pseudomonadaceae-Pseudomonas-lini
Firmicutes-Clostridia-Clostridiales-Lachnospiraceae-Roseburia-intestinalis
Firmicutes-Clostridia-Clostridiales-Ruminococcaceae-Ruminococcus-gauvreauii
Firmicutes-Negativicutes-Veillonellales-Veillonellaceae-Veillonella-dispar

information is considered when analyzing the results, as one group or variable can only increase in relative terms if some other group or groups reduce. In this work we focus on the case of compositional data being included in the predictor variables. The effect of increasing one of the variables in relative terms in the predictor therefore depends on which other variables are decreased when this occur. In log-ratio parlance, the effect of increasing one log-ratio is interpreted while keeping all other log-ratios constant as the same log-ratio may have different meaning depending on the way that the other log-ratios in the model are assembled. Thus, the interpretation of log-ratios as explanatory variables is usually different from other approaches. Several different approaches of building and interpreting the log-ratios have been considered in the literature, often leading to the same predictions and residuals [27]. Among the different parametrizations, in this work, we have chosen centred log-ratios [28]. In our analysis, we consider a vector of 6-dimensional real space that carries information on the relative importance of its components,

$$\mathbf{x}_i = \left(x_{i0}, x_{i1}, \ldots, x_{i5}\right) \in \mathbb{R}_+^6, \text{ with } x_{ij} > 0,$$
$$j = 0, 1, 2, \ldots, 5, \sum_{j=0}^{5} x_{ij} = 1, \tag{1}$$

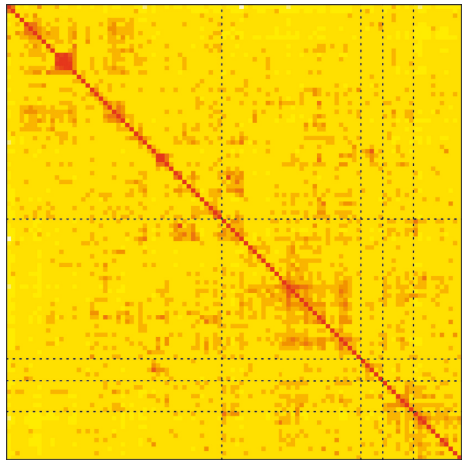


Figure 1: Heatmap of coabundance groups. This figure represents the distance among the abundances of the different genus of bacteria characterised in the sequencing analyses. The dashed lines depict the boundaries between the five different clusters. Darker colours represent closer distances (coabundance) between genus whereas lighter colours displays larger distances.

Table 2: Mean, median, and standard deviation for each coabundance group.

Coabundance group	Mean	Median	Standard deviation
Group 1	0.5548	0.5669	0.1542
Group 2	0.1899	0.1670	0.1024
Group 3	0.0816	0.0624	0.0688
Group 4	0.0899	0.0813	0.0517
Group 5	0.0603	0.0239	0.0859
Group 0	0.0235	0.0120	0.0632

where $i = 1, \ldots, 1004$. Note that the explanatory variables x_{ij} represent the proportion of the bacterial coabundance proportion of Group j in individual i. Centred log-ratios are calculated by using a quotient between each variable and the geometric mean of all components (see [28]),

$$\log_2\left(\frac{x_j}{\sqrt[6]{\prod_{j=0}^{5} x_j}}\right), \quad \text{with } j = 0, 1, 2, \ldots, 5. \tag{2}$$

The fact that we are using logarithms to base 2 means that a unit increase in this logarithm leads to a double increase in the original magnitude. In order to avoid perfect collinearity one centred log-ratio must be deleted from the regression equation. Since all six centred log-ratios add-up to zero, by increasing a fixed centred log-ratio while keeping the other four remaining log-ratios in the regression equation (with regressors β_j with $j = 0, 1, 2, \ldots, 5$) constant implies increasing the given centred log-ratio whilst reducing the omitted centred log-ratio by the same amount. In this regard, a positive statistically significant regression β_j coefficient indicates an increasing value of the covariate x_j at the expense of decreasing the amount of the omitted component has a significant positive effect on expected value of the response variable. This is equivalent to say that in terms of logarithm of base 2, β_j is interpreted as the expected change in the response variable when the ratio between x_j and the omitted explanatory variable is multiplied by six. Finally, in order to obtain the estimates and their corresponding p values for all possible pair combinations, the model needs to be repeated six different times by ignoring each time a different centred log-ratio.

3. Relation of Age, BMI, and Level of Physical Activity with Microbiome

In this section, we firstly consider the marginal relation of age given a level of physical activity with microbiome through a beta regression model. Next, we examine the connection of BMI with the microbiome given the level of physical activeness by using a mixture of logistics regression model. Finally, the joint relationship of these variables is examined via a t-copula.

3.1. Relation of Age and Level of Physical Activity with Microbiome. It is our interest to model the relationship between the age of the subject and the proportion of each coabundance genus-like groups at taxonomic level of species via a beta regression model (see Ferrari and Cribari-Neto [29]). This model assumes that the response variable is beta distributed using a parametrization of the beta law that is indexed by mean and dispersion parameters. This regression family is useful for modeling rates and proportions, that is, in situations where the dependent variable of interest is continuous and restricted to a bounded interval (a, b) where a and b are known scalars with $a < b$. This model is related to other variables through a regression structure. Our goal is to explain a continuous response variable Y_1 with $a < y_1 < b$. The density of Y_1 is defined as follows:

$$f_1(y_{1i}|\omega_i, \phi) = \frac{\Gamma(\phi)(b-a)^{1-\phi}}{\Gamma(\omega_i\phi)\Gamma((1-\omega_i)\phi)} \left(\frac{y_{1i}-a}{b-a}\right)^{\omega_i\phi-1} \left(\frac{b-y_{1i}}{b-a}\right)^{(1-\omega_i)\phi-1}, \tag{3}$$

with $0 < \omega_i < 1$ and $\phi > 0$ with $i = 1, \ldots, n$. This parametrization allows us to obtain a regression structure for the mean of the response along with a dispersion parameter ϕ. Here, n is the sample size and $E(Y_{1i}) = \omega_i(b-a) + a$. The variance of the response variable can be easily explained in terms of its mean by the following expression $\mathrm{Var}(Y_{1i}) = (b-a)^2 \omega_i(1-\omega_i)/1+\phi$. The variance decreases with the value of the dispersion parameter.

Let us now consider that a random variable Y_{1i} denoting age of the individual i in the sample is related to a compositional data predictor related to each one the coabundance groups, $\mathbf{x}_i = (1, u_{i1}, \ldots, u_{i5})^\top$ where $(u_{i1}, \ldots, u_{i5})$ are chosen among all combinations without repetition from the vector $(x_{i0}/\sqrt[6]{\prod_{j=0}^{5} x_j}, \ldots, x_{i5}/\sqrt[6]{\prod_{j=0}^{5} x_j})^\top$ taking 5 components at a time. Then, by using the logit link (i.e., $h(\omega_i) = \log\omega_1/1-\omega_i$), we have that $\omega_i = \exp(\mathbf{x}_\mathbf{i}^\top \underline{\beta})/1 + \exp(\mathbf{x}_\mathbf{i}^\top \underline{\beta})$, where $\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_5)^\top$ is

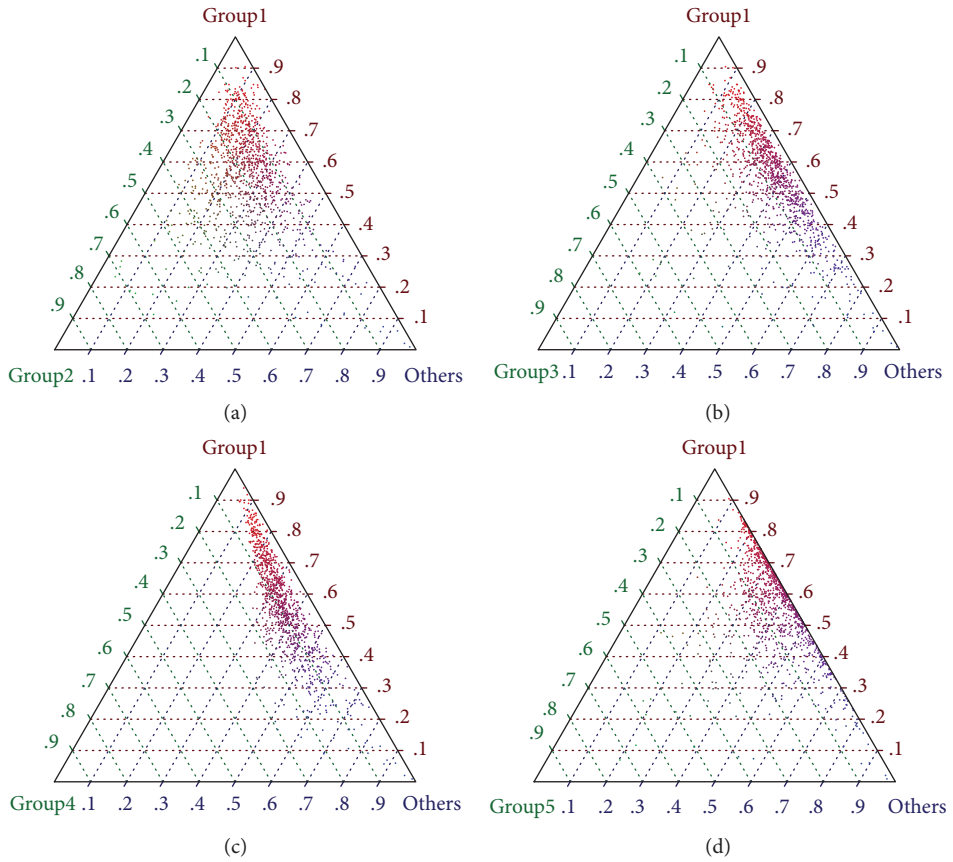


Figure 2: Ternary plots associated to different coabundance groups. (a) Group1/Group 2/Other groups. (b) Group 1/Group 3/Other groups. (c) Group 1/Group 4/Other groups. (d) Group 1/Group 5/Other groups.

Table 3: Parameter estimates (first row) and p values (second row) for the regressors associated with the six predictors for individuals classified as sedentary. The response variable is age.

	Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5	Predictor 6
Group 1	−0.0341	—	0.1369	0.0047	−0.0050	0.0921
	0.6026	—	0.1388	0.9593	0.9473	0.1262
Group 2	−0.1708	−0.1368	—	−0.1313	−0.1500	−0.0448
	0.0117	0.1388	—	0.0861	0.0686	0.4267
Group 3	−0.0404	−0.0064	0.1305	—	−0.0076	0.0857
	0.49404	0.9439	0.0880	—	0.7957	0.0956
Group 4	−0.0448	−0.0112	0.1253	−0.0046	—	0.0809
	0.4992	0.9017	0.1923	0.9515	—	0.2157
Group 5	−0.1260	−0.0921	0.0446	−0.0863	−0.1023	—
	0.0044	0.1261	0.4286	0.0931	0.0284	—
Group 0	—	0.0341	0.1708	0.0397	0.0145	0.1261
	—	0.6025	0.0117	0.5010	0.6334	0.0044
Intercept	−0.1209	−0.1210	−0.1214	−0.1178	−0.1088	−0.1210
	0.5175	0.5173	0.5157	0.5285	0.5659	0.5173
ϕ	2.9797	2.9795	2.9796	2.9789	2.9807	2.9795
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
AIC	3481.07	3481.07	3481.07	3481.07	3480.86	3481.07

a vector of regressors. Other choices for the link function link functions for the response model are feasible.

We have fitted this beta regression model to this dataset to explain the response variable *Age* by considering two levels of physical activity: sedentary and nonsedentary by assuming $a = 17.5$ and $b = 65.5$. Below in Table 3, the estimates and p values associated with the six predictors for individuals classified as sedentary for each microbiome coabundance group's proportion obtained under the regression model (1). Similarly, in Table 4 estimates and p

Table 4: Parameter estimates (first row) and p values (second row) for the regressors associated with the six predictors for individuals classified as nonsedentary. The response variable is age.

	Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5	Predictor 6
Group 1	−0.0465	—	0.1209	−0.0105	−0.0745	0.0159
	0.4022	—	0.0283	0.8851	0.3583	0.7459
Group 2	−0.1673	−0.1209	—	−0.1328	−0.1954	−0.1052
	0.0063	0.1151	—	0.0461	0.0344	0.0459
Group 3	−0.0358	0.0108	0.1318	—	−0.0637	0.0265
	0.4957	0.8814	0.0477	—	0.4014	0.5497
Group 4	0.0277	0.0745	0.1954	0.0641	—	0.0899
	0.6623	0.3583	0.0344	0.3983	—	0.1771
Group 5	−0.0623	−0.0157	0.1052	−0.0263	−0.0902	—
	0.1196	0.7484	0.0458	0.5522	0.1753	—
Group 0	—	0.0465	0.1674	0.0349	−0.0280	0.0622
	—	0.4021	0.0063	0.5069	0.6589	0.1206
Intercept	−0.1641	−0.1642	−0.1641	−0.1667	−0.1643	−0.1649
	0.2890	0.2888	0.2892	0.2820	0.2887	0.2869
ϕ	2.9002	2.8999	2.9000	2.8981	2.8999	2.8999
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
AIC	4155.08	4155.08	4155.08	4155.08	4155.08	4155.08

values results for each predictor are shown for nonsedentary subjects. From these tables, it is discernible that for the first predictor the regressor associated to *Group 5* for sedentary individuals is statistically significant at the 5% level whereas it is not for nonsedentary subjects. Its value, −0.1260, is interpreted as the decrease in the covariate x_5 (i.e., *Pasteurellaceae*) at the expense of increasing the amount of x_0 has a significant negative effect on $h(\omega_i)$ while keeping the remaining four log-ratios in the equation constant. In a similar fashion for the third predictor for the nonsedentary subjects the explanatory variables, x_1, x_3, x_4, and x_5 are significant at the same level while they are not for the sedentary individuals. For all these covariates, the sign of their regressors is positive; therefore, an increase in these regressors at the expense of decreasing the value of x_2 has a significant effect on the transformation of the expected value of the mean of the model. For the fourth predictor, the regression coefficient associated to the second group is only significant for the nonsedentary party. Similarly for the fifth predictor, x_5 is significant for the sedentary individuals whereas the regressor associated to *Group 2*, i.e., *Ruminococcaceae-Bifidobacteriaceae* is significant for the nonsedentary individuals. Similar situation is also verified for the sixth predictor. On the other hand, for the sedentary subjects, the variable x_0 is a positive significant variable.

In Figure 3, we have plotted the histograms of the empirical distribution of the response variable *Age* for the nonsedentary (top left panel) and sedentary (bottom left panel) subjects. For both histograms, we have superimposed the probability density function of the beta distribution. It is observable that this distribution provides a better fit to empirical data for the nonsedentary group than for the sedentary party. Furthermore, we have performed a diagnostic analysis to check the goodness-of-fit of the estimated model by providing a global measure of explained variation and graphical tools based on QQ-plots, to detect departures from the given model and influential observations. Residuals are used to check the appropriateness of a chosen model and to identify outliers. For that reason, randomized quantile residuals (Dunn and Smyth [30]) are used since other type of residuals, i.e., Pearson's and deviance residuals are far from normality when the parameters of the model are known and they fail to provide useful information of the inadequacy of the model. The ith randomized quantile residuals for a discrete response variable is defined as $r_{q,i} = \Phi^{-1}(F(y_{1i}; \widehat{\omega}_i, \widehat{\phi}))$ where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution and $F(y_{1i}; \widehat{\omega}_i, \widehat{\phi})$ is the cumulative distribution function associated to the beta regression model evaluated at the estimated parameters for $i = 1, \ldots, n$. In the right panels of Figure 3, the QQ-plots of the randomized quantile residuals of the beta regression models when the predictor 1 is considered for the nonsedentary (top right) and sedentary (bottom right) subjects. Each dot on the plots represents an empirical residual. A perfect alignment with the 45° line implies that the residuals are normally distributed. In general, it is observable that the residuals for the nonsedentary group adhere closer to the line in the whole distribution.

3.2. Relation of BMI and Level of Physical Activity with Microbiome. Regardless of the gender, the empirical distribution of BMI in humans is bimodal. Then, finding appropriate statistical models that have the capacity to explain bimodal datasets is an issue of vital importance. In this work, we use a mixture of two logistic distributions with different locations and scale parameters. We have chosen this family for its flexibility and simplicity. It is now our interest to explain the BMI in the population in terms of a random variable $Y_2 \in \mathbb{R}$. The probability density function of this random variable is

$$f_2(y_{2i}|w_i, \mu_1, \sigma_1, \mu_2, \sigma_2) = \frac{w_i}{4\sigma_1}\operatorname{sech}^2\left(\frac{y_{2i} - \mu_1}{2\sigma_1}\right), \quad (4)$$

$$+\frac{1 - w_i}{4\sigma_2}\operatorname{sech}^2\left(\frac{y_{2i} - \mu_2}{2\sigma_2}\right), \quad (5)$$

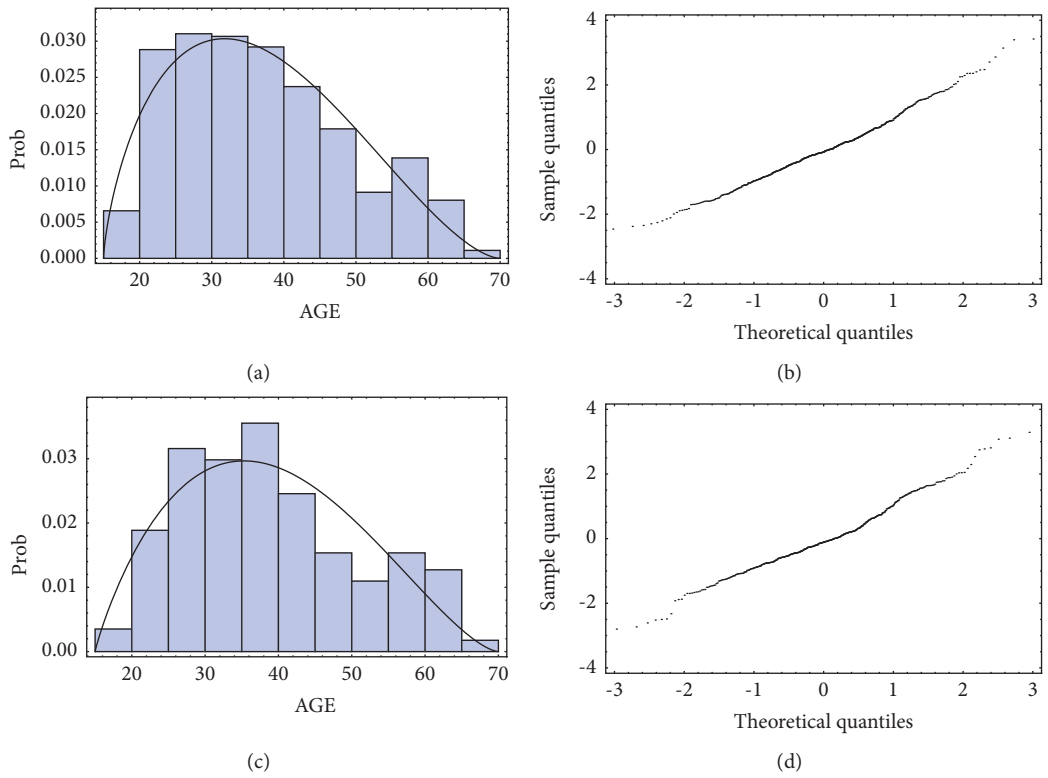


Figure 3: Histograms of the *Age* distributions and the beta density superimposed and QQplots of the randomized quantile residuals (RQRs) for each regression model by using Predictor 1. (a) Histogram of age and nonsedentary. (b) RQR of age and nonsedentary against Predictor 1 regression model. (c) Histogram of age and sedentary. (d) RQR of age and sedentary against Predictor 1 regression model.

where $\mu_1, \mu_2 \in \mathbb{R}$ are location parameters and $\sigma_1, \sigma_2 > 0$ are scale parameters. In this model, it will also be assumed that the weight parameter for each individual in the sample is again expressed as a function of the same group of covariates, $w_i = \exp(\mathbf{x}_i^\top \underline{\alpha})/1 + \exp(\mathbf{x}_i^\top \underline{\alpha})$, where $\underline{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_5)^\top$ is a vector of regressors. Other choices for the link function link functions for the response model are also possible. This parametrization enable us to obtain a regression structure for the mean of the response in the following way, $E(Y_{2i}) = w_i\mu_1 + (1 - w_i)\mu_2$. The variance of the response variable is written in terms of the following linear combination of the scale parameters:

$$\mathrm{Var}(Y_{2i}) = \frac{\pi^2}{3}\left\{(w_i\sigma_1)^2 + ((1 - w_i)\sigma_2)^2\right\}. \tag{6}$$

We have now fitted the mixture of logistics regression model given by (5) to this dataset to explain the dependent variable *BMI* by again considering two levels of physical activity: sedentary and nonsedentary. In Table 5, the estimates and p values associated with the six predictors for individuals classified as sedentary for each microbiome coabundance groups proportion obtained under this mixture of logistics regression model (5). In a similar way, in Table 6, estimates and p values results for each predictor are shown for nonsedentary subjects. From these tables, it is apparent that for the first predictor the regressor associated to *Group 1* for sedentary individuals is statistically significant at the 5% level whereas it is not for non-sedentary subjects. The estimated value is −0.3631, that it is interpreted as the decrease in the covariate x_1 (i.e., *Bacteroidales-Bacteroidaceae*) at the expense of increasing the amount of x_0 has a significant negative effect on $h(\omega_i)$ while keeping the remaining four log-ratios in the equation of the predictor constant. Similarly, for the second predictor and the sedentary subjects, the explanatory variables, x_0 is statistically significant at the 10% significance level while it is not for the nonsedentary individuals. The sign of this regression coefficient is positive; therefore, an increase in this regressor at the expense of decreasing the value of x_1, while keeping the other four log-ratios in the equation constant has a significant effect on the transformation of the expected value of the mean of the model. Finally, for predictor 3, the regression coefficient associated to the first group is only significant at the 10% significance level for the sedentary individuals. The sign of this regressor is negative.

In Figure 4, we have plotted the histograms of the empirical distribution of the response variable *BMI* for the nonsedentary (top left panel) and sedentary (bottom left panel) individuals. For both histograms, we have superimposed the density function of the mixture of logistics distributions. It can be seen that this distribution is able to reproduce the two modes of the empirical distribution for both cohorts. Note that for the group of sedentary individuals, the second modal value located around the *BMI* value of 31 is clearly more predominant. Once again, we have plotted the QQ-plots of the randomized quantile residuals of this mixture of logistics regression when the first predictor 1

Table 5: Parameter estimates (first row) and p values (second row) for the regressors associated with the six predictors for individuals classified as sedentary. The response variable is BMI.

	Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5	Predictor 6
Group 1	−0.3631	—	−0.5055	−0.2915	−0.1853	−0.1829
	0.0492	—	0.0534	0.2393	0.4711	0.2635
Group 2	0.2200	0.3914	—	0.2154	0.3215	0.1421
	0.2429	0.1217	—	0.3117	0.2173	0.3455
Group 3	−0.0267	0.2575	−0.2111	—	0.1062	−0.0071
	0.8716	0.2966	0.3209	—	0.6199	0.9604
Group 4	−0.1777	0.0996	−0.3113	−0.1064	—	−0.1145
	0.3450	0.6939	0.2281	0.6195	—	0.5254
Group 5	−0.0527	0.2175	−0.2280	−0.0181	0.0881	—
	0.6648	0.1641	0.1379	0.9004	0.6246	—
Group 0	—	0.3143	−0.1292	0.0849	0.1911	0.1057
	—	0.0819	0.4838	0.6061	0.3049	0.3826
Intercept	0.3101	0.3894	0.3626	0.3633	0.3634	0.3374
	0.5688	0.4873	0.4983	0.4989	0.4989	0.5406
μ_1	24.4923	24.70000	24.4250	24.4537	24.4532	24.6451
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
μ_2	31.3471	31.4683	31.3505	31.3616	31.3613	31.4405
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
s_1	2.1269	2.1761	2.0905	2.0984	2.0984	2.1679
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
s_2	1.6660	1.6123	1.6664	1.6643	1.6644	1.6371
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
AIC	2717.15	2717.25	2716.83	2716.83	2716.83	2717.54

Table 6: Parameter estimates (first row) and p values (second row) for the regressors associated with the six predictors for individuals classified as nonsedentary. The response variable is BMI.

	Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5	Predictor 6
Group 1	0.0419	—	0.1453	0.0264	0.2168	0.0526
	0.6145	—	0.4500	0.8847	0.2696	0.6742
Group 2	−0.1095	−0.1466	—	−0.1185	0.0662	−0.1277
	0.4624	0.4459	—	0.4661	0.7654	0.3291
Group 3	0.0164	−0.0282	0.1186	—	0.1793	0.0202
	0.8980	0.8769	0.4660	—	0.3258	0.8540
Group 4	−0.1438	−0.1902	−0.0425	−0.1618	—	−0.1669
	0.3464	0.3322	0.8481	0.3746	—	0.2921
Group 5	0.0084	−0.0252	0.1204	0.0022	0.1827	—
	0.9295	0.8366	0.3558	0.9843	0.2488	—
Group 0	—	−0.0328	0.1126	−0.0056	0.1725	−0.0027
	—	0.8068	0.4497	0.9650	0.2594	0.9771
Intercept	0.2075	0.2282	0.2308	0.2307	0.1902	0.1731
	0.6145	0.5798	0.5754	0.5754	0.6438	0.6438
μ_1	23.9065	23.8923	23.8953	23.8921	23.8906	23.8869
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
μ_2	31.2608	31.2582	31.2606	31.2576	31.2563	31.2567
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
s_1	1.9478	1.9444	1.9434	1.9417	1.9408	1.9418
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
s_2	1.4769	1.4785	1.4779	1.4788	1.4790	1.4803
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
AIC	3264.49	3264.45	3264.45	3264.45	3264.47	3264.58

is considered for the nonsedentary (top right) and sedentary (bottom right) individuals. In general, it is observable that the residuals for the nonsedentary group adhere closer to the line in the whole distribution but it underestimates the top part of the distribution of residuals.

3.3. Joint Relation of Age, BMI, and Level of Physical Activity with Microbiome. The degree of association between the two variables age and BMI in the sample for the different levels of physical activity can be summarized in terms of some measures of correlation for bivariate data. In Table 7,

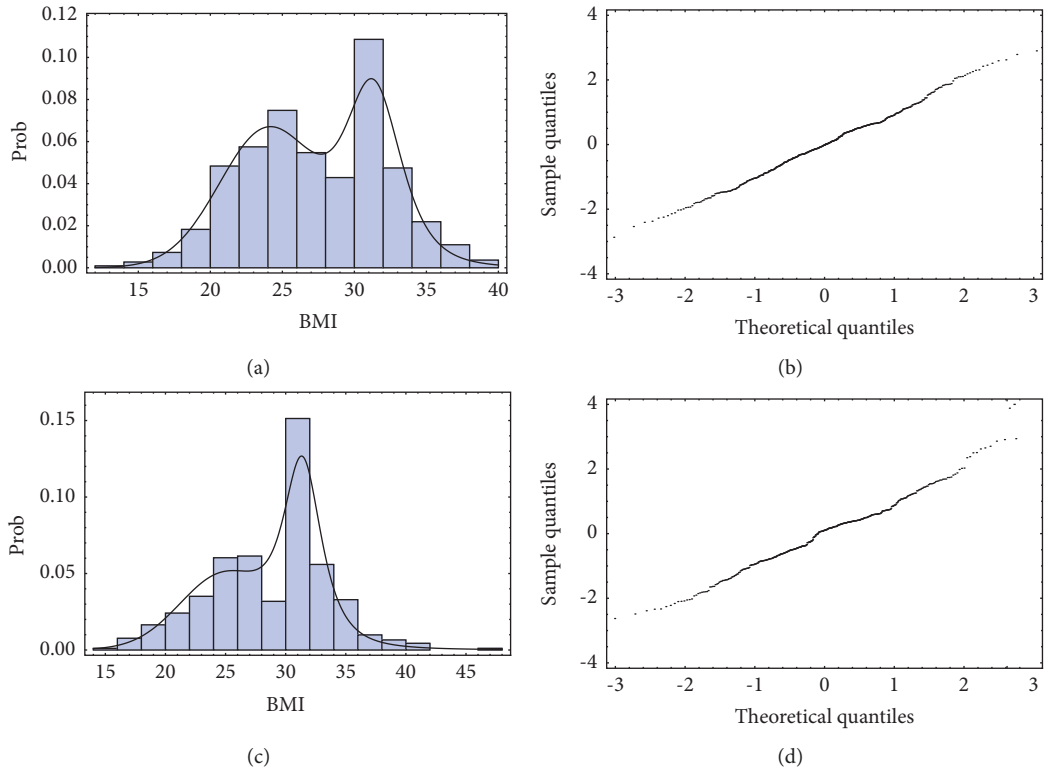


Figure 4: Histograms of the BMI distributions and the mixture of logistic density superimposed and QQplots of the randomized quantile residuals (RQR) for each regression model by using Predictor 1. (a) Histogram of BMI and nonsedentary. (b) RQR of BMI and nonsedentary against Predictor 1 regression model. (c) Histogram of BMI and sedentary. (d) RQR of BMI and sedentary against Predictor 1 regression model.

Table 7: Pearson's, Spearman's, and Kendall's measures of correlation for the variables age and BMI and different levels of physical activity.

Measure of correlation	Physical activeness level	
	Sedentary	Nonsedentary
Pearson's	0.2238	0.1738
Spearman's	0.2060	0.1828
Kendall's	0.1421	0.1242

Pearson's, Spearman's, and Kendall's measures of correlation for these continuous random variables are displayed. It is noticeable that there exists weak positive correlation between these two variables. The degree of association is less intense for the nonsedentary individuals.

We model the joint dependence of age and BMI for different level of physical activity and their relationship with the proportion of each coabundance genus-like groups at taxonomic level of species via a t-copula with degrees of freedom (df) parameter $\nu > 1$ with marginal distributions given by the beta regression model given in (1) and the mixture of logistics distributions provided in (5). The density of this of this multivariate distribution is defined as

$$g\left(\underline{y}_i|\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2,\boldsymbol{\Theta}_3\right) = \frac{\Gamma(\nu+1/2)\Gamma(\nu/2)\left(1+\underline{z}\,\Sigma^{-1}\underline{z}^{\top}/\nu\right)^{-\nu+2/2}}{\Gamma(\nu+1/2)^2|\Sigma|^{1/2}\prod_{j=1}^{2}\left(1+z_j^2/\nu\right)^{-\nu+1/2}} \times f_1\left(y_{i1}|\boldsymbol{\Theta}_1\right)\times f_2\left(y_{i2}|\boldsymbol{\Theta}_2\right), \tag{7}$$

where $\underline{y}_i := (y_{i1}, y_{i2})^{\top}$, $\Theta_1 = (\underline{\beta}, \phi)$, $\Theta_2 = (\underline{\alpha}, \mu_1, \sigma_1, \mu_2, \sigma_2)$, and $\Theta_3 = (\nu, \Sigma)$. Here, $\Sigma := (\rho_{ij})_{1\le i,j\le 2}$ is a symmetric and positive definite scatter matrix with dimension 2×2 with unit diagonal entries and $-1 < \rho_{ij} < 1$, $|\cdot|$ denotes the determinant of a matrix, and $\Gamma(\cdot)$ is the complete gamma function. Also, $\underline{z} := (z_1, z_2)^{\top}$ with $z_j = t_\nu^{-1}(F_j(y_j|\Theta_j))$ with $j = 1, 2$, where $t_\nu^{-1}(\cdot)$ is the quantile function of univariate t-distribution with ν df and $F_j(y_{ij}|\Theta_j)$ is the cdf

associated to the regression models presented above with $i = 1, \ldots, n$.

The corresponding log-likelihood function, given a sample $\mathbf{y} := \underline{y}_1, \ldots, \underline{y}_n$ is provided by

$$\begin{aligned}
\ell(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \boldsymbol{\Theta}_3 | \mathbf{y}) &= \ell_C(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \boldsymbol{\Theta}_3 | \mathbf{y}) + \ell_M(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2 | \mathbf{y}) \\
&= n \log \Gamma\left(\frac{\nu+1}{2}\right) + n \log \Gamma\left(\frac{\nu}{2}\right) - 2n \log \Gamma\left(\frac{\nu+1}{2}\right) \\
&\quad - \frac{n}{2}\log|\Sigma| - \frac{\nu+2}{2}\sum_{i=1}^{n}\log\left(1 + \frac{\underline{z}_i \Sigma^{-1} \underline{z}_i^{\top}}{\nu}\right) \\
&\quad + \frac{\nu+1}{2}\sum_{i=1}^{n}\sum_{j=1}^{2}\log\left(1 + \frac{z_{ij}^2}{\nu}\right) + n \log \Gamma(\phi) + n(1-\phi)\log(b-a) \\
&\quad - \sum_{i=1}^{n}\left(\log \Gamma(\omega_i \phi) + \log \Gamma((1-\omega_i)\phi\right) + \sum_{i=1}^{n}(\omega_i \phi - 1)\log\left(\frac{y_i - a}{b-a}\right) \\
&\quad + \sum_{i=1}^{n}((1-\omega_i)\phi - 1)\log\left(\frac{b-y_i}{b-a}\right) \\
&\quad + \sum_{i=1}^{n}\log\left(\frac{w_i}{4\sigma_1}\operatorname{sech}^2\left(\frac{y_{2i}-\mu_1}{2\sigma_1}\right) + \frac{1-w_i}{4\sigma_2}\operatorname{sech}^2\left(\frac{y_{2i}-\mu_2}{2\sigma_2}\right)\right),
\end{aligned} \tag{8}$$

where $\ell_C(\cdot)$ and $\ell_M(\cdot)$ are the log-likelihood functions of the copula and marginal model, respectively. Maximum likelihood estimation can be used to estimate the parameters of expression (8) via an adaptive maximization by parts (MBP) algorithm as described in [31], by using initial estimates $(\widehat{\Theta}_1^{(0)}, \widehat{\Theta}_2^{(0)}, \widehat{\Theta}_3^{(0)})$ generated by inference for margins algorithm. In the step k of this algorithm for $k = 1, 2, \ldots$, we find,

$$\begin{aligned}
\left(\widehat{\boldsymbol{\Theta}}_1^{(k)}, \widehat{\boldsymbol{\Theta}}_2^{(k)}\right) &= \arg\max\left\{\ell_M(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2 | \mathbf{y}) + \ell_C\left(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2 | \mathbf{y}, \widehat{\boldsymbol{\Theta}}_3^{(k-1)}\right)\right\}, \\
\boldsymbol{\Theta}_3^{(k)} &= \arg\max \ell_C\left(\boldsymbol{\Theta}_3 | \mathbf{y}, \widehat{\boldsymbol{\Theta}}_1^{(k)}, \widehat{\boldsymbol{\Theta}}_2^{(k)}\right).
\end{aligned} \tag{9}$$

The algorithm stops when a terminating condition between two consecutive iterations is reached, i.e.,

$$\left\|\ell\left(\widehat{\boldsymbol{\Theta}}_1^{(k)}, \widehat{\boldsymbol{\Theta}}_2^{(k)}, \widehat{\boldsymbol{\Theta}}_3^{(k)} | \mathbf{y}\right) - \ell\left(\widehat{\boldsymbol{\Theta}}_1^{(k-1)}, \widehat{\boldsymbol{\Theta}}_2^{(k-1)}, \widehat{\boldsymbol{\Theta}}_3^{(k-1)} | \mathbf{y}\right)\right\|_1 < 10^{-3}. \tag{10}$$

Finally, we have fitted the bivariate distribution given in (5) to the bivariate data set. Once again, the two levels of physical activeness have been considered. Results are shown in Table 8 for the sedentary case and Table 9 for the nonsedentary situation. When using the first predictor, i.e., the omitted covariate is x_0, the variable x_5 and x_1 are statistically significant at the 5% level of significance for the variables age and BMI, respectively, for sedentary individuals while they are not for the nonsedentary group. Also, the regressor associated with the covariates x_2 and x_0 for the *BMI* are significant at the same level of significance for the sedentary individuals whereas they are not for the nonsedentary group. The sign of these regression coefficients is negative. Conversely, for the third predictor and the variable *Age*, the regressors for the variables x_3 and x_5 are positive significant only for the nonsedentary subjects. In addition, the variable x_3 is positive significant for the variable BMI for the same level of physical activeness. Regression coefficients associated to x_2 in the fourth and fifth predictors (omitted variables x_3 and x_4, respectively) are negative significant for the response variable age. Finally, when the explanatory variable x_5 is deleted, the regressor linked to x_0 is positive significant at the 5% level for the sedentary subjects and response variable *Age*. Similarly, for this sixth predictor, the regressor associated to the variable x_2 for the same response variable is negative significant for individuals classified as nonsedentary.

4. Discussion and Extensions

Although genetic information can reliably inform about the future appearance of certain diseases and it is an element of great interest for different stakeholders, this genomic information is considered in the highest level of confidentiality and protected from disclosure. In this sense, in the insurance industry, a particularity of genomic information in this context is that it does not only provide information and knowledge about the individual taking the insurance but also in respect to their ancestors and descendants. As a consequence of these limitations, international regulators, e.g., the Council of Europe encourages insurers to update their actuarial bases according to relevant and new scientific knowledge and this may open the gates to explore new avenues and data types and information sources. As part of

Table 8: Parameter estimates (first row) and p values (second row) for the regressors associated with the six predictors for individuals classified as sedentary (bivariate regression model).

		Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5	Predictor 6
Age	Group 1	−0.0051	—	0.0603	−0.0000	0.0346	0.0857
		0.9316	—	0.5081	0.9997	0.6975	0.1475
	Group 2	−0.1580	−0.1524	—	−0.1233	−0.0983	−0.0437
		0.0178	0.0944	—	0.1011	0.2985	0.4303
	Group 3	−0.0462	−0.0293	0.1203	—	0.0167	0.0824
		0.4265	0.7436	0.1103	—	0.8232	0.1031
	Group 4	−0.0430	−0.0371	0.0880	−0.0065	—	0.0765
		0.5101	0.6761	0.0116	0.0745	—	0.2339
	Group 5	−0.1125	−0.1045	0.0161	−0.0825	−0.0603	—
		0.0099	0.0779	0.7711	0.1027	0.3485	—
	Group 0	—	0.0126	0.1463	0.0440	0.0645	0.1257
		—	0.8448	0.0282	0.4480	0.3238	0.0039
	Intercept	−0.2183	−0.1924	−0.0095	−0.1094	−0.1543	−0.1174
		0.2361	0.2958	0.9587	0.5517	0.4014	0.5226
	ϕ	3.1241	3.1261	3.1143	3.1354	3.1258	3.1411
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
BMI	Group 1	−0.3887	—	−0.4573	−0.2960	−0.1711	−0.2530
		0.0292	—	0.0642	0.2154	0.4850	0.1132
	Group 2	0.1077	0.4876	—	0.1827	0.3142	0.2227
		0.5429	0.0500	—	0.3680	0.2120	0.1314
	Group 3	−0.0647	0.3060	−0.1777	—	0.1355	0.0421
		0.6835	0.2018	0.3813	—	0.5090	0.7622
	Group 4	−0.2022	0.1727	−0.3104	−0.1363	—	−0.0920
		0.2598	0.4813	0.2169	0.5057	—	0.5961
	Group 5	−0.1169	0.2616	−0.2148	−0.0417	0.0902	—
		0.3182	0.1023	0.1444	0.7640	0.6033	—
	Group 0	—	0.3713	−0.1111	0.0640	0.1979	0.1079
		—	0.0362	0.5286	0.6857	0.2685	0.3546
	Intercept	0.4546	0.3999	0.3452	0.3700	0.3962	0.3764
		0.3794	0.4365	0.4995	0.4692	0.4403	0.4631
	μ_1	24.5370	24.5337	24.5521	24.5309	24.5345	24.5355
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	μ_2	31.4511	31.4477	31.4557	31.4509	31.4486	31.4552
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	s_1	1.9793	1.9764	1.9786	1.9708	1.9752	1.9791
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	s_2	1.5562	1.5576	1.5530	1.5526	1.5572	1.5599
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	ρ_{12}	0.2554	0.2550	0.2571	0.2557	0.2558	0.2557
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	ν	17.1151	17.1042	16.9295	16.8143	17.0653	16.9971
		0.0050	0.0050	0.0046	0.0043	0.0049	0.0047

this new vision, we have examined the potential use of microbiome information in some variables associated with the insurance underwriting. Recent investigations have shown that changes in the gut microbiome are associated to certain risk of pathologies could be a potential proximal predictor of disease onset.

Recently, in an unpublished work by using text mining techniques in life insurance literature and microbiome research, a significant overlap between certain diseases and health conditions and other elements that are considered in insurance underwriting. One of these elements is the body mass index (BMI). This is one of the variables considered in the standard health declaration. Traditionally, this declaration is the first step in the risk assessment in health insurance underwriting practice. Certainly, depending on the level of insured capital and age of the policyholder, extra medical examination will be the obligatory required guarantee regardless of the outcome of the standard health declaration. However, medical examinations are expensive, disturbing for the applicant and time-consuming in the underwriting process (see [32]). The importance of BMI is linked to obesity that will lead to large number of chronic diseases, and consequently increase health expenditures and claims costs. Therefore, an early detection of obesity is crucial to safeguard the financial structure of the health insurance provider. Similar conclusions can be drawn about the early detection of cardiovascular, mental metabolic or immune diseases. Then, it is extremely important for the private health insurers to monitor their policyholders' health status in order to reduce future claims costs.

Table 9: Parameter estimates (first row) and p values (second row) for the regressors associated with the six predictors for individuals classified as nonsedentary (bivariate regression model).

		Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5	Predictor 6
Age	Group 1	−0.0256	—	0.1122	−0.0160	−0.0783	0.0123
		0.6443	—	0.1419	0.8252	0.3330	0.8009
	Group 2	−0.1632	−0.1279	—	−0.1313	−0.1980	−0.1032
		0.0076	0.0952	—	0.0478	0.0317	0.0493
	Group 3	−0.0324	0.0075	0.1344	—	−0.0650	0.0283
		0.5360	0.9174	0.0425	—	0.3906	0.5205
	Group 4	0.0353	0.0749	0.1917	0.0617	—	0.0913
		0.5767	0.3550	0.0371	0.4146	—	0.1690
	Group 5	−0.0546	−0.0205	0.1022	−0.0299	−0.0950	—
		0.1714	0.6748	0.0513	0.4980	0.1527	—
	Group 0	—	0.0420	0.1640	0.0323	−0.0328	0.0622
		—	0.4480	0.0072	0.5383	0.6037	0.1197
	Intercept	−0.2207	−0.1789	−0.1543	−0.1667	−0.1726	−0.1595
		0.1532	0.2471	0.3170	0.2802	0.2636	0.3015
	ϕ	2.9231	2.9215	2.9357	2.9269	2.9234	2.9294
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
BMI	Group 1	0.0183	—	0.1122	0.0182	0.1886	0.0139
		0.8906	—	0.1419	0.9198	0.3338	0.9090
	Group 2	−0.1063	−0.1419	—	−0.1120	0.0545	−0.1103
		0.4730	0.4582	—	0.4883	0.8052	0.3944
	Group 3	0.0070	−0.0286	0.1344	—	0.1682	0.0003
		0.9561	0.8746	0.0426	—	0.3537	0.9933
	Group 4	−0.1635	−0.1953	−0.0549	−0.1703	—	−0.1644
		0.2825	0.3168	0.8036	0.3474	—	0.2959
	Group 5	0.0037	−0.0281	0.1133	−0.0023	0.1680	—
		0.9687	0.8168	0.3815	0.9831	0.2856	—
	Group 0	—	−0.0344	0.1074	−0.0067	0.1623	−0.0054
		—	0.7963	0.4682	0.9576	0.2860	0.9546
	Intercept	0.2518	0.2081	0.2263	0.2352	0.2327	0.2471
		0.5380	0.6100	0.5790	0.5641	0.5689	0.5455
	μ_1	23.8809	23.8742	23.8818	23.8866	23.8793	23.8795
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	μ_2	31.2662	31.2593	31.2647	31.2716	31.2667	31.2640
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	s_1	1.9226	1.9203	0.9188	1.9169	1.9224	1.9238
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	s_2	1.4653	1.4656	1.4624	1.4634	1.4665	1.4655
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	ρ_{12}	0.1970	0.1968	0.1965	0.1968	0.1970	0.1967
		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	ν	114.9900	110.6990	98.4523	103.0250	110.6760	106.2180
		0.5978	0.5840	0.5393	0.5879	0.5816	0.5676

The main findings of our analysis show that the second bacterial coabundance group associated to *Ruminococcaceae-Bifidobacteriaceae* has a significant negative effect on the expected value of the response variable *Age* for the nonsedentary individuals. This issue is verified not only for the marginal model associated with the response variable age but also for the joint regression model. Concerning the fifth coabundance group related to *Pasteurellaceae*, it was observed a positive impact on the expected value of age for the sedentary individuals in the marginal model; in contrast, for some predictors, a negative impact in the mean of the response variable age is noticed under the bivariate regression model. This fact is somehow consistent with the recent work of Jollet et al. [33], where it is described a randomized clinical trial that shows in two of the bacterial groups of the *Pasteurellaceae* group are related to the level of physical activity. However, the degree that this conclusion is valid is arguable since, in general it appears that the majority of people have a unique baseline microbiome that is influenced by environmental conditions. The standard microbiome composition is affected not only by the level of physical activity but also for other factors such as age, diet, and medication. In addition, we have only analyzed a single observation of the gut microbiome per individual limited to the age interval 18–65. It should be highly recommended to perform a longitudinal analysis to relate microbiome information to risk mortality factors and probability of developing certain pathologies. For example, to deal with the high proportion of zeroes in the operational taxonomic units observed, a two part zero-inflated regression model with

random effects could be used [26]. In this regard, the *American Gut Project* (a citizen science project containing more than 10,000 samples not only from the USA but also from several other countries around the world including Australia, UK, or Spain) represents an opportunity to access microbiome data for a variety of age group. This source of information together with the *Human Mortality Database* available in https://www.mortality.org could be used to analyze how changes in the gut microbiome are related to human longevity in different countries.

Data Availability

In our analyses, we use a dataset available in Dubey et al.'s [21] *LogMPIE* study. This dataset is freely accessible, and it may be downloaded from the European Nucleotide Archive (ENA) portal of the European Bioinformatics Institute (https://www.ebi.ac.uk/ena/data/view/PRJEB25642). Datasets are also included in the submission (Abundance.txt and Metadata.txt).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

E.C.O and G.L.C acknowledge Fundación MAPFRE for "Ignacio H. de Larramendi 2017" research grant that partially funded this work.

Supplementary Materials

Metadata: demographic information of the individuals. Table S1: microbiome abundance for each individual. (*Supplementary Materials*)

References

[1] G. S. Ginsburg and K. A. Phillips, "Precision medicine: from science to value," *Health Affairs*, vol. 37, no. 5, pp. 694–701, 2018.

[2] B. Chen and W. Xu, "Generalized estimating equation modeling on correlated microbiome sequencing data with longitudinal measures," *PLoS Computational Biology*, vol. 16, no. 9, Article ID c1008108, 2020.

[3] D. MacFabe, "Autism: metabolism, mitochondria, and the microbiome," *Global Advances in Health and Medicine*, vol. 2, no. 6, pp. 52–66, 2013.

[4] Y. Sanz, M. Olivares, A. Moya-Perez, and C. Agostoni, "Understanding the role of gut microbiome in metabolic disease risk," *Pediatric Research*, vol. 77, no. 1-2, pp. 236–244, 2015.

[5] B. Singh, N. Qin, and G. Reid, "Microbiome regulation of autoimmune, gut and liver associated diseases," *Inflammation and Allergy - Drug Targets*, vol. 14, no. 2, pp. 84–93, 2016.

[6] G. B. Stefano, R. Ptacek, J. Raboch, and R. M. Kream, "Microbiome: a potential component in the origin of mental disorders," *Medical Science Monitor*, vol. 23, pp. 3039–3043, 2017.

[7] H. Tilg and A. Kaser, "Gut microbiome, obesity, and metabolic dysfunction," *Journal of Clinical Investigation*, vol. 121, no. 6, pp. 2126–2132, 2011.

[8] R. L. Walker, H. Vlamakis, J. W. J. Lee et al., "Population study of the gut microbiome: associations with diet, lifestyle, and cardiometabolic disease," *Genome Medicine*, vol. 13, no. 1, p. 188, 2021.

[9] A. S. Meijnikman, O. Aydin, A. Prodan et al., "Distinct differences in gut microbial composition and functional potential from lean to morbidly obese subjects," *Journal of Internal Medicine*, vol. 288, no. 6, pp. 699–710, 2020.

[10] J. Bai, Y. Hu, and D. W. Bruner, "Composition of gut microbiota and its association with body mass index and lifestyle factors in a cohort of 7-18 years old children from the American Gut Project," *Pediatric obesity*, vol. 14, no. 4, Article ID e12480, 2019.

[11] F. Valeriani, F. Gallé, M. S. Cattaruzza et al., "Are nutrition and physical activity associated with gut microbiota? a pilot study on a sample of healthy young adults," *Annali di Igiene: Medicina Preventiva e di Comunita*, vol. 32, no. 5, pp. 521–527, 2020.

[12] M. A. Stanislawski, D. Dabelea, L. A. Lange, B. D. Wagner, and C. A. Lozupone, "Gut microbiota phenotypes of obesity," *Npj-Biofilms and Microbiomes*, vol. 5, no. 1, p. 18, 2019.

[13] J. Aragón-Vela, P. Solis-Urra, F. J. Ruiz-Ojeda, A. I. Álvarez-Mercado, J. Olivares-Arancibia, and J. Plaza-Diaz, "Impact of exercise on gut microbiota in obesity," *Nutrients*, vol. 13, no. 11, p. 3999, 2021.

[14] S. Dua, M. Bhuker, P. Sharma, M. Dhall, and S. Kapoor, "Body mass index relates to blood pressure among adults," *North American Journal of Medical Sciences*, vol. 6, no. 2, pp. 89–95, 2014.

[15] L. Shamai, E. Lurix, M. Shen et al., "Association of Body Mass Index and lipid profiles: evaluation of a broad spectrum of body mass index patients including the morbidly obese," *Obesity Surgery*, vol. 21, no. 1, pp. 42–47, 2011.

[16] K. M. Flegal, B. I. Graubard, D. F. Williamson, and M. H. Gail, "Cause-specific excess deaths associated with underweight, overweight, and obesity," *JAMA*, vol. 298, no. 17, p. 2028, 2007.

[17] K. Bhaskaran, I. Douglas, H. Forbes, I. dos Santos-Silva, D. A. Leon, and L. Smeeth, "Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults," *The Lancet*, vol. 384, no. 9945, pp. 755–765, 2014.

[18] C. Wang, J. Hu, M. J. Blaser, and H. Li, "Estimating and testing the microbial causal mediation effect with high-dimensional and compositional microbiome data," *Bioinformatics*, vol. 36, no. 2, pp. 347–355, 2020.

[19] H. Zhang, J. Chen, Y. Feng, C. Wang, H. Li, and L. Liu, "Mediation effect selection in high-dimensional and compositional microbiome data," *Statistics in Medicine*, vol. 40, no. 4, pp. 885–896, 2021.

[20] D. Dumuid, T. E. Stanford, J. A. Martin-Fernández et al., "Compositional data analysis for physical activity, sedentary time and sleep research," *Statistical Methods in Medical Research*, vol. 27, no. 12, pp. 3726–3738, 2018.

[21] A. K. Dubey, N. Uppadhyaya, P. Nilawe et al., "LogMPIE, pan-India profiling of the human gut microbiome using 16S rRNA sequencing," *Scientific Data*, vol. 5, no. 1, 2018.

[22] K. Shestopaloff, M. D. Escobar, B. Graubard, and W. Xu, "Analyzing differences between microbiome communities using mixture distributions," *Statistics in Medicine*, vol. 37, no. 27, pp. 4036–4053, 2018.

[23] V. Jonsson, T. Osterlund, O. Nerman, and E. Kristiansson, "Modelling of zero-inflation improves inference of metagenomic gene count data," *Statistical Methods in Medical Research*, vol. 28, no. 12, pp. 3712–3728, 2019.

[24] B. Chen and W. Xu, "Functional response regression model on correlated longitudinal microbiome sequencing data," *Statistical Methods in Medical Research*, vol. 31, no. 2, pp. 361–371, 2022.

[25] J. J. Egozcue and V. Pawlowsky-Glahn, "Compositional data: the sample space and its structure," *Test*, vol. 28, no. 3, pp. 599–638, 2019.

[26] X. Yingling, J. Sun, and D. Chen, *Statistical Analysis Of Microbiome Data With R. ICSA Book Series in Statistics*, Springer Nature, Singapore, 2018.

[27] J. Aitchison, "The statistical analysis of compositional data," *Monographs on Statistics and Applied Probability*, Chapman and Hal, London, UK, 1986.

[28] J. Aitchison, "Principal component analysis of compositional data," *Biometrika*, vol. 70, no. 1, pp. 57–65, 1983.

[29] S. Ferrari and F. Cribari-Neto, "Beta regression for modelling rates and proportions," *Journal of Applied Statistics*, vol. 31, no. 7, pp. 799–815, 2004.

[30] P. K. Dunn and G. K. Smyth, "Randomized quantile residuals," *Journal of Computational & Graphical Statistics*, vol. 5, no. 3, pp. 236–244, 1996.

[31] R. Zhang, C. Czado, and A. Min, "Efficient maximum likelihood estimation of copula based meta *t*-distributions," *Computational Statistics & Data Analysis*, vol. 55, pp. 1196–1214, 2011.

[32] M. Rothstein, *Genetics and Life Insurance: Medical Underwriting and Social Policy*, MIT Press, Boston, 2004.

[33] M. Jollet, K. Nay, A. Chopard et al., "Does physical inactivity induce significant changes in human gut microbiota? New answers using the dry immersion hypoactivity model," *Nutrients*, vol. 13, no. 11, p. 3865, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 9538295, 9 pages
https://doi.org/10.1155/2022/9538295

Research Article

Fast Recognition Algorithm for Human Motion Posture Using Multimodal Bioinformation Fusion

Xiangbing Zhao and Jianhui Zhou

School of Computer and Network Engineering, Shanxi Datong University, Datong 037009, China

Correspondence should be addressed to Jianhui Zhou; zhoujianhui@sxdtdx.edu.cn

Received 22 February 2022; Accepted 28 March 2022; Published 15 April 2022

Academic Editor: Xiaofeng Li

To address the problems of low feature extraction accuracy, large bias of human motion pose recognition and posture recognition error, poor recognition effect, and low recognition rate of traditional human motion posture fast recognition algorithm, we propose a human motion posture fast recognition algorithm using multimodal bioinformation fusion. First, wavelet packet decomposition with sample entropy is used to extract the human motion posture hand features such as kurtosis, time domain feature skewness, and frequency domain feature electromyogram (EMG) integral value and time domain features such as mean, standard deviation, and interquartile distance of leg motion amplitude. Second, after normalizing the two features, the human hand and leg motion feature set is obtained, and finally the feature set is used to construct a human motion posture fast recognition model based on multimodal bioinformation fusion, and the feature set is input into the recognition model, which completes the fusion of human motion posture information by improving the typical correlation analysis method, and the fusion result is used as the input of the minimum distance classifier to achieve human motion posture fast recognition. The results show that the proposed algorithm has high accuracy of feature extraction, small bias of human motion posture recognition, the posture recognition error is -0.21~0.02, the recognition rate is always above 95%, and the practical application effect is good.

1. Introduction

Data fusion technology is used mainly to classify and process multiple data sources. After be processed, data can be classified into clearer and more perfect categories [1, 2]. Multimodal biological information fusion mainly uses more biological features to obtain more category information [3], that is, it can identify biological information more comprehensively. As there are a variety of human motion postures and different motion features will lead to different motion postures, so how to accurately identify human motion postures has become a topic that needs urgent study at present [4]. The recognition of human motion posture can help to solve problems in many fields, such as the analysis of the motion status of athletes, elderly people, and young people, in order to enhance the monitoring of human health levels [5]. Therefore, it is very important to study the method of human motion posture recognition.

More scholars have researched on posture recognition and more excellent research results have been achieved; for example, Mou et al. [6] studied attention-based CNN-LSTM multimodal fusion driver stress detection algorithm using attention-based convolutional neural network (CNN) and long and short-term memory (LSTM) models to fuse nonintrusive data, including eye data, vehicle data, and environmental data. Then, the features are automatically extracted from each modality separately, and the features of different modalities are given different levels of attention by self-attention mechanisms. The multimodal fusion driver stress detection is achieved using different attention mechanisms. However, the recognition effect of this algorithm is not perfect, and the information of biometric features is not fully fused when fusing motion features. Zhang et al. [7] studied fall detection based on the spatiotemporal evolution of human posture, using an improved two-branch multilevel convolutional neural

network to extract and construct the inverted pendulum structure of human posture in real complex scenes. Multimedia analysis is used to observe the time series changes of the human inverted pendulum structure and construct the spatio-temporal evolution of human posture motion. The visual features of the spatio-temporal evolution of human posture under potential instability are analyzed, and two key features of human falling behavior are explored, which are used for posture recognition and fall recognition. However, the algorithm is only able to recognize human fall posture and is less in recognizing other human motion situations. Han et al. [8] studied a two-stage fall recognition algorithm based on human posture features, which extracted human posture features when the human was in an unstable state by classifying the human state into three states: stable state, fluctuating state, and disordered state based on the trend sign and stable sign variables integrated by the scattered key features. Support vector machine, K-nearest neighbor, decision tree, and random forest are used to classify and combine the classification results to achieve human fall posture detection. However, the algorithm cannot recognize the human at rest and therefore has a weak recognition capability. Ding et al. [9] studied online adaptive prediction of human motion intention based on surface electromyogram (EMG) signals and designed a surface EMG feature extraction network and an online adaptive network. Highly compressed surface EMG features were obtained using a convolutional self-coding network combined with muscle synergistic features to assist motion prediction, and human motion posture recognition was achieved by combining the motion prediction results. However, the algorithm cannot accurately extract human motion features, leading to errors in recognition. Olivas-Padilla et al. [10] studied wearable devices for stochastic biomechanical modeling and recognition of human motion, constructing a gesture-based manipulation model, which uses an autoregressive model to learn the dynamics of joints by assuming associations between them. The statistical significance of each model hypothesis is calculated to identify the joints most involved in the motion, and this result is combined to achieve human motion posture recognition. However, the algorithm is more based on the device to achieve recognition, and if the device is abnormal, it leads to a decrease in the level of recognition.

To solve the problems of the above algorithms, we propose a fast human motion posture recognition algorithm using multimodal bioinformation fusion. The main contributions of this paper are as follows: (1) accurate recognition of posture is achieved by fusion of multimodal biometric information. (2) The time domain features of human motion posture hand features and leg motion amplitude are extracted, using these features can lay a solid foundation for the subsequent accurate recognition of human motion posture. (3) The fusion of human motion posture information is completed by improving the typical correlation analysis method, which improves the fusion accuracy and speed, and combines the fusion results to achieve fast human posture recognition, thus improving the recognition accuracy and efficiency.

2. Methodology

2.1. Human Motion Pose Feature Extraction

2.1.1. Hand Motion Feature Extraction in Human Motion Postures. In this study, the characteristic peak value and skewness of hand characteristic EMG signal and intensity in human motion are taken as the characteristics of human motion postures, and these features are extracted by wavelet packet decomposition and sample entropy method, which can be used as the basic information of human motion posture recognition [11, 12].

Suppose the power signal of hand muscle is *sEMG*. Since bioenergy is usually in the low-frequency posture of 10 ~ 250 Hz, the "db5" wavelet basis function is used. The muscle power signal *sEMG* is decomposed by four layers of wavelet packet. At the same time, the signals in the first eight subspaces during the fourth layer decomposition are set as the extraction content of sample entropy, that is, the sample entropy of superficial flexor muscle of hand and brachioradialis muscle are extracted and set as Sove1 and Sove2.

When the number of wavelet decomposition layers increases, the spatial resolution will decline [13, 14]. Therefore, this paper selects the kurtosis k, time-domain characteristic skewness η, and frequency-domain characteristic iEMG integral value in the muscle power signal *sEMG* as the features extracted in this paper. The feature parameters are defined as follows:

$$\eta_i = \frac{N}{[(N-1)(N-2)]} \sum_{i=0}^{N} \left[\frac{x(t_i) - \widehat{M}_u}{\widehat{\sigma}_u} \right]^3, \tag{1}$$

$$k_i = \frac{\left[1/N \sum_{i=0}^{N} \left(u(t_i) - \widehat{M}_u \right)^4 \right]}{\left[\left(1/N \sum_{i=0}^{N} \left(x_i - \widehat{M}_u \right)^2 \right)^2 \right]}, \tag{2}$$

$$i\mathrm{EMG}_i = \int_0^N u(t_i)\mathrm{d}t, \tag{3}$$

where the sequence value of muscle electronic signal is described by $u(t_i)$; the values of i are 1 and 2, respectively, which represent the sEMG signals of superficial flexor muscle and brachioradialis muscle in turn; N represents the discrete EMG sequence; at the same time, the mean and variance of EMG signal sequence are represented by $\widehat{M}_u$ and $\widehat{\sigma}_u$, respectively.

Suppose $\mathrm{Feature}_1$ is the combined characteristics of each hand motion. The vector composition of the combined features is expressed by the following equation:

$$\mathrm{Feature}_1 = \{\mathrm{Sove1}\eta_1 k_1 i\mathrm{EMG}_1, \mathrm{Sove2}\eta_2 k_2 i\mathrm{EMG}_2\}. \tag{4}$$

2.1.2. Leg Motion Feature Extraction in Human Motion Postures. Human motion postures are composed of motion characteristics of various parts. This paper extracts the leg motion features during human motion, mainly analyzes the time-domain features of the leg motion amplitude during

motion, which are expressed as mean, standard deviation, interquartile distance, and so on, and extracts the detailed features in the following ways:

The number of quantities trending in a certain group of data sets [15, 16]. , the mean, can be obtained by adding all the data within the group and dividing it by the number of data in the whole group, and the mean is calculated as shown in the equation (5). The specific calculation is shown as follows:

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n} o_i, \tag{5}$$

where n denotes the number of data in the group and o_i denotes the sum of the data. The standard deviation refers to the arithmetic square root of the square of the deviation between the standard value and its average in all data [17].The dispersion of the data set can be displayed through the standard deviation [18]. Therefore, the standard deviation of the data in the window can represent the intensity of human leg activity at this stage. Set the standard deviation as Std, which can be calculated by

$$Std\sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i - \overline{o})^2}. \tag{6}$$

Skewness can describe the symmetry of sample data distribution, which can analyze the symmetry of time series data. If the skewness is greater than zero, it indicates that the sample data distribution is positive. If the skewness is less than zero, it indicates that the data distribution is negative [19]. Set the skewness as Skewness, which can be calculated by

$$\text{Skewness} = \frac{\sum_{i=1}^{n}(o_i - \overline{o})^3}{(n-1)std^3}. \tag{7}$$

Suppose Feature_2 is the combined characteristics of each leg motion. The vector composition of the combined features is expressed by

$$\text{Feature}_2 = \{\text{Mean}, Std, \text{Skewness}\}. \tag{8}$$

2.2. Fast Fusion Recognition Model for Human Motion Posture. In this paper, based on the extraction of two kinds of features, we propose a fusion recognition method based on multimodal bioinformation, and the whole fusion recognition process is represented in Figure 1.

According to Figure 1, in the process of human motion posture fast recognition, the hand and leg motion data are extracted first, the motion features of the hand and leg are extracted separately [20], then the features are normalized to obtain the human hand and leg motion feature sets, the fusion of human motion posture information is completed by improving the typical correlation analysis method, and the fusion results are used as the input of the minimum distance classifier to output the human motion posture fast recognition results.

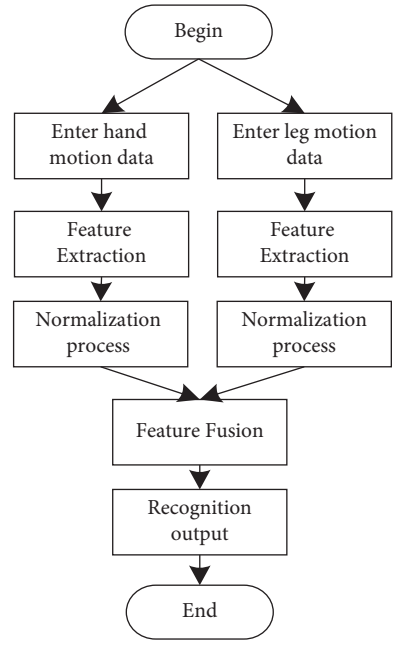


Figure 1: Process of rapid recognition of human motion posture.

2.2.1. Multimodal Biometric Fusion Recognition Based on Improved Typical Correlation Analysis. Canonical correlation analysis method mainly analyzes the feature vectors of samples. In this paper, the method is improved, and the samples are analyzed based on the characteristic matrix of the samples. It is assumed that the characteristic combination matrices of hand and leg motion and are $X = (X \in R^{m_x \times n_x})$ and $Y = (Y \in R^{m_y \times n_y})$. The typical correlation matrix form is expressed through

$$\begin{bmatrix} Z_X' \\ Z_Y' \end{bmatrix}\begin{bmatrix} X \\ Y \end{bmatrix}\begin{bmatrix} R_X \\ R_Y \end{bmatrix}. \tag{9}$$

This paper improves the canonical correlation analysis matrix, so that the category information of training samples can be further used and the recognition results can be more accurate.

Suppose that the left transformation matrices of X and Y are Z_X and Z_Y, and the right transformation matrices are R_X and R_Y. If the first pair of typical related variables of X and Y are $a\prime X$ and $b\prime Y$, after improvement, the maximum correlation between $a\prime X$ and $b\prime Y$ is adjusted to the maximum correlation between matrix $Z_X'\text{XR}_X$ and $Z_Y'\text{YR}_Y$. The improved objective function of the method can be obtained by introducing matrices $Z_X'\text{XR}_X$ and $Z_Y'\text{YR}_Y$ into the criterion function in the typical correlation analysis method:

$$J(Z_X', R_X, Z_Y', R_Y) = \frac{\text{cov}(Z_X'\text{XR}_X, Z_Y'\text{YR}_Y)}{\sqrt{\text{var}(Z_X'\text{XR}_X)\text{var}(Z_Y'\text{YR}_Y)}}. \tag{10}$$

Replace X and Y in (9) with S_{WX} and S_{WY}. After replacement, the improved canonical correlation analysis method (MDCCA) criterion function can be obtained according to the criterion function in the canonical correlation analysis method, as shown in

$$J_{\mathrm{MDCCA}}\left(Z_X', R_X, Z_Y', R_Y\right) = \frac{\mathrm{cov}\left(Z_X' S_{\mathrm{WX}} R_X, Z_Y' S_{\mathrm{WY}} R_Y\right)}{\sqrt{\mathrm{var}\left(Z_X' \mathrm{XR}_X\right)\mathrm{var}\left(Z_Y' \mathrm{YR}_Y\right)}}. \quad (11)$$

Due to the theory of traditional canonical correlation analysis method, certain constraints need to be met when solving (11), as shown in:

$$\mathrm{var}\left(Z_X' X R_X\right) = \mathrm{var}\left(Z_Y' Y R_Y\right) = 1 \quad (12)$$

Therefore, the improved criterion function in this paper can be expressed by

$$J_{\mathrm{MDCCA}}\left(Z_X', R_X, Z_Y', R_Y\right) = \mathrm{cov}\left(Z_X' S_{\mathrm{WX}} R_X, Z_Y' S_{\mathrm{WY}} R_Y\right). \quad (13)$$

At this moment, the solutions to the projection vector a_i and b_i are transformed into the solutions to the projection matrix Z_X, Z_Y, R_X, and R_Y.

2.2.2. Solution to Improved Canonical Correlation Analysis. When improving the traditional canonical correlation analysis method, this paper first adds the intraclass divergence matrix to the original method and then looks for the left-right transformation matrix Z_X, Z_Y, R_X, and R_Y, which can effectively improve the correlation of canonical correlation matrices $Z_X' S_{\mathrm{WX}} R_X$ and $Z_Y' S_{\mathrm{WY}} R_Y$. At the same time, the left-right transformation matrix also represents the projection matrix. According to the constraints of (12) and the criterion function of (13), at the same time, it is assumed that the $p * $ p-order covariance matrix of X is $\sum_{11}$, the $q * q$ covariance matrix of Y is $\sum_{12}$, $p < q$, and the covariance matrix between X and Y is $\sum_{12} = \sum_{21}'$, and then

$$\begin{aligned}
\sum_{11}^{R} &= \frac{\sum_{i=1}^{N}\left(S_{\mathrm{WX}_i} - \overline{S}_{\mathrm{WX}}\right) R_X R_X'\left(S_{\mathrm{WX}_i} - \overline{S}_{\mathrm{WX}}\right)'}{N}, \\
\sum_{22}^{R} &= \frac{\sum_{i=1}^{N}\left(S_{\mathrm{WY}_i} - \overline{S}_{\mathrm{WY}}\right) R_Y R_Y'\left(S_{\mathrm{WY}_i} - \overline{S}_{\mathrm{WY}}\right)'}{N}, \\
\sum_{12}^{R} &= \frac{\sum_{i=1}^{N}\left(S_{\mathrm{WX}_i} - \overline{S}_{\mathrm{WX}}\right) R_X R_Y'\left(S_{\mathrm{WY}_i} - \overline{S}_{\mathrm{WY}}\right)'}{N}.
\end{aligned} \quad (14)$$

Thus, it can be seen that $\mathrm{cov}\left(Z_X' S_{\mathrm{WX}} R_X, Z_Y' S_{\mathrm{WY}} R_Y\right) = Z_X' \sum_{12}^{R} Z_Y$. When the criterion function of (13) is in a state that can satisfy $Z_X' \sum_{11}^{R} Z_X = 1$ and $Z_Y' \sum_{22}^{R} Z_Y = 1$, it can be equivalent to

$$\begin{aligned}
&J_{\mathrm{MDCCA}}\left(Z_X', R_X, Z_Y', R_Y\right) \\
&= \max J\left(Z_X' \sum_{12}^{R} Z_Y\right).
\end{aligned} \quad (15)$$

From the above calculation process, the following results can be obtained:

$$\sum_{11}^{Z} = \frac{\sum_{i=1}^{N}\left(S_{\mathrm{WX}_i} - \overline{S}_{\mathrm{WX}}\right)' Z_X Z_X'\left(S_{\mathrm{WX}_i} - \overline{S}_{\mathrm{WX}}\right)}{N}, \quad (16)$$

$$\sum_{22}^{Z} = \frac{\sum_{i=1}^{N}\left(S_{\mathrm{WY}_i} - \overline{S}_{\mathrm{WY}}\right)' Z_Y Z_Y'\left(S_{\mathrm{WY}_i} - \overline{S}_{\mathrm{WY}}\right)}{N}, \quad (17)$$

$$\sum_{12}^{Z} = \frac{\sum_{i=1}^{N}\left(S_{\mathrm{WX}_i} - \overline{S}_{\mathrm{WX}}\right)' Z_X Z_Y'\left(S_{\mathrm{WY}_i} - \overline{S}_{\mathrm{WY}}\right)}{N}. \quad (18)$$

Meanwhile, $\sum_{21}^{Z} \sum_{21}^{Z}{}'$. When $R_X' \sum_{11}^{Z} R_X = 1$ and $R_Y' \sum_{22}^{Z} R_Y = 1$, (13) is equivalent to

$$\begin{aligned}
&J_{\mathrm{MDCCA}}\left(Z_X', R_X, Z_Y', R_Y\right) \\
&= \max J\left(R_X' \sum_{12}^{Z} R_Y\right).
\end{aligned} \quad (19)$$

Introduce the set R_X and R_Y into (15) to get the solutions to Z_X and Z_Y. Then, use the solutions to get Z_X and Z_Y and introduce them into (19) to calculate R_X and R_Y. Iteration is realized after continuous cyclic calculation.

When R_X and R_Y are given, the optimization of the criterion function of (16) is changed to the generalized eigenvalue solution, as shown in

$$\begin{bmatrix} 0 & \sum_{12}^{R} \\ \sum_{21}^{R} & 0 \end{bmatrix} \begin{bmatrix} Z_X \\ Z_Y \end{bmatrix} = \lambda \begin{bmatrix} \sum_{11}^{R} & 0 \\ 0 & \sum_{22}^{R} \end{bmatrix} \begin{bmatrix} Z_X \\ Z_Y \end{bmatrix}. \quad (20)$$

When R_X and R_Y are given, the optimization of the criterion function of (16) is changed to the generalized eigenvalue solution, as shown in

$$\begin{bmatrix} 0 & \sum_{12}^{Z} \\ \sum_{21}^{Z} & 0 \end{bmatrix} \begin{bmatrix} R_X \\ R_Y \end{bmatrix} = \lambda \begin{bmatrix} \sum_{11}^{Z} & 0 \\ 0 & \sum_{22}^{Z} \end{bmatrix} \begin{bmatrix} Z_X \\ Z_Y \end{bmatrix}. \quad (21)$$

Thus, the solution process is repeated until the convergence effect is achieved, that is, the projection matrix Z_X, Z_Y, R_X, and R_Y can be obtained. Obtain the maximum value of (22) by using Lagrange multiplier method:

$$G_Z = Z_X' \sum_{12}^{R} Z_Y - \frac{\lambda_{Z_X}\left(Z_X' \sum_{11}^{R} Z_X - 1\right)}{2} - \frac{\lambda_{Z_Y}\left(Z_Y' \sum_{22}^{R} Z_Y - 1\right)}{2}. \quad (22)$$

Lagrange multiplier factors are represented by λ_{Z_X} and λ_{Z_Y}, respectively. Use G_Z to get the partial derivative Z_X and Z_Y to make the final value as zero, as shown in

$$\begin{cases} \dfrac{\partial G_Z}{\partial L_X} = \sum_{12}^{R} Z_Y - \lambda_{Z_X} \sum_{11}^{R} Z_X = 0, \\ \dfrac{\partial G_Z}{\partial L_Y} = \sum_{21}^{R} Z_X - \lambda_{Z_Y} \sum_{22}^{R} Z_Y = 0. \end{cases} \tag{23}$$

Z_X' and Z_Y' are multiplied by the left multiplication in (23) in turn to obtain $Z_X' \sum_{12}^{R} Z_Y = \lambda_{Z_X} Z_X' \sum_{11}^{R} Z_X = \lambda_{Z_X}$, $Z_Y' \sum_{21}^{R} Z_X = \lambda_{Z_Y} Z_Y' \sum_{22}^{R} Z_Y = \lambda_{Z_Y}$.The above calculation shows that $\lambda = \lambda_{Z_X} = \lambda_{Z_Y}$.

Therefore, the solution of (19) can be realized by using (23).

If matrix $\sum_{22}^{R}$ is a convertible matrix, $Z_Y = \lambda^{-1} (\sum_{22}^{R})^{-1} \sum_{21}^{R} Z_X$ and $\sum_{12}^{R} (\sum_{22}^{R})^{-1} \sum_{21}^{r} Z_X = \lambda^2 \sum_{11}^{R} Z_X$, i.e., through the calculation of Z_X and Z_Y, the generalized eigenvector of (20) can be obtained. If the given contents are Z_X and Z_Y, through the calculation of (21), R_X and R_Y can be obtained.

Z_X, Z_Y, R_X, and R_Y are iterated continuously through the above process until the result converges. Then, the projection matrix is calculated, and the matrix X and y are projected into the projection matrix in turn to form a typical correlation matrix.

2.2.3. Fast Fusion Recognition Algorithm Process of Human Motion Postures. Fast recognition of human motion postures based on multimodal biological information fusion is realized through the following steps:

(1) Two groups of normalized features can be obtained by using the algorithm.

(2) The left and right transformation matrices Z_X, Z_Y, R_X, and R_Y are given according to the MDCCA theory. For the training samples X and Y, the intraclass divergence matrix S_{WX} and S_{WY} is given in turn, and the objective function is set.

(3) Equations (20) and (21) are iteratively calculated through the criterion function until the result converges, and the value determined Z_X, Z_Y, R_X, and R_Y in turn.

(4) Map samples X and Y to the projection matrices Z_X, Z_Y, R_X, and R_Y to obtain the typical correlation matrix $Z_X'\mathrm{XR}_X$ and $Z_Y'\mathrm{YR}_Y$ after projection.

(5) Taking the matrix $Z_X'\mathrm{XR}_X$ and $Z_Y'\mathrm{YR}_Y$ and the data X as the input of the minimum distance classifier, the minimum distance classifier is used to realize the rapid recognition of human motion postures. In the minimum distance classifier, each category w_j is represented by the mean vector ϑ_j, that is, the vector is described by the mean vector of the feature vector family of each sample, that is: $\vartheta_j = 1/\tau_j \sum_{x \in w_j} x$, in which τ_j refers to the number of training vectors, through which the summation is realized. At the same time, the category of unknown sample vector x is determined through the vector two norm, that is, the distance between sample vectors is analyzed, and the recognition process of human motion postures is realized through the measurement of distance.

2.3. Dataset and Experimental Index. Two data sets were selected for experimental analysis, Human3.6 M dataset and CMU Panoptic data set. Human3.6 M data set includes 3.6 million 3D human postures and corresponding images, a total of 11 experimenters, and a total of 17 action scenes. To use this data set for posture recognition can improve the simulation accuracy. The CMU Panoptic data set is produced by CMU University and consists of 480 VGA cameras, 30+ HD cameras, and 10 Kinnect sensors, which are installed in the lab or in the subject's body to collect the subject's motion data. The overall number of experimental samples collected during the experiment was 4000 image samples, of which 3000 images were used as training samples and the others were used as 1000 images as test samples.

The proposed algorithm is compared with the multimodal fusion driver stress detection algorithm based on attention CNN LSTM in literature [6], the fall detection algorithm based on the spatiotemporal evolution of human posture in literature [7], the two-stage fall recognition algorithm based on human posture features in literature [8], the online adaptive prediction algorithm of human motion intention based on surface EMG signal in literature [9], and the stochastic biomechanical modeling and recognition algorithm of wearable devices for human motion in literature [10] and analyze the recognition performance of different algorithms.

Feature extraction: the general case hand motion feature amplitude is between [-1.8,1.5] and leg motion feature value amplitude is between [-3,3], if the feature value extracted in proposed algorithm is closer to the interval, the higher the feature extraction accuracy.

Human motion posture skewness: the equation for calculating this index is as follows:

$$K = k_1 - k_2. \tag{24}$$

Posture recognition error: this value is the difference between the actual human motion posture and the posture recognized by the proposed algorithm, which is calculated as follows:

$$a = b - c. \tag{25}$$

ROC curve: receiver operating characteristic curve is the line of points drawn under a specific stimulus condition with the probability of false alarm P(y/N) obtained by the subject under different judgment criteria as the horizontal coordinate and the probability of hit P(y/SN) as the vertical coordinate.

Recognition rate: this index refers to the ratio of the amount of data of the correctly recognized human motion posture to the total amount of experimental data, and the equation for calculating this index is as follows:

$$Z = \frac{z_i}{z_j} \times 100\%. \tag{26}$$

3. Results and Discussion

The hand and leg features during human movement in the two data sets are extracted, respectively, and the extracted feature amplitude is counted, asshown in Figure 2.

According to Figure 2, there is a significant difference in the amplitude of leg and hand motion features at different times, in which the leg motion features show large up and down fluctuations, while the hand features are in slow fluctuations. Therefore, there is a large gap between them, and there are disadvantages in using a single feature for human motion posture recognition, and it is necessary to obtain more accurate human motion biometric features through reasonable bioinformation fusion, the proposed algorithm uses the use of wavelet packet decomposition and sample entropy to extract the kurtosis of muscle power signal, time domain feature skewness, frequency domain feature myoelectric integral value and other hand features of human motion posture, and extract the time domain features of leg motion amplitude such as mean, standard deviation, and quartile distance. After normalizing the two features to obtain the set of human hand and leg motion features, a solid foundation can be laid for the subsequent rapid recognition of human motion posture.

The change in skewness of human motion posture recognition using hand, leg, and post-fusion motion features at different times is analyzed, and the results are shown in Figure 3.

According to the data in Figure 3, the bias of human motion posture recognition using leg motion and hand motion features is relatively large at different times, in which the bias of human motion posture recognition using leg features is larger than that of hand motion in walking, sitting, standing, squatting, falling, and lying down motion postures, while the bias of recognition using hand motion features is relatively small, indicating that the magnitude of hand motion is relatively low. The proposed algorithm uses feature sets to construct a fast recognition model of human motion posture based on multimodal bioinformation fusion and inputs the feature sets into the recognition model, which completes the fusion of human motion posture information by improving the typical correlation analysis method, improves the fusion accuracy, and can lay the foundation for subsequent accurate and fast recognition of human motion posture.

Comparing the posture errors of different algorithms in performing human motion posture recognition, the analysis results are shown in Figure 4.

In Figure 4, when the number of frames keeps increasing, the error of different algorithms in posture recognition decreases, in which, the posture error of the algorithm of literature [6] decreases the most, but the final error of this algorithm reaches about −0.82, and the recognition effect is still poor, the posture recognition error of the algorithm of literature [7] is between −0.52 and 1.15, the

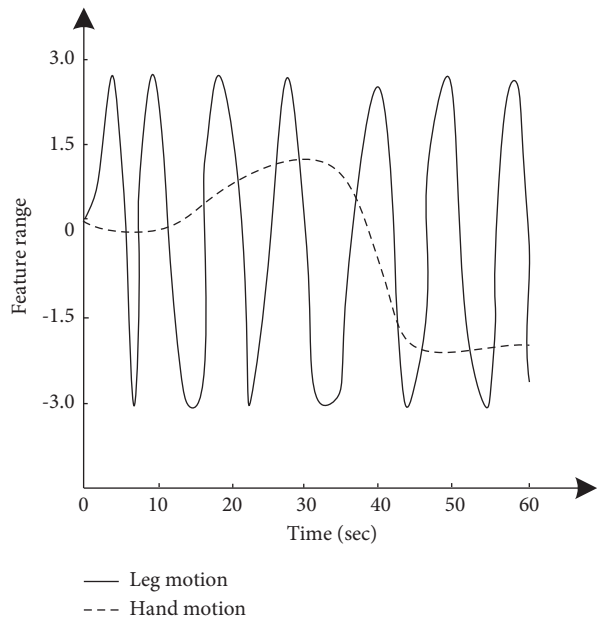


Figure 2: Feature extraction situation analysis.

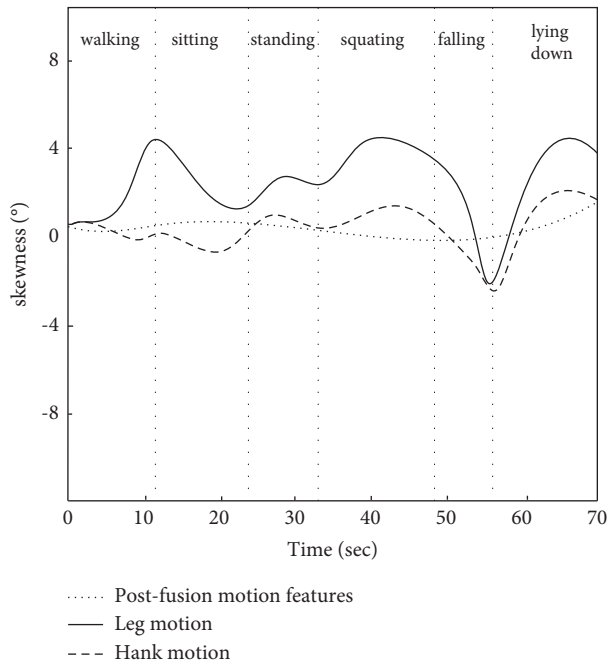


Figure 3: Results of the change of bias of human motion posture by different features recognition.

posture recognition error of the algorithm of literature [8] is between −0.09 and 1.27. The posture recognition error of the algorithm in literature [9] is between −0.31 and 1.25, the posture recognition error of the algorithm in literature [10] is between 0.86 and 1.36, which means that the recognition error of these algorithms is higher than that of the proposed algorithm at the initial stage, and the recognition error of proposed algorithm maintains a decreasing trend at the same time, but the posture recognition error fluctuates slightly between −0.21 and 0.02. It can be seen that the

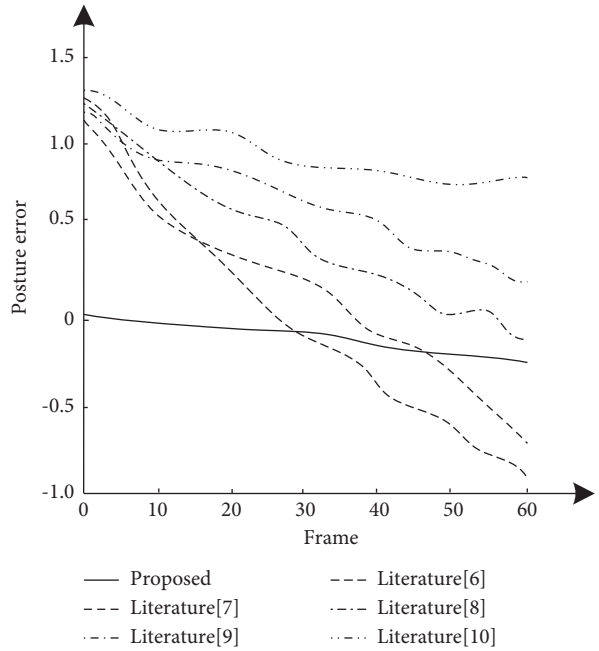


Figure 4: Comparison and analysis of posture recognition error.

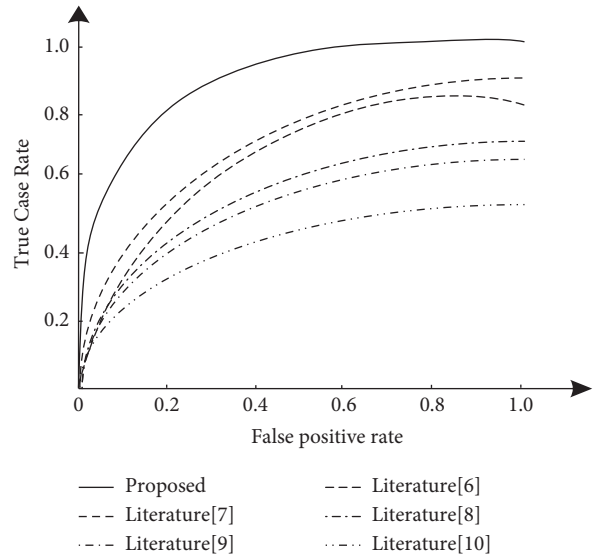


Figure 5: Comparison of ROC curves for different algorithms.

proposed algorithm can effectively reduce the error in posture recognition.

ROC curves were used to verify the recognition effects of different algorithms, and the analysis results are shown in Figure 5.

According to Figure 5, there are some differences in the ROC curve changes of different algorithms, in which the ROC curve of the algorithm of literature [10] keeps the lowest, and the ROC curves of other algorithms are higher than the algorithm of literature [10]. It means that the algorithms of literature [6], literature [7], literature [8], literature [9], and literature [10] have lower recognition real case rate, and the ROC curve of the method calculated in this paper always keeps the highest, and when the false positive rate keeps increasing, the real case rate of the proposed algorithm also keeps at a higher level. The proposed algorithm has a high recognition effect. The method uses the human hand and leg motion feature set to construct a human motion posture fast recognition model using multimodal bioinformation fusion, and inputs the feature set into the recognition model, which completes the fusion of human motion pose information by improving the typical correlation analysis method, and uses the fusion result as the input of the minimum distance classifier, so as to achieve human motion posture fast recognition. Therefore, the proposed algorithm has a better recognition effect.

Table 1: Analysis of recognition rate of different algorithms (%).

Motion posture	Proposed	Literature [6]	Literature [6]	Literature [8]	Literature [9]	Literature [10]
Walking	97.4	91.4	87.4	89.5	78.5	83.4
Sitting	96.5	92.1	86.5	89.3	76.4	84.6
Falling	97.5	90.3	85.7	84.3	75.6	85.3
Squatting	94.8	91.4	89.6	81.4	74.3	88.3
Standing	95.6	90.7	88.6	89.5	73.3	81.3
Lying down	96.5	89.5	89.6	88.5	71.5	83.4

The results of the analysis of the variation of the recognition rate of different algorithms in the case of different postures of human motion are shown in Table 1.

It is seen form Table 1 that the recognition rate of the proposed algorithm is always the highest in posture recognition, the recognition rate of the proposed algorithm is always above 95%, the human motion posture recognition rate of the algorithm in literature [6] fluctuates in the range of 89.5% to 92.1%, the human motion pose recognition rate of the algorithm in literature [7] fluctuates in the range of 85.7% to 89.6%, the human motion posture recognition rate of the algorithm in literature [8] fluctuates in the range of 81.4% to 89.5%, the human motion posture recognition rate of the algorithm in literature [9] fluctuated in the range of 71.5% to 78.5%, and the human motion posture recognition rate of the algorithm in literature [10] fluctuated in the range of 81.3% to 88.3%. The recognition rate of the other algorithms is also always lower than that of the proposed algorithm. Therefore, the proposed algorithm has the best results for human motion posture recognition.

4. Conclusions

Due to single-modal bioinformation is one sided, the application of such information to human motion gesture recognition process leads to increased recognition bias, so we design a fast human motion posture recognition algorithm using multimodal bioinformation fusion, extracting biofeatures of multiple human motion modalities and fusing these features for posture recognition. The posture recognition error is −0.21~0.02, and the recognition rate is always above 95%. The feasibility of the proposed algorithm is verified using experiments, and it is confirmed that the proposed algorithm achieves better application results. However, the recognition process of proposed algorithm is too complicated, which may lead to the degradation of recognition efficiency. Therefore, in future work, the optimization can be continued for the current algorithm, so that the algorithm can recognize more human motion posture changes faster.

Data Availability

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

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 Grant no. 61572343, the 13th Five-Year Plan Project of Shanxi Education Science under Grant no. GH-18044, and the Application Basic Research Project of Datong Science and Technology Bureau under Grant no. 2021172.

References

[1] H. Song, A. Li, T. Wang, and M. Wang, "Multimodal deep reinforcement learning with auxiliary task for obstacle avoidance of indoor mobile robot," *Sensors*, vol. 21, no. 4, pp. 1363–1372, 2021.

[2] M. S. S. Syed, E. Pirogova, and M. Lech, "Prediction of public trust in politicians using a multimodal fusion approach," *Electronics*, vol. 10, no. 11, pp. 1259–1270, 2021.

[3] J. Zhao, L. Yu, and Z. Liu, "Research based on multimodal deep feature fusion for the auxiliary diagnosis model of infectious respiratory diseases," *Scientific Programming*, vol. 2021, Article ID 5576978, 6 pages, 2021.

[4] Q. Qi, L. Lin, and R. Zhang, "Feature extraction network with attention mechanism for data enhancement and recombination fusion for multimodal sentiment analysis," *Information (Switzerland)*, vol. 12, no. 9, pp. 342–351, 2021.

[5] C. Zhang, "Convolution analysis operator for multimodal image fusion," *Procedia Computer Science*, vol. 183, no. 5, pp. 603–608, 2021.

[6] L. Mou, C. Zhou, P. Zhao et al., "Driver stress detection via multimodal fusion using attention-based cnn-lstm," *Expert Systems with Applications*, vol. 173, no. 12, Article ID 114708, 2021.

[7] J. Zhang, C. Wu, and Y. Wang, "Human fall detection based on body posture spatio-temporal evolution," *Sensors*, vol. 20, no. 3, pp. 946–957, 2020.

[8] K. Han, Q. Yang, and Z. Huang, "A two-stage fall recognition algorithm based on human posture features," *Sensors*, vol. 20, no. 23, pp. 6966–6975, 2020.

[9] Z. Ding, C. Yang, Z. Wang, X. Yin, and F. Jiang, "Online adaptive prediction of human motion intention based on semg," *Sensors*, vol. 21, no. 8, pp. 2882–2889, 2021.

[10] B. E. Olivas-Padilla, S. Manitsaris, D. Menychtas, and A. Glushkova, "Stochastic-biomechanic modeling and recognition of human movement primitives, in industry, using wearables," *Sensors*, vol. 21, no. 7, pp. 2497–2510, 2021.

[11] S. Mu, M. Cui, and X. Huang, "Multimodal data fusion in learning analytics: a systematic review," *Sensors*, vol. 20, no. 23, pp. 6856–6866, 2020.

[12] J. Ding and G. Zhao, "Fast recognition method of text features in dynamic video images," *International Journal of Reasoning-Based Intelligent Systems*, vol. 12, no. 4, pp. 248–259, 2020.

[13] A. P. M. Diniz, K. F. Côco, F. S. V. Gomes, and J. L. F. Salles, "Forecasting model of silicon content in molten iron using wavelet decomposition and artificial neural networks," *Metals-Open Access Metallurgy Journal*, vol. 11, no. 7, pp. 1001–1012, 2021.

[14] Y. Yu, W. Zhao, S. Li, and S. Huang, "A two-stage wavelet decomposition method for instantaneous power quality indices estimation considering interharmonics and transient disturbances," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 21, pp. 1–13, 2021.

[15] J. Brünger, M. Gentz, I. Traulsen, and R. Koch, "Panoptic segmentation of individual pigs for posture recognition," *Sensors (Basel, Switzerland)*, vol. 20, no. 13, pp. 3710–3722, 2020.

[16] S. Xu, H. Lu, C. Ference, and Q. Zhang, "An accuracy improvement method based on multi-source information fusion and deep learning for tssc and water content nondestructive detection in "luogang" orange," *Electronics*, vol. 10, no. 1, pp. 80–92, 2021.

[17] X. Zhu, T. Shi, X. Jin, and Z. Du, "Multi-sensor information fusion based control for vav systems using thermal comfort constraints," *Building Simulation*, vol. 14, no. 4, pp. 1047–1062, 2021.

[18] Y. Li, H. Jia, J. Qi et al., "An acquisition method of agricultural equipment roll angle based on multi-source information fusion," *Sensors*, vol. 20, no. 7, pp. 2082–2094, 2020.

[19] Y. Wu, X. Li, and Z. Cao, "Effective doa estimation under low signal-to-noise ratio based on multi-source information meta fusion," *Journal of Beijing Institute of Technology (Social Sciences Edition)*, vol. 30, no. 4, pp. 377–396, 2021.

[20] Y. Feng, J. Hu, R. Duan, and Z. Chen, "Credibility assessment method of sensor data based on multi-source heterogeneous information fusion," *Sensors*, vol. 21, no. 7, pp. 2542–2555, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 8544591, 9 pages
https://doi.org/10.1155/2022/8544591

Research Article

Multimodal Biometrics Fusion Algorithm Using Deep Reinforcement Learning

Quan Huang

Institute of Information, Hunan University of Humanities, Science and Technology, Loudi 417000, China

Correspondence should be addressed to Quan Huang; hnldhq@126.com

Received 4 February 2022; Revised 1 March 2022; Accepted 8 March 2022; Published 24 March 2022

Academic Editor: Xiaofeng Li

Multimodal biometrics fusion plays an important role in the field of biometrics. Therefore, this paper presents a multimodal biometrics fusion algorithm using deep reinforcement learning. In order to reduce the influence of user behavior, user's personal characteristics, and environmental light on image data quality, data preprocessing is realized through data transformation and single-mode biometric image region segmentation. A two-dimensional Gobar filter was used to analyze the texture of local sub-blocks, qualitatively describe the similarity between the filter and the sub-blocks and extract the phase information and local amplitude information of multimodal biometrics features. Deep reinforcement learning was used to construct the classifier of different modal biometrics, and the weighted sum fusion of different modal biometrics was implemented by fractional information. The multimodal biometrics fusion algorithm was designed. The Casia-iris-interval-v4 and NFBS datasets were used to test the performance of the proposed algorithm. The results show that the fused image quality is better, the feature extraction accuracy is between 84% and 93%, the average accuracy of feature classification is 97%, the multimodal biometric classification time is only 110 ms, the multimodal biometric fusion time is only 550 ms, the effect is good, and the practicability is strong.

1. Introduction

With the development of biometric extraction and its related technologies, biometric extraction products are becoming more and more mature and are gradually entering every aspect of society, which has brought great changes in people's living habits. Among them, fingerprint identification attendance meter, face recognition access control, iris customs clearance system, and other products have played a great role in security and brought great convenience to people's lives. However, with the increasing demand for security, there is an urgent need for a better biometric recognition system that must meet various requirements, including uniqueness, universality, stability, collectability, ease of use, and security. Universality means that the attribute of use needs to be the attribute of normal and healthy human beings; uniqueness means that the biometric information used is unique to the individual; stability means that the biometric mode needs to be stable and constant for a long time; collectability means that biometric modal information can be collected by external equipment; and ease of use means that the system needs to meet the needs of users for convenience [1]. Security means that the mode needs to be not easy to leave so as to avoid theft and forgery by others. The above requirements can be met through multimodal biometric fusion. Research on this problem plays a positive role in the improvement of research results in the field of biometric extraction and has high research value. Therefore, the multimodal biometric fusion algorithm is studied [2].

With more and more attention paid to security in various fields, research on multimodal biomedical feature fusion algorithms has gradually attracted much attention. Dinh Phu-Hung [3] proposed a multimodal medical image fusion algorithm based on the marine predator algorithm and three-scale image decomposition. The algorithm uses the Kirsch compass operator based on the local energy function for detailed level fusion so that the output image retains important information. Moreover, the algorithm uses the marine predator algorithm (MPA) to fuse the

bottom layer by optimizing the parameters so that the output image has good quality, but the algorithm has the problem of low accuracy in feature extraction. Xinhua Li and Jing Zhao [4] proposed a new algorithm for multimodal medical image fusion. Firstly, CT images and MRI images are decomposed into low and high-frequency sub-bands by NSCT with multiscale geometric transformation; secondly, for the low-frequency sub-band, the local standard deviation algorithm or fusion is selected. For the high-frequency sub-band, the adaptive pulse-coupled neural network model is constructed, and the fusion rules are set according to the cumulative ignition times of iterative operations in the network; finally, the fused image is obtained by image reconstruction. However, the algorithm has the problem of a low error rate. M. Zhu and X. Yu [5] proposed a multifeature fusion algorithm in detail enhancement of VR panoramic images. The shadow detection results are obtained by using HSV color features and texture features, and then the final detection results are obtained by fusion. Experimental results show that the algorithm greatly reduces the false detection rate. However, the algorithm has the problem of low average classification accuracy. Geng Peng and Xiuming Sun [6] proposed a multimodal medical image fusion algorithm based on the quaternion wavelet transform. The algorithm can fuse not only CT and MR image pairs, but also CT and proton density-weighted MR image pairs, as well as multispectral MR images such as T1 and T2. However, the algorithm has the problem of a high average authentication time for a single user. Castro et al. [7] proposed a multimodal feature fusion algorithm for gait recognition based on CNN. Based on the original pixels and their derived simple functions, the algorithm uses advanced learning technology to extract relevant features and fuse the original pixel information with the information from optical flow and depth map to realize multimodal feature fusion. However, the algorithm has the problem of a low error rate. Therefore, a new multimodal biometrics fusion algorithm based on deep reinforcement learning is proposed. The main contributions to this paper are as follows: (1) data preprocessing is realized through data transformation, single-mode biometric image region segmentation, and other steps to reduce the impact of user behavior, user personal characteristics, and ambient light on image data quality. (2) The two-dimensional Gobar filter is used to analyze the local sub-blocks of texture, qualitatively describe the similarity between the filter and the sub-blocks, extract the phase information and local amplitude information of multimodal biometrics, and the accuracy and efficiency of feature extraction are proposed. (3) Deep reinforcement learning is used to construct different modal biometric classifiers, and the weighted sum fusion of different modal biometrics is implemented through the score information to realize the design of a multimodal biometric fusion algorithm so as to achieve the multiple objectives of improving the fusion efficiency and quality.

2. Materials and Methods

2.1. Dataset Description. In the process of testing the designed multimodal biomedical feature fusion algorithm based on deep reinforcement learning, the experimental datasets used are the CASIA-Iris-Interval-v4 and NFBS datasets, respectively.

The CASIA-Iris-Interval-v4 includes binocular iris images of 100 users. The image size is $640 * 480$, and the acquisition times on the left and right sides of each person are 7 times, for a total of 1400 images. The NFBS (neurofeedback skull-stripped) dataset, the dataset includes 125 sets of brain image data. Each set of data includes a defaced T1 MR image, a skull removed MR image, brain segmentation results, and its dataset modal T1 weighted MRI scan. In the performance test of the algorithm, the experimental parameters are set as follows: the corresponding input batch size of a convolutional neural network is set to 16. Adam is the optimizer; the initial learning rate is set to 0.01; the probability of random inactivation of neurons is set to 0.8. The two datasets are divided into verification set, test set, and training set, with a division ratio of 1 : 2 : 7.

2.2. Preprocessing of Single-Modal Biometric Image Data. Preprocess the single-modal biometric image data to eliminate the adverse effects of user behavior, user personal characteristics, and ambient light. The specific data preprocessing steps are data transformation and single-modal biometric image region segmentation.

Data transformation is to transform the value range of single-modal biometric image data according to demand so as to better implement the processing of single-modal biometric image data [8]. The processing method of data transformation is the minimum-maximum normalization method, which mainly implements the linear transformation of value range data through the linear transformation method [9] so that the transformed data is in a unified value range. The equation is as follows:

$$\text{new\_value} = \frac{\text{old\_value} - \text{old\_Min} \times (\text{new\_Max} - \text{new\_Min})}{\text{old\_Max} - \text{old\_Min}} + \text{new\_Min}, \quad (1)$$

where new_value refers to the sample data within the normalized single-modal biometric range; old_value refers to the sample data of single-modal biometric range; old_Min refers to the minimum value of sample data in value range; new_Max refers to the maximum value of sample data in value range; new_Min refers to the minimum value of the data classification range within the value range; and old_Max refers to the maximum value of the data classification range within the value range [10].

The single-modal biometric region segmentation mainly detects the edge between the background region and the single-modal biometric region; that is, the single-modal biometric contour is detected, and the single-modal biometric region is segmented in the image background [11].

The human body is affected by infrared, and the brightness of the single-modal biometric region is often lower than that of other background regions. Edge detection is implemented by the Canny operator. The detection steps are as follows:

(1) The smoothing filter is constructed by a one-dimensional Gaussian function $G(x)$. The convolution operation of $f(x, y)$ is performed by column and row, respectively, to obtain the smooth image $I(x, y)$. The equation for constructing the filter is as follows [12]:

$$G(x) = \frac{\exp\left(-x^2/2\sigma^2\right)}{2\pi\sigma^2}, \tag{2}$$

(i) where σ is the standard deviation corresponding to the Gaussian function.

(2) We pass a 2×2 matrix, where the gradient direction $H(x, y)$ and gradient amplitude $M(x, y)$ of $I(x, y)$ are calculated by the neighborhood first-order partial derivative equation [13]. The equation is as follows:

$$\begin{aligned} H(x, y) &= \arctan\left[k_x(x, y), k_y(x, y)\right], \\ M(x, y) &= \sqrt{k_x^2(x, y) + k_y^2(x, y)}, \end{aligned} \tag{3}$$

(i) where k_x refers to the result of $I(x, y)$ convolution operation of the smooth image by line through the filter; k_y refers to the result of $I(x, y)$ convolution operation of the smooth image by column through the filter. The convolution template of k_x is as follows:

$$f_x = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}. \tag{4}$$

(ii) The convolution template of k_y is as follows:

$$f_y = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}. \tag{5}$$

(3) The nonmaximum suppression of gradient amplitude is implemented to obtain all possible edge points in the smooth image $I(x, y)$.

(4) The result of step (3) is segmented by high and low thresholds, respectively, to obtain two threshold edge images. For the image obtained by low threshold segmentation, the edges need to be collected continuously until all the gap connections of the image obtained by high threshold segmentation are realized [14].

Through the above steps, the preprocessing of single-modal biometric image data is completed.

2.3. Multimodal Biometric Extraction. The texture local sub-block analysis of the segmented single-modal biometric region is implemented through the two-dimensional Gobar filter. The similarity between the filter and the local sub-block is qualitatively described [15], and the phase information and local amplitude information of multimodal biometrics are extracted [16]. When the convolution result is positive, the local texture phase coding of the sub-block is 1. On the contrary, the local texture phase coding is 0.

The definition of a two-dimensional Gobar filter is as follows:

$$R(x, y) = e^{\pi\left[\frac{(x-x_0)^2 + (y-y_0)^2}{\lambda \cdot \tau}\right]} e^{\pi^2\left[\frac{(x-x_0)^2 + (y-y_0)^2}{\lambda_0 \cdot \tau_0}\right]}, \tag{6}$$

where x_0 refers to the time domain frequency of two-dimensional Gabor wavelet; y_0 refers to the frequency of two-dimensional Gabor wavelet in the frequency domain; λ refers to the localization frequency of two-dimensional Gabor wavelet; τ_0 refers to the local texture threshold of sub-block; λ_0 refers to the wavelet iterative function; and τ_0 refers to the scale difference function of texture extreme points.

The texture coordinates in the sub-block are transformed into a fixed scale rectangular region to describe each pixel in the single-modal biometric region.

$$R(r, \theta) = e^{-\omega(\theta-\theta')} e^{\left(-(r-r')^2/\lambda^2\right)} e^{\pi\left(-(\theta-\theta')^2/\tau^2\right)}, \tag{7}$$

where r refers to coordinate transformation distance; θ refers to coordinate transformation angle; ω refers to rectangular area's scale; r' refers to the normalized value of coordinate transformation distance; and θ' refers to the normalized value of coordinate transformation angle [17, 18].

Through $R(x, y)$, filter waves of $R(r, \theta)$. The local sub-block is qualitatively described as 1 and 0. The single-modal biometrics are described through the description of the local sub-block, and the corresponding multibit feature codes of various single-modal are created to realize the extraction of multimodal biometrics [19, 20].

2.4. Multimodal Biological Feature Fusion Algorithm. Deep reinforcement learning is used to construct different modal biometric classifiers, and the weighted sum fusion of different modal biometric features is implemented through the score information to realize the design of a multimodal biometric fusion algorithm.

(i) Input: test sample set of multimodal biometrics.

(ii) Output: fusion results of multimodal biometrics.

(1) The deep reinforcement learning algorithm based on convolution neural networks is used to construct different modal biometric classifiers. The classifier is composed of two modules: Inception-ResNet module and residual connection module [21].

The residual connection module can accumulate the output result of the upper layer network and the copy data and directly input it to the lower layer network. The accumulation of the output and the original image feature map helps to reduce the training error. The module network is composed of multiple residual blocks. It is a stacked deepening network that can prevent the occurrence of the overfitting phenomenon. The middle of the network is the residual structure, which can realize jump transmission [22, 23].

The Inception-ResNet module uses the inception structure and can find the optimal sparse structure unit. The single-modal between the two activation functions is expanded to become a new neural network, and multiple convolution kernels of different sizes are used to realize feature classification on different scales. The loss function of the classifier is as follows:

$$l = \frac{1}{2}\sum_{i=1}^{N}\left|t - C_i\right|^2, \tag{8}$$

where l refers to the loss value; N refers to the number of samples; t refers to the feature value; and C_i refers to the feature center point of category corresponding to sample i.

When the value of the loss function is the smallest, the network can reach the most convergent state.

(2) m classifiers are represented by the following equation:

$$M = \{m_1, m_2, \ldots, m_n\}, \tag{9}$$

where m_n represents the n th classifier.

The classification results of m classifiers are as follows:

$$H = \sum_{n=1}^{M} \alpha m_n, \tag{10}$$

where α represents the average classification threshold:

(3) Let us estimate the optimal weight of classification results as follows:

$$\delta_n = \left(\frac{\kappa(\mathrm{Gen}_n) - S}{\kappa(\mathrm{Im}p_n) - S}\right)_2^2, \tag{11}$$

where $\kappa(Gen_n)$ is the mean value of the truth function $\kappa(\mathrm{Im}p_n)$ is the mean of false value scores. S is the mean of the matching score.

(4) Let us quantify the uncertain area in the feature information through the mean closure strategy as follows:

$$\mu = F(z_i)S, \tag{12}$$

where μ refers to the quantized value of uncertain area and $\mathrm{F}(z_i)$ refers to the quantized function of uncertain value z_i.

(5) Let us assign weight to uncertain areas according to δ_n value as follows:

$$s_i = \frac{\delta_n}{\sum_{n=1}^{M}\delta_n}. \tag{13}$$

The assigned weight value satisfies the following equation:

$$\begin{aligned} &0 \le s_i \le 1, \quad \forall M, \\ &\sum_{n=1}^{M} s_i = 1, \quad \forall n. \end{aligned} \tag{14}$$

Through the reasonable distribution of weights, the summation and weighted fusion of the matching score information of different modal features are realized. The process of the proposed algorithm is shown in Figure 1.

2.5. Experimental Index. In the experiment, multimodal medical image fusion algorithms based on marine predator algorithm and three-scale image decomposition, the new algorithm for multimodal medical image fusion, multifeature fusion algorithm in detail enhancement of VR panoramic images, multimodal medical image fusion algorithm based on quaternion wavelet transform, and multimodal feature fusion algorithm for gait recognition based on CNN are used as the comparison test algorithms, which are represented by algorithms 1, 2, 3, 4 and 5, respectively.

The various performances of the proposed algorithm are tested. The experimental indexes are as follows: fusion effect: we take brain CT images and MRI images as examples for multimodal biometric fusion, and compare the fusion results, that is, the quality of the image. Accuracy of multimodal biometric extraction: the calculation equation of multimodal biometric extraction accuracy is as follows:

$$A = \frac{a_i}{a_j} \times 100\%, \tag{15}$$

where a_i refers to the amount of data correctly extracted from multimodal biometric samples, and a_j refers to the total number of samples for multimodal biometrics.

Average classification accuracy (ACA): it represents the average probability of correct classification of multimodal biometrics in multiple experiments, and its calculation equation is as follows:

$$\mathrm{ACA} = \frac{\sum_{i=1}^{n}\left(\left(b_{ik}/b_{ij}\right) \times 100\%\right)}{n}, \tag{16}$$

where a_{ik} refers to sample data size of correctly classified multimodal biometrics, b_{ij} refers to the experimental data, and n refers to the times of experiments.

The time-consuming of multimodal biometric classification: the time-consuming of multimodal biometric classification refers to the time spent in classifying multimodal biometrics. The calculation equation of the index is as follows:

$$T_1 = \left|t_1 - t_2\right|, \tag{17}$$

where t_1 represents the classification start time and t_2 represents the classification end time.

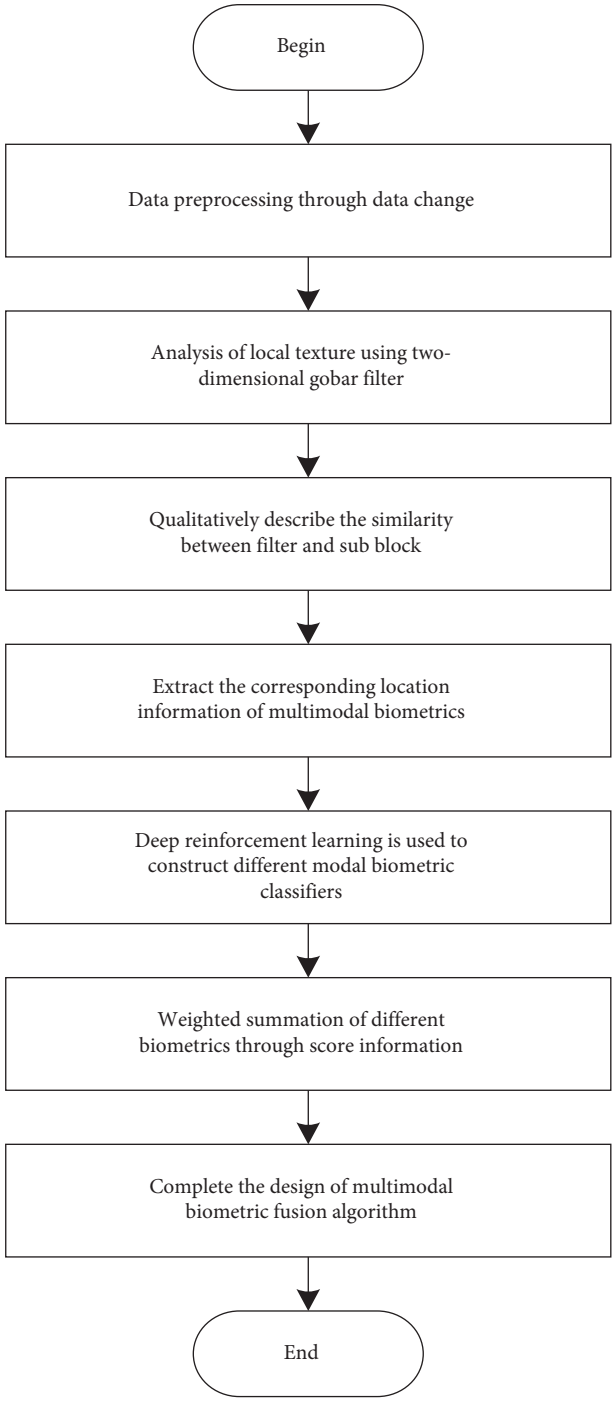


Figure 1: Process of multimodal biometric fusion algorithm.

Multimodal biometric fusion time: it refers to the sum of the time taken to complete all multimodal biometric fusion steps. It is expressed as follows:

$$T_2 = \sum_{i=1}^{n} t_i, \tag{18}$$

where t_i represents the time taken for the i th multimodal biometric fusion step.

3. Results and Discussion

This paper takes brain CT images and MRI images as an example for multimodal biometric fusion, in which the original image is shown in Figure 2.

The fusion results of different algorithms are shown in Figure 3.

The fusion results of brain CT images and MRI images are shown in Figure 3. The fusion images with algorithms 1–5 are of low quality and have poor definition. Therefore, the fusion results cannot accurately reflect the detailed features of the two images. Compared with these algorithms, the proposed algorithm almost retains the texture of the soft tissue area in most brain MRI images and obviously retains the shadow area of a suspected blood clot in brain CT images. The image definition is higher and the fusion effect is better.

We test the feature extraction accuracy of the designed multimodal biometric fusion algorithm based on deep reinforcement learning and five comparison algorithms. The test results are shown in Figure 4.

According to the data in Figure 4, the process of increasing dimension, the feature extraction accuracy of the proposed algorithm also increases. The feature extraction accuracy of algorithm1 is 74%–86%, that of algorithm 2 is 63%–73%, that of algorithm 3 is 52%–64%, that of algorithm 4 is 57%–70%, and that of algorithm 5 is 56%–67%. The feature extraction accuracy of the proposed algorithm is between 84% and 93%, and its feature extraction accuracy is higher than that of the five comparison algorithms, which shows that the proposed algorithm can achieve more accurate multimodal biometric extraction.

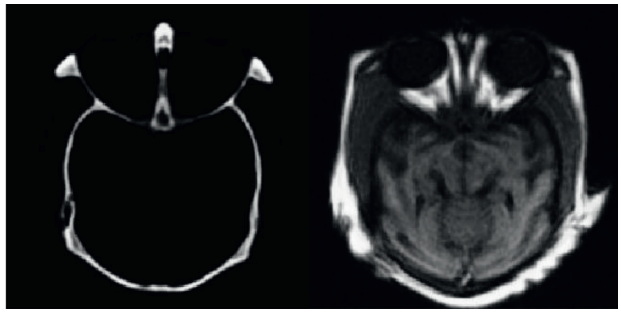
We test the average classification accuracy of the designed algorithm and the five comparison algorithms. The test results are shown in Figure 5.

The average classification accuracy test results in Figure 5 show that the highest average classification accuracy was 82% for algorithm 1, 73% for algorithm 2, 76% for algorithm 3, 72% for algorithm 4, and 75% for algorithm 5. The average classification accuracy of multimodal biometrics of the proposed algorithm is high, up to 97%, which is higher than the five comparison algorithms, indicating that the ACA performance of multimodal biometrics of the proposed algorithm is strong.

The time-consuming test results of multimodal biometric classification are shown in Figure 6.

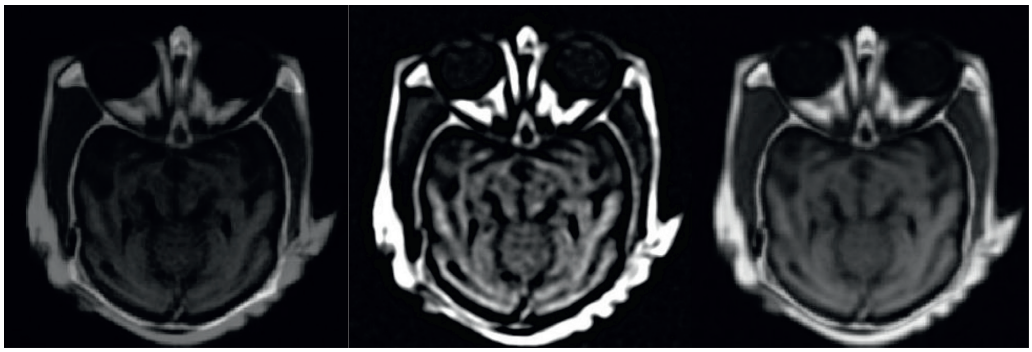
The test results in Figure 6 show that the multimodal biometric classification of algorithm 1 takes 270 s, the multimodal biometric classification of algorithm 2 takes 340 s, the multimodal biometric classification of algorithm 3 takes 250 s, the multimodal biometric classification of algorithm 4 takes 260 s, and the multimodal biometric classification of algorithm 5 takes 460 s. The multimodal biometric classification of the proposed algorithm takes the shortest time, and it only takes 110 ms to realize the multimodal biometric classification with the highest efficiency.

The test results of multimodal biometric fusion time of six algorithms are shown in Figure 7.

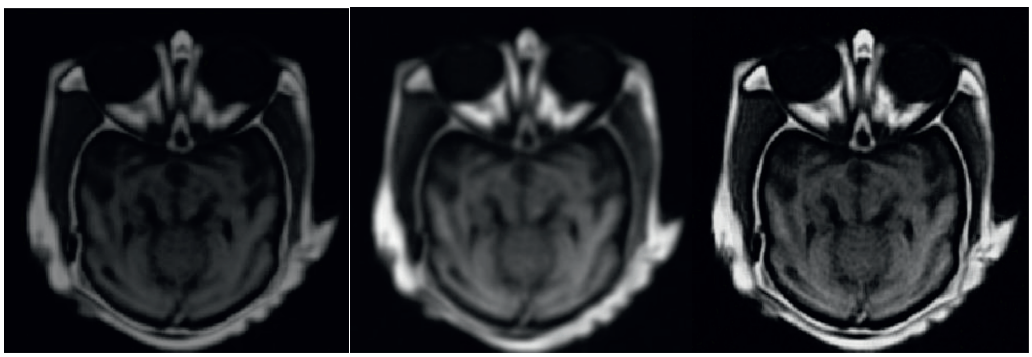


(a) (b)

Figure 2: Original images: (a) CT image and (b) Mr-t2 image.



(a) (b) (c)



(d) (e) (f)

Figure 3: Comparison of fusion effects: (a) algorithm 1, (b) algorithm 2, (c) algorithm 3, (d) algorithm 4, (e) algorithm 5, and (f) the proposed algorithm.

According to the multimodal biometric fusion time test data in Figure 7, the multimodal biometric fusion time of algorithm 1 is 700 ms, the multimodal biometric fusion time of algorithm 2 is 740 ms, the multimodal biometric fusion time of algorithm 3 is 640 ms, the multimodal biometric fusion time of algorithm 4 is 700 ms, and the multimodal biometric fusion time of algorithm 5 is 650 ms. The multimodal biometric fusion time of the proposed algorithm is the shortest, and the multimodal biometric fusion can be realized in only 550 ms, which is more efficient.

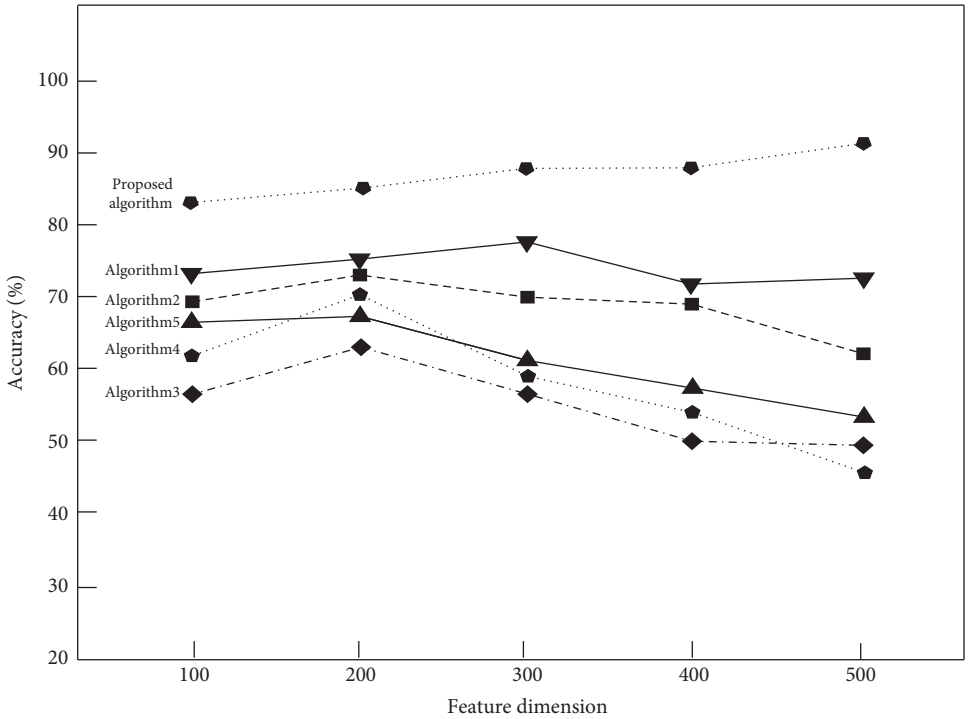


Figure 4: Comparison of feature extraction accuracy results.

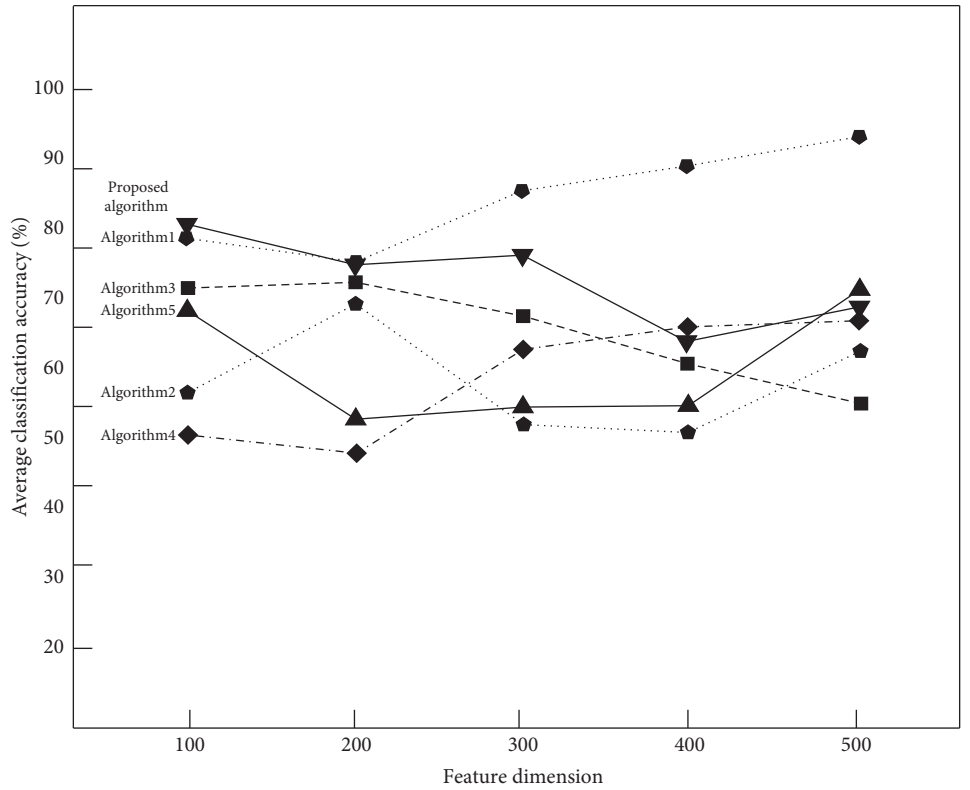


Figure 5: Comparison of average classification accuracy results.

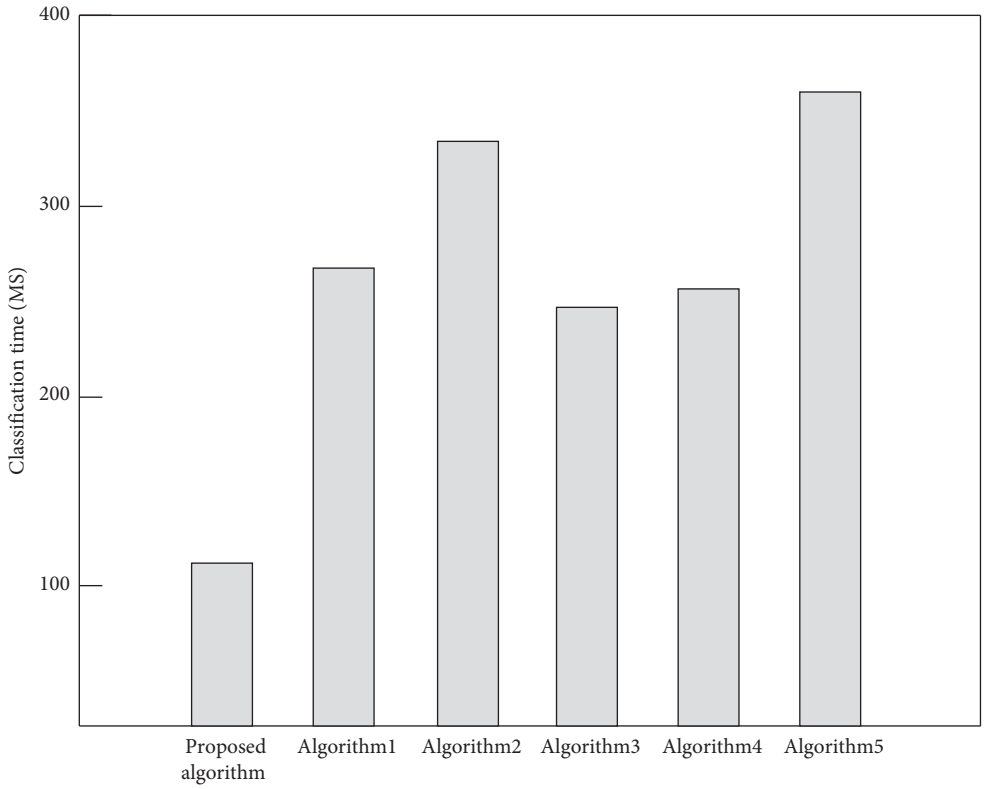


Figure 6: Comparison of classification time-consuming test results.

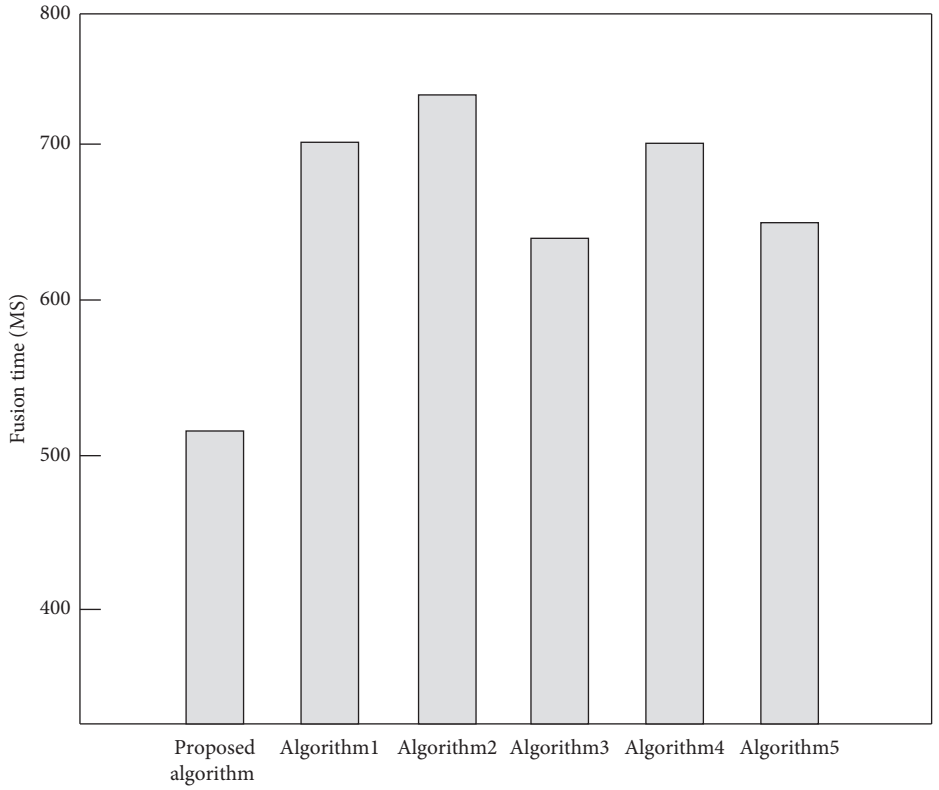


Figure 7: Comparison of fusion time test results.

4. Conclusions

In the research, the problem of multimodal biometric fusion is solved, a multimodal biometric fusion algorithm based on deep reinforcement learning is proposed, and the algorithm is verified in practice. The following research results were obtained: (1) the unification of the value range of single-modal biometric image data and the segmentation of single-modal biometric regions are realized; (2) the extraction of multimodal biometric phase information and local amplitude information is realized by two-dimensional Gobar filter; and (3) different modal biometric classifiers are designed. In the research, due to the limitations of time and energy, the training of classifiers and other issues have not been studied in detail. More in-depth and detailed research will be carried out in the future so as to maximize the classification accuracy and the effect of multimodal biometric fusion.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest with this manuscript.

Acknowledgments

This work was supported by the Natural Science Foundation of Hunan Province under Grant no. 2017JJ2124, the Teaching Reform Project in Hunan Province under Grant no. 2018436, and the Cooperative Education Project in Ministry of Education under Grant no. 201802.

References

[1] G. S. Walia, G. Jain, N. Bansal, and K. Singh, "Adaptive weighted graph approach to generate multimodal cancelable biometric templates," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1945–1958, 2020.

[2] D. Zhao, Z. Chang, and S. Guo, "A Multimodal fusion approach for image captioning," *Neurocomputing*, vol. 329, pp. 476–485, 2019.

[3] P. H Dinh, "A novel approach based on Three-scale image decomposition and Marine predators algorithm for multi-modal medical image fusion," *Biomedical Signal Processing and Control*, vol. 67, no. 2, pp. 1–14, 2021.

[4] X. Li and J. Zhao, "A novel multimodal medical image fusion algorithm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 1–8, 2020.

[5] M. Zhu and X. Yu, "Multi-feature fusion algorithm in VR panoramic image detail enhancement processing," *IEEE Access*, no. 99, p. 1, 2020.

[6] G. Peng, X. Sun, and J. Liu, "Adopting quaternion wavelet transform to fuse multimodal medical images," *Journal of Medical and Biological Engineering*, vol. 37, no. 5, pp. 230–239, 2017.

[7] F. M Castro, M. J Marín-Jiménez, N. Guil, and N. Blanca, "Multimodal feature fusion for CNN-based gait recognition: an empirical comparison," *Neural Computing & Applications*, vol. 32, no. 11, pp. 14173–14193, 2020.

[8] C. Taouche and H. Belhadef, "Multimodal biometric system combining left and right palmprints," *Information Discovery and Delivery*, vol. 48, no. 1, pp. 2–13, 2019.

[9] D. Zhao, "Rapid multimodal image registration based on the local edge histogram," *Mathematical Problems in Engineering*, vol. 2021, no. 9, pp. 1–9, 2021.

[10] Z. Wang, J. Zhen, Y. Li, G. Li, and Q. Han, "Multi-feature multimodal biometric recognition based on quaternion locality preserving projection," *Chinese Journal of Electronics*, vol. 28, no. 4, pp. 789–796, 2019.

[11] N. Tawfik, H. A. Elnemr, M. Fakhr, M. I. Dessouky, and F. E. A. El-Samie, "Hybrid pixel-feature fusion system for Multimodal medical images," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 42, pp. 1–18, 2021.

[12] D. Wang, Y. Li, L. Jia, Y. Song, and Y. Liu, "Novel three-stage feature fusion method of multimodal data for bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 99, pp. 1–10, 2021.

[13] S. Bhagat, S. D. Joshi, B. Lall, and G. Smriti, "Multimodal sensor fusion using symmetric skip autoencoder via an adversarial regulariser," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 99, p. 1, 2020.

[14] L. C. O. Tiong, S. T. Kim, and Y. M. Ro, "Implementation of Multimodal biometric recognition via multi-feature deep learning networks and feature fusion," *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 22743–22772, 2019.

[15] C. Zhang, "Convolution analysis operator for Multimodal image fusion," *Procedia Computer Science*, vol. 183, no. 5, pp. 603–608, 2021.

[16] L.-W. Ko, C.-T. Lin, Y.-C. Lu et al., "Multimodal fuzzy fusion for enhancing the motor-imagery-based brain computer interface," *IEEE Computational Intelligence Magazine*, vol. 14, no. 1, pp. 96–106, 2019.

[17] X. Gui, X. Liu, Q. Tian, and W. Guo, "Multimodal Data Fusion in 3-D Printing Quality Prediction," *IEEE Sensors Letters*, vol. 3, no. 1, pp. 1–4, 2019.

[18] N. Bakalos, A. Voulodimos, N. Doulamis et al., "Protecting water infrastructure from cyber and physical threats: using multimodal data fusion and adaptive deep learning to monitor critical systems," *IEEE Signal Processing Magazine*, vol. 36, no. 2, pp. 36–48, 2019.

[19] A. Shamsan, W. Dan, and C. Cheng, "Multimodal data fusion using multivariate empirical mode decomposition for automatic process monitoring," *IEEE Sensors Letters*, vol. 3, no. 1, pp. 1–4, 2019.

[20] Y. R. Pandeya and J. Lee, "Deep learning-based late fusion of Multimodal information for emotion classification of music video," *Multimedia Tools and Applications*, vol. 80, no. 38, pp. 1–19, 2021.

[21] Z. Wu, K. Mao, and G.-W. Ng, "Enhanced feature fusion through irrelevant redundancy elimination in intra-class and extra-class discriminative correlation analysis," *Neurocomputing*, vol. 335, pp. 105–118, 2019.

[22] Y. Wu, Y. Zhao, and X. Lu, "Modeling incongruity between modalities for multimodal sarcasm detection," *IEEE Multimedia*, vol. 28, no. 99, p. 1, 2021.

[23] H. Mehraj and A. H. Mir, "A multi-biometric system based on multi-level hybrid feature fusion," *Herald of the Russian Academy of Sciences*, vol. 91, no. 2, pp. 176–196, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 2856818, 12 pages
https://doi.org/10.1155/2022/2856818

Research Article

Motor Imagery EEG Decoding Based on New Spatial-Frequency Feature and Hybrid Feature Selection Method

Yuan Tang,[1] Zining Zhao,[1] Shaorong Zhang,[1,2] Zhi Li,[1,2] Yun Mo,[2] and Yan Guo[2]

[1]*School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi, China*
[2]*School of Electronic Information and Automation, Guilin University of Aerospace Technology, Guilin, Guangxi, China*

Correspondence should be addressed to Shaorong Zhang; zsrong@guat.edu.cn

Received 21 October 2021; Accepted 27 November 2021; Published 7 January 2022

Academic Editor: Xiaofeng Li

Feature extraction and selection are important parts of motor imagery electroencephalogram (EEG) decoding and have always been the focus and difficulty of brain-computer interface (BCI) system research. In order to improve the accuracy of EEG decoding and reduce model training time, new feature extraction and selection methods are proposed in this paper. First, a new spatial-frequency feature extraction method is proposed. The original EEG signal is preprocessed, and then the common spatial pattern (CSP) is used for spatial filtering and dimensionality reduction. Finally, the filter bank method is used to decompose the spatially filtered signals into multiple frequency subbands, and the logarithmic band power feature of each frequency subband is extracted. Second, to select the subject-specific spatial-frequency features, a hybrid feature selection method based on the Fisher score and support vector machine (SVM) is proposed. The Fisher score of each feature is calculated, then a series of threshold parameters are set to generate different feature subsets, and finally, SVM and cross-validation are used to select the optimal feature subset. The effectiveness of the proposed method is validated using two sets of publicly available BCI competition data and a set of self-collected data. The total average accuracy of the three data sets achieved by the proposed method is 82.39%, which is 2.99% higher than the CSP method. The experimental results show that the proposed method has a better classification effect than the existing methods, and at the same time, feature extraction and feature selection time also have greater advantages.

1. Introduction

Motor imagery electroencephalogram (EEG) signal is widely used in brain-computer interface (BCI) system, but it has strong randomness and low signal-to-noise ratio and is easily disturbed by physiological and nonphysiological noises, which makes it difficult to decode [1]. In EEG decoding, feature extraction and selection are the core components [2]. On the one hand, extracting discriminative and stable features can effectively improve the performance of EEG decoding [3]. On the other hand, the extracted features usually contain noise and redundant information, so feature selection is required to eliminate invalid information [4]. In addition, feature selection can reduce the feature dimension and the complexity of the classification model and avoid dimension disaster and overfitting. Therefore, feature extraction and selection have always been the focus and difficulty of BCI system research.

Common spatial pattern (CSP) is a relatively effective method for feature extraction of motor imagery EEG among many methods [5]. The traditional CSP method extracts logarithmic variance as features after spatial filtering [6], but some studies have shown that this feature extraction method is not necessarily optimal. For example, literature [7] proposed the logarithmic band power (LBP) feature based on the CSP transform, which is called CSP-LBP in this paper. The experimental results show that CSP-LBP is superior to the traditional CSP method. In addition, the traditional CSP method lacks frequency domain information. Therefore, a lot of work has been done to select the optimal frequency band for CSP. For example, in literature [8], the original EEG signal was filtered into multiple subbands by filter bank method, and then CSP was used to extract features. Finally, mutual information was used to select the features of the optimal frequency band. Zhang et al. [9] proposed a sparse

filter band common spatial pattern (SFBCSP) method. SFBCSP carried out band-pass filtering on original EEG signals through multiple frequency subbands with a frequency range of 4–40 Hz, a bandwidth of 4 Hz, and an overlap rate of 2 Hz between subbands. CSP was used to extract features on each subband, and the least absolute shrinkage and selection operator (LASSO) was used for frequency band feature selection. Finally, the support vector machine (SVM) was used to classify selected features. Subsequently, Zhang et al. [10] proposed a subband optimization method that implements sparse Bayesian learning of frequency bands (SBLFB) for motor imagery classification. The subbands filtering method was the same as that in literature [9], but sparse Bayesian learning was used to select sparse frequency band features. The above spatial-frequency feature extraction methods filter the original EEG signals into multiple subbands, which requires a large amount of computation and a long time.

Existing feature selection methods mainly include filter, wrapper, and embedded [11]. The filter feature selection method uses evaluation criteria such as information measurement and distance measurement to select features. The wrapper feature selection method generates feature subsets in a specific way and then uses the results of classifiers as the evaluation criteria for feature selection. The embedded feature selection method can automatically remove some features during classifier training, so feature selection and classification can be carried out simultaneously. The above three types of methods have their advantages and disadvantages, and the organic combination of these methods can achieve complementary advantages. Therefore, the hybrid feature selection method has been studied widely in recent years. Moradi et al. [12] proposed a novel hybrid feature selection algorithm based on particle swarm optimization (PSO) and the local search strategy, and the local search strategy was embedded in the PSO to select the less correlated and salient feature subset. Jain et al. [13] proposed a hybrid model for gene selection and cancer classification, and the optimal gene subset was selected by correlation-based feature selection method combined with improved binary PSO. Lu et al. [14] proposed a hybrid feature selection algorithm for gene expression data classification. The algorithm combined the mutual information maximization and the adaptive genetic algorithm to reduce the dimension of gene expression data and remove the redundancies for classification. Ghareb et al. [15] combined six filtering feature selection methods and an improved genetic algorithm to form a new hybrid feature selection method. Literature [16] is a review article that comprehensively introduces the hybrid feature selection method for cancer classification. Although the hybrid feature selection method has been widely used, as far as we know, few hybrid feature selection methods have been applied to EEG decoding. In addition, the existing hybrid feature selection methods are mostly based on intelligent optimization algorithms such as PSO and genetic algorithms, so the feature selection time is relatively long.

In order to further improve the performance of motor imagery EEG decoding, a new spatial-frequency feature extraction method and hybrid feature selection method are proposed in this paper. First, considering the effectiveness of the CSP-LBP method [7] and the frequency defects of CSP, a new spatial-frequency feature extraction method based on CSP transform, filter bank (FB), and logarithmic band power (LBP) is proposed; we call it CSP-FBLBP. The original EEG signals are preprocessed and then spatially filtered by CSP transform. After that, the spatially filtered signals are decomposed into multiple subbands using a filter bank, and the logarithmic band power of each subband is extracted as the feature. Second, a new hybrid feature selection method based on Fisher score (F-score) and SVM is proposed to select subject-specific spatial-frequency features; we call it F-score-h. The Fisher score of each feature is calculated, then a series of threshold parameters are set to generate different feature subsets, and finally, SVM and 10-fold cross-validation are used to select the optimal feature subset. After feature extraction and selection are completed, SVM is used for classification. Two public data sets and a self-collected data set are used to verify the proposed method.

The main contributions of this paper are in two aspects:

First, a new spatial-frequency feature extraction method is proposed. CSP is used for dimensionality reduction, and then the spatial projection signal is band-pass filtered using the filter bank method. Finally, logarithmic band power is used for feature extraction. CSP dimensionality reduction effectively reduces the number of signal channels, thereby reducing the calculation amount and time of band-pass filtering, which greatly improves the timeliness of feature extraction. In addition, the experimental results show that logarithmic band power is more effective than the logarithmic variance of the traditional CSP method.

Second, a new hybrid feature selection method is proposed. The Fisher score is used for feature sorting, and the threshold method, SVM, and cross-validation are combined for optimal feature subset selection. The proposed method takes full advantage of the simple calculation of the filtering method and the supervised selection of the wrapped method, which not only reduces the feature selection time but also improves the classification performance of EEG decoding.

2. Materials and Methods

2.1. EEG Data Description. Data set 1: data set IIa of BCI competition IV (2008) [17]: this data set contains 22 electrode channels, and the sampling rate is 250 Hz. Nine healthy subjects (A01, A02, A03, A04, A05, A06, A07, A08, and A09) performed left-hand, right-hand, foot, and tongue motor imagery tasks, respectively. Since we only consider binary classification tasks, $C_4^2 = 6$ groups of binary classification tasks are obtained by permutation and combination of four types of tasks. Since there are nine subjects, $9 \times 6 = 54$ data subsets could be obtained. The number of samples in the training set and test set of each subject is 144, respectively.

Data set 2: data set IIb of BCI competition IV (2008) [18]: this data set contains 3 electrode channels, and the sampling rate is 250 Hz. Nine healthy subjects (B01, B02, B03, B04, B05, B06, B07, B08, and B09) performed left-hand and right-hand motor imagery tasks, respectively. This data set has five

sessions, and we only analyzed the data of the third session [10]. The number of samples in the training set and test set of each subject is 80, respectively.

Data set 3: data set self-collected from our laboratory: NuAmps amplifier and electrode cap from Neuroscan company are used for scalp EEG signals collection, and the sampling rate is 250 Hz. A total of 36 electrode channel data in this data set, including 30 channels of EEG data, 4 channels of electrooculogram data, and two reference channels, and only 30 channels of EEG data, are analyzed in this paper. Six healthy subjects (S01, S02, S03, S04, S05, and S06) performed left-hand and right-hand motor imagery tasks, respectively.

2.2. The Proposed Method. The data processing flow of the proposed method is shown in Figure 1, which mainly includes preprocessing, feature extraction, feature selection, and feature classification. In the preprocessing stage, all data sets performed 8–30 Hz band-pass filtering using a 6-order Butterworth filter. The time window with 0.5–2.5 s is selected for single-trial data extraction. In the following content, we will introduce the core work of the proposed method in detail.

2.2.1. The New Spatial-Frequency Feature Extraction Method. The new spatial-frequency feature extraction method includes CSP dimensionality reduction, filter bank band-pass filtering, and logarithmic band power feature extraction.

(1) CSP Dimensionality Reduction. The solution of the CSP objective function can be equivalent to a generalized eigenvalue problem [19]. After the eigenvector matrix is obtained, the eigenvectors corresponding to the first m largest eigenvalues and the last m smallest are selected to form the final spatial filter. Assuming that the spatial filter is $\mathbf{W}$ and the single-trial data is $\mathbf{D}$, the spatial projection signal $\mathbf{Z}$ can be calculated by the following calculation formula:

$$\mathbf{Z} = \mathbf{W}^T\mathbf{D}, \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{C\times 2m}$, $\mathbf{D} \in \mathbb{R}^{C\times K}$, C represents the total number of electrode channels, m represents the pair number of the spatial filters, and K represents the number of sampling points for each electrode channel.

After the single-trial data $\mathbf{D}$ is transformed by CSP, the EEG signal has only $2m$ channels. The value of m is usually set to 3 or 1, so the number of EEG signal channels is significantly reduced after CSP transformation. For example, data set 1 has 22 electrode channels; if m is set to 3, the EEG signal has only 6 channels after CSP dimensionality reduction. The specific form of the signal $\mathbf{Z}$ is as follows:

$$Z = \begin{bmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,K} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ z_{2m,1} & z_{2m,2} & \cdots & z_{2m,K} \end{bmatrix}. \tag{2}$$

(2) Filter Bank Band-Pass Filtering. The signal $\mathbf{Z}$ is band-pass filtered using a filter bank with frequency subbands of 4–8 Hz, 6–10 Hz, ..., 26–30 Hz. Specifically, band-pass filtering is performed on each channel of the signal $\mathbf{Z}$, which is shown in Figure 2.

(3) The Logarithmic Band Power Feature Extraction. The logarithmic variance is extracted as the feature in the traditional CSP method [6]. However, the experimental results in literature [7] prove that the logarithmic band power is more effective. Therefore, after band-pass filtering, the logarithmic band power is extracted as the feature in this paper, specifically as follows [7]:

$$LBP_p = \log\left(\frac{1}{K}\sum_{i=1}^{K}\left|Z_p(i)\right|^2\right), p = 1, 2, \ldots 2m, \tag{3}$$

where $Z_p(i)$ represents the $i-th$ sample point of the $p-th$ channel of the signals $\mathbf{Z}$.

In the newly proposed spatial-frequency feature extraction method, CSP spatial filtering is performed first, and then band-pass filtering and feature extraction are performed on the spatially filtered signal. This processing has two advantages. On the one hand, after the signal is spatially filtered by the CSP, the signal quality is improved, and the extracted features are more stable and more discriminative. On the other hand, after CSP dimensionality reduction, the signal channel is greatly reduced, thereby reducing the calculation amount of band-pass filtering. Therefore, the time of feature extraction is greatly reduced, and it is not affected by the actual number of electrode channels.

2.2.2. Hybrid Feature Selection Method. The Fisher score can measure the distinguishing ability of features between two categories [20]. The Fisher score is obtained by calculating the variance ratio between-classes and within-classes of each feature, details as follows:

$$F(i) = \frac{\left(\overline{x}_i^{(+)} - \overline{x}_i\right)^2 + \left(\overline{x}_i^{(-)} - \overline{x}_i\right)^2}{\left(1/\left(n_+ - 1\right)\right)\sum_{k=1}^{n+}\left(x_{k,i}^{(+)} - \overline{x}_i^{(+)}\right)^2 + \left(1/\left(n_- - 1\right)\right)\sum_{k=1}^{n-}\left(x_{k,i}^{(-)} - \overline{x}_i^{(-)}\right)^2}, \tag{4}$$

where $F(i)$ represents the Fisher score of $i-th$ future. n_+ and n_-, respectively, represent the number of positive samples and negative samples, and $n = n_+ + n_-$ represents the number of total samples. $\overline{x}_i$, $\overline{x}_i^{(+)}$, and $\overline{x}_i^{(-)}$ are, respectively,

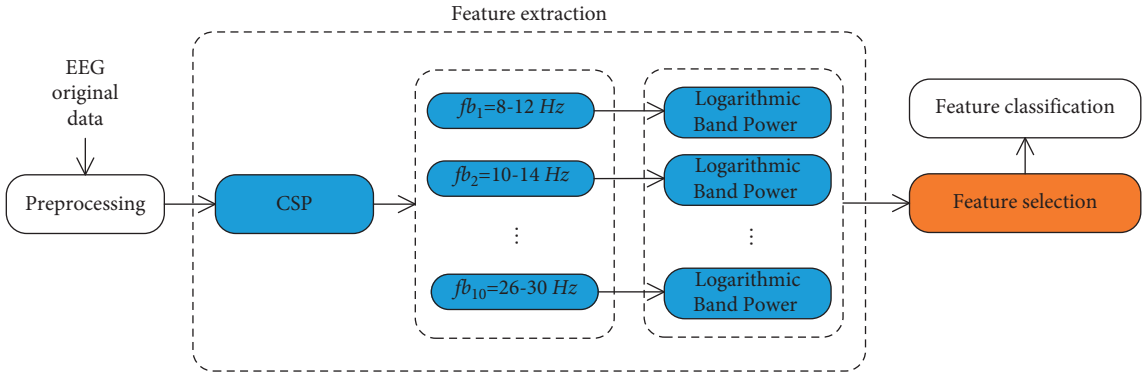


Figure 1: The data processing flow of the proposed method.

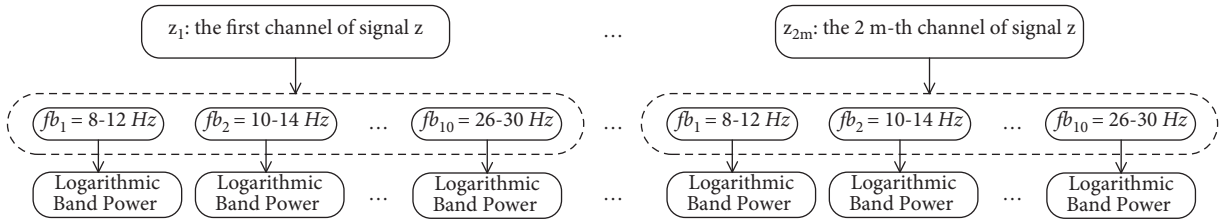


Figure 2: The band-pass filtering on each channel of the signal **Z**.

the mean value of the $i-th$ feature in the whole data sample set, the mean value of the $i-th$ feature in the positive sample set, and the mean value of the $i-th$ feature in the negative sample set. $x_k \in \mathbb{R}^p$ represents the $k-th$ feature sample, $x_{k,i}^{(+)}$ is the feature value of the $i-th$ feature of the $k-th$ positive class sample, $x_{k,i}^{(-)}$ is the feature value of the $i-th$ feature of the $k-th$ negative class sample, and p is the feature dimension. The larger the F value, the stronger the discrimination of the corresponding features [21]. The traditional F-score method sorts the features according to the Fisher score and then selects the top K features for subsequent classification.

However, it is difficult to determine exactly how many features should be selected to achieve the best classification effect. Therefore, a hybrid feature selection method based on F-score and SVM classifier is proposed in this paper; we call it F-score-h. Unlike the filter feature selection method based on F-score, F-score-h uses feature weights (i.e., the Fisher score of the feature) to select the optimal feature subset, as shown in Figure 3. Specifically, after the features are sorted by the Fisher score, we set a series of thresholds to generate different feature subsets; the features greater than the set threshold will be selected. The set of candidate parameters for the threshold is $Th \in \{0, 0.05, 0.1 \ldots, 0.8\}$. For each threshold parameter, the average verification accuracy of each feature subset is calculated by combining SVM and 10-fold cross-validation (CV). The threshold corresponding to the highest average accuracy is selected, and the optimal feature subset is further selected according to the optimal threshold.

The newly proposed hybrid feature selection method takes advantage of the small amount of calculation of the filtering method and the supervised selection of the wrapped method, which can take into account the time efficiency and the classification performance for feature selection at the same time.

2.2.3. SVM Classification. SVM is used as the classifier. The SVM classification model used in this paper is as follows [22]:

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{n}\xi_i, \\ &\text{subject to } y_i\left(\mathbf{w}^T\phi\left(\mathbf{x}_i\right) + b\right) \geq 1 - \xi_i, \\ &\xi_i \geq 0, i = 1, \ldots, l, n, \end{aligned} \tag{5}$$

where $\mathbf{x}_i \in \mathbb{R}^p$ represents the $i-th$ feature sample (feature vector). y_i represents the $i-th$ label. ξ_i represents the $i-th$ slack variable. x_i is mapped into a higher-dimensional space by $\phi(x_i)$, and $C > 0$ is the regularization parameter. Using the primal-dual algorithm to solve (5), the following decision function can be obtained:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{n} y_i\alpha_i K(x, x_i) + b\right), \tag{6}$$

where α_i is the Lagrange multiplier, $K(x, x_i) = \phi(x)^T\phi(x_i)$ is the kernel function, and $\operatorname{sgn}(\cdot)$ represents a symbolic function. SVM is implemented with the linear kernel using the LIBSVM toolbox [22]. The model parameter of SVM adopts the default setting of the toolbox [22].

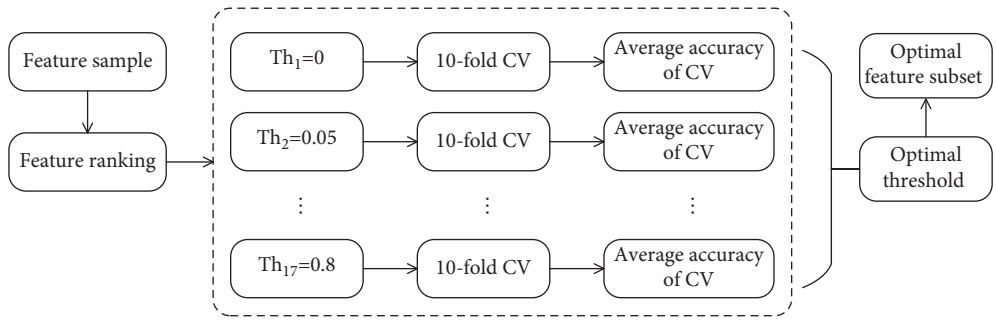


Figure 3: The hybrid feature selection method based on F-score and SVM.

3. Results

3.1. Results for Feature Extraction Methods

3.1.1. Comparison Methods and Parameter Settings. In order to verify the effectiveness of the proposed feature extraction method, the proposed method is compared with the other four CSP methods, which are the traditional CSP method [6, 19], CSP-FB [11], SFBCSP [9], SBLFB [10], and CSP-LBP [7]. If there is no special instruction, the pair number of spatial filters for CSP and its improvement methods are set as follows: $m = 3$ for data set 1 and data set 3; $m = 1$ for data set 2; SVM is used for classification. The comparison algorithms and their parameter settings are as follows:

CSP: CSP feature extraction refers to literature [6, 19].

CSP-FB: the parameter setting of the CSP-FB algorithm refers to literature [11]. F-score-h is used to select features.

SFBCSP: the parameter setting of the SFBCSP algorithm refers to literature [9]. Seventeen subbands (4–8 Hz, 6–10 Hz, . . ., 36–40 Hz) with a bandwidth of 4 Hz and an overlap rate of 2 Hz are used for band-pass filtering. A 6-order Butterworth filter is used. LASSO is used to select sparse band features.

SBLFB: the parameter setting of the SBLFB algorithm refers to literature [10]. The subbands setting is the same as SFBCSP. Sparse Bayesian learning is used to select sparse band features.

CSP-LBP: CSP-LBP feature extraction refers to literature [7].

CSP-FBLBP: CSP-FBLBP is used for feature extraction, and F-score-h is used for feature selection.

3.1.2. Experimental Results. Tables 1–3, respectively, show the classification accuracy and the total average classification accuracy of all the subjects in the three data sets. The highest accuracy is marked in bold. In Table 1, the left-hand, right-hand, foot, and tongue motor imagery tasks in data set 1 were represented by letters *L*, R, F, and *T*, respectively. *L* versus *R* means left-hand and right-hand binary classification tasks, and the others can be deduced by analogy. Due to space constraints, only the average classification accuracy of each binary classification task is given. It can be seen from Table 1 that CSP-FBLBP achieves the highest average classification accuracy on data set 1, and the accuracy is 3.24% higher than that of CSP. CSP-LBP and CSP-FB are better than CSP, but SFBCSP and SBLCSP are lower than CSP. Similarly, CSP-FBLBP also achieved the highest average classification accuracy on data sets 2 and 3; see Table 2 and Table 3 for details.

In order to make a more intuitive comparison of the classification effect achieved by various methods, the average classification accuracy achieved by different feature extraction methods is shown in Figure 4. It can be seen from Figure 4 that the classification effect of CSP-FBLBP is significantly better than other methods. The total average classification accuracy of CSP, CSP-FB, SFBCSP, SBLFB, CSP-LBP, and CSP-FBLBP in all data is 79.40, 80.53, 75.88, 75.63, 80.01, and 82.39, respectively.

Furthermore, the distribution of classification accuracy achieved by various feature extraction methods is shown in Figure 5. The red line represents the median value of classification accuracy. It can be seen that the median value of CSP-FBLBP is higher than that of other methods. The maximum value of CSP-FBLBP is 100%, and the minimum value of CSP-FBLBP is also higher than other methods. In addition, the accuracy distribution of CSP-FBLBP is relatively compact and close to the top. Therefore, CSP-FBLBP is superior to other methods.

In order to fully reflect the advantages of CSP-FBLBP, we further studied the time efficiency of CSP-FBLBP; the running time of various feature extraction methods is shown in Table 4. The training sets of the three subjects (A01, B01, and S01) are selected to calculate the feature extraction time. The feature extraction time of CSP-FB, SFBCSP, SBLFB, and CSP-FBLBP includes two parts, namely, CSP spatial filtering time and band-pass filtering time. For CSP-FB, SFBCSP, SBLFB, and CSP-FBLBP methods, two types of time are listed in brackets. The first one represents the spatial filtering time, and the last one represents the band-pass filtering time. It can be seen from Table 4 that the feature extraction time of SFBCSP and SBLFB is the longest, mainly because their band-pass filtering is relatively time-consuming. It is worth pointing out that the feature extraction process of SFBCSP and SBCSP is the same, so the feature extraction time is the same. Although the feature extraction time of CSP-FBLBP is longer than that of CSP-LBP and CSP, such time does not affect the use of CSP-FBLBP in a real-time BCI system. In addition, we can see that CSP-FBLBP has a greater time advantage than CSP-FB, SFBCSP, and SBLFB.

Table 1: The classification accuracy achieved by various feature extraction methods (data set 1).

Subject	CSP	CSP-FB	SFBCSP	SBLFB	CSP-LBP	CSP-FBLBP
L versus R	77.62	**79.01**	76.39	77.01	77.47	77.78
L versus F	82.41	85.88	79.32	80.25	82.33	**87.58**
L versus T	85.26	85.42	82.87	81.25	**85.80**	85.65
R versus F	81.02	84.72	76.93	76.62	84.10	**86.96**
R versus T	82.95	81.79	79.78	79.32	83.34	**85.80**
F versus T	72.61	72.61	71.30	70.68	74.46	**77.55**
Mean ± Std	80.31 ± 14.25	81.57 ± 12.56	77.76 ± 12.06	77.52 ± 12.03	81.25 ± 12.73	**83.55 ± 12.32**

Table 2: The classification accuracy achieved by various feature extraction methods (data set 2).

Subject	CSP	CSP-FB	SFBCSP	SBLFB	CSP-LBP	CSP-FBLBP
B01	80.00	**83.75**	78.75	77.50	80.00	76.25
B02	60.00	55.00	50.00	42.50	56.25	**61.25**
B03	43.75	41.25	47.50	46.25	43.75	**50.00**
B04	97.50	97.50	97.50	**98.75**	92.50	97.50
B05	87.50	**93.75**	81.25	80.00	91.25	92.50
B06	**81.25**	76.25	73.75	75.00	80.00	78.75
B07	81.25	**87.50**	85.00	82.50	82.50	**87.50**
B08	92.50	92.50	80.00	76.25	91.25	**95.00**
B09	80.00	80.00	70.00	70.00	**83.75**	78.75
Mean ± Std	78.19 ± 16.61	78.61 ± 18.86	73.75 ± 16.12	72.08 ± 17.62	77.92 ± 16.84	**79.72 ± 15.86**

Table 3: The classification accuracy achieved by various feature extraction methods (data set 3).

Subject	CSP	CSP-FB	SFBCSP	SBLFB	CSP-LBP	CSP-FBLBP
S01	70.31	68.75	51.56	50.00	67.19	**76.56**
S02	68.75	**72.92**	65.63	68.75	64.06	70.83
S03	68.75	**84.62**	64.58	62.50	72.92	78.46
S04	**78.13**	75.00	70.31	73.44	79.69	**78.13**
S05	75.00	**78.13**	43.75	45.31	73.44	76.56
S06	77.08	64.62	77.08	**83.33**	75.00	75.38
Mean ± Std	73.00 ± 4.25	74.01 ± 7.04	62.15 ± 12.31	63.89 ± 14.37	72.05 ± 5.61	**75.99 ± 2.77**

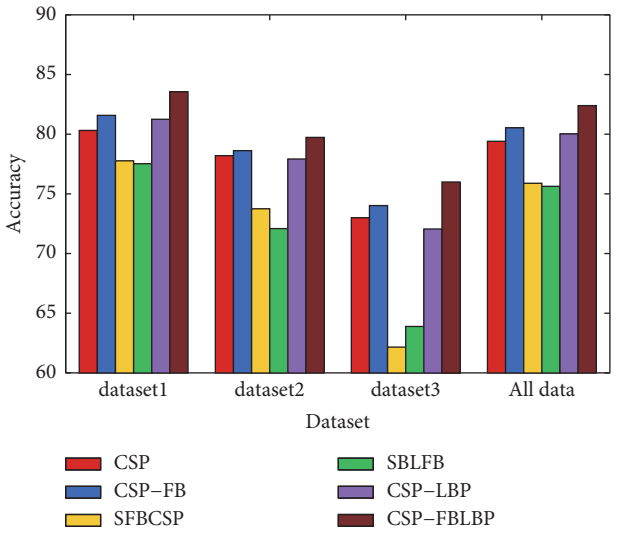


Figure 4: The average classification accuracy achieved by various feature extraction methods.

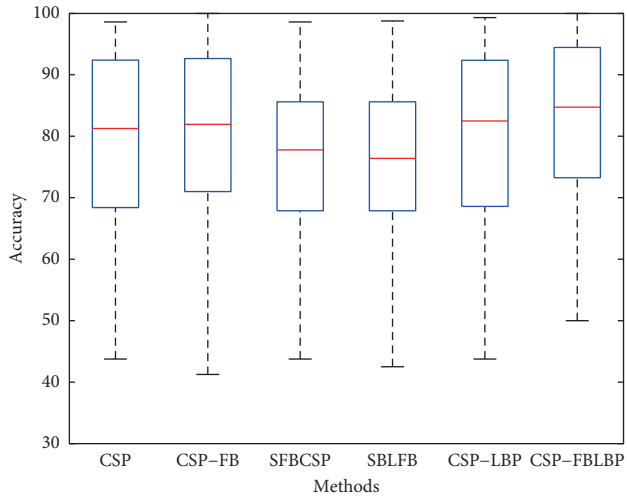


Figure 5: The distribution of classification accuracy achieved by various feature extraction methods.

Table 4: The feature extraction time of various methods (unit: second).

Subject	CSP	CSP-FB	SFBCSP (SBLFB)	CSP-LBP	CSP-FBLBP
A01	0.076	0.912 (0.041, 0.871)	55.380 (0.774, 54.606)	0.052	0.542 (0.040, 0.502)
B01	0.034	0.264 (0.025, 0.239)	11.769 (0.499, 11.270)	0.027	0.173 (0.024, 0.149)
S01	0.050	0.537 (0.026, 0.511)	54.765 (0.756, 54.009)	0.031	0.326 (0.027, 0.299)

3.2. Results for Feature Selection Methods

3.2.1. Comparison Methods and Parameter Settings. In order to verify the effectiveness of the proposed feature selection method, the proposed method was compared with four other feature selection methods, namely, LASSO [23], genetic algorithm (GA) [24], binary particle swarm optimization (BPSO) algorithm [25], and binary differential evolution (BDE) algorithm [26]. CSP-FBLBP is used for feature extraction, and SVM is used for classification.

LASSO: the alternative parameter set for LASSO is $\lambda \in 0.1 \times \{1, 2, \ldots, 30\}$. The optimal parameter is selected by 10-fold cross-validation. LASSO with regression model is implemented by the SLEP toolbox [27]. After the LASSO model is determined, the features with a weight coefficient greater than 0 are selected as the optimal feature subset.

GA: the parameter setting of GA refers to literature [24]. The binary encoding is selected as the feature encoding method. The fitness function is the classification accuracy of the k-nearest neighbor classifier, where $k = 5$. The population size is 10, the number of iterations is used as the termination condition of the algorithm, and the maximum number of iterations is 100. The crossover probability is 0.8, and the mutation probability is 0.01.

BPSO: the implementation of BPSO refers to literature [25], and the parameter setting is consistent with literature [25]. The fitness function is the classification accuracy of the k-nearest neighbor classifier, where $k = 5$. The population size is 10, the number of iterations is used as the termination condition of the algorithm, and the maximum number of iterations is 100. The acceleration coefficients of the BPSO are set as $c_1 = 2, c_2 = 2$. The maximum and minimum velocities are 6 and −6, respectively. The maximum and minimum inertial weights are 0.9 and 0.4, respectively.

BDE: the parameter setting of BDE refers to literature [26]. The population size is 10, and the number of iterations is used as the termination condition of the algorithm, and the maximum number of iterations is set to 100. The crossover probability is 0.9.

3.2.2. Experimental Results. The average classification accuracy achieved by different feature selection methods is shown in Figure 6. The total average classification accuracy of LASSO, BPSO, GA, BDE, and F-score-h in all data is 77.57, 80.62, 80.56, 80.43, and 82.39, respectively. F-score-h is significantly better than other feature selection methods. BPSO, GA, and BDE are equally effective, and the effect of LASSO is relatively poor.

The distribution of classification accuracy achieved by various feature selection methods is shown in Figure 7. It can be seen that the median value of F-score-h is higher than that of other methods. The maximum value of F-score-h is 100%, and the minimum value of F-score-h is also higher than other methods. In addition, the overall classification accuracy distribution of F-score-h is relatively compact and close to the top. These results fully prove the superiority of F-score-h.

The running time of various feature selection methods is shown in Table 5. The training sets of the three subjects (A01, B01, and S01) are selected to calculate the feature selection time. The feature selection time of F-score-h is the

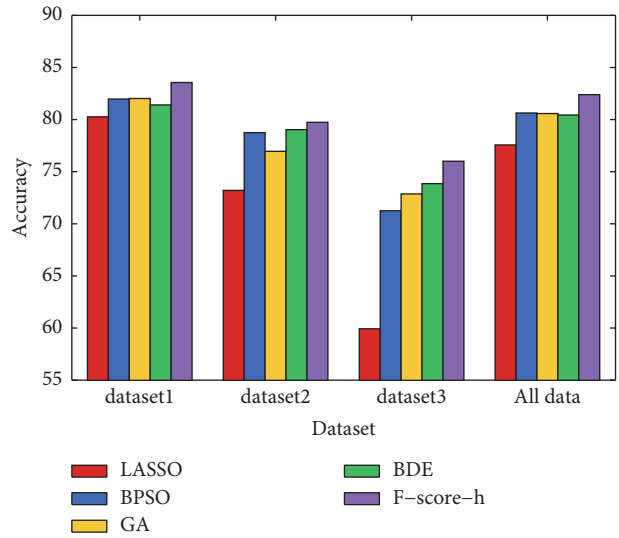


Figure 6: The average classification accuracy achieved by various feature selection methods.

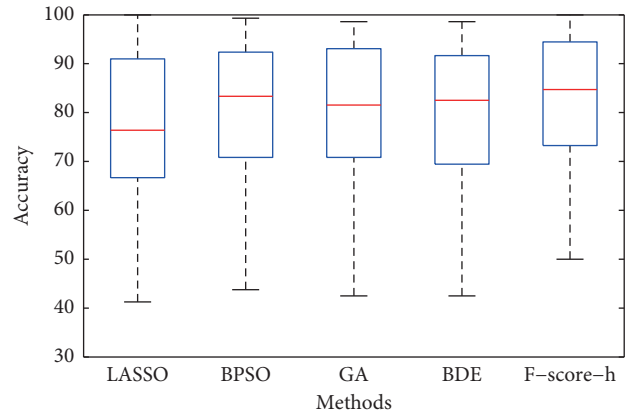


Figure 7: The distribution of classification accuracy achieved by various feature selection methods.

Table 5: The feature selection time of various feature selection methods (unit: second).

Subject	LASSO	BPSO	GA	BDE	F-score-h
A01	2.865	39.917	74.561	47.257	0.070
B01	1.879	40.337	77.432	45.693	0.050
S01	2.271	47.711	73.017	46.408	0.037

shortest, which is much lower than other methods. The feature selection time of BPSO, GA, and BDE methods is relatively long. In contrast, F-score-h has a huge time advantage.

In summary, F-score-h has great advantages in classification performance and feature selection time.

3.3. Results Compared with Other Existing Methods. In order to more fully reflect the advantages of the proposed method, the classification results of the proposed method are compared with that of the recently published papers. The classification results of data set 1 (L versus *R* binary classification task) are shown in Table 6, and the classification results of data set 2 are shown in Table 7. In data set 1, the proposed method is superior to other existing methods. In data set 2, the proposed method is second only to the NCFS method [35] and is better than most existing methods. From the above experimental results, it can be seen that the classification effect of the proposed method has certain advantages.

4. Discussion

Comparing the classification results of CSP-FB and CSP as well as CSP-FBLBP and CSP-LBP can prove that selecting a subject-specific frequency band can improve the classification performance of CSP. The reason why CSP-FB and CSP-FBLBP obtain better classification results is mainly that these two methods use the filter bank method to make up for

Table 6: The classification accuracy achieved by various methods (data set 1: left-hand versus right-hand).

Subject	GRU-RNN [28] (2018)	IST-TSVM [29] (2019)	ICA + PSR + CSP [30] (2020)	p-LTCSP [31] (2020)	CSP-FBLBP
A01	84.82	80.14	80	82.6	**92.36**
A02	65.32	51.55	65.36	**70.23**	54.17
A03	83.54	95.54	87.14	70.23	**93.75**
A04	**67.67**	53.6	67.5	55.15	67.36
A05	64	51.65	55.54	54.36	**64.58**
A06	**70.87**	56.83	50.18	60.14	64.58
A07	84.96	56.58	**91.79**	73.38	77.78
A08	71.95	93.42	84.11	85.29	**93.75**
A09	68.9	**92.66**	87.86	74.62	91.67
Mean ± Std	73.56 ± 8.53	70.22 ± 19.74	74.39 ± 15.18	69.56 ± 11.10	**77.78 ± 15.53**

Table 7: The classification accuracy achieved by various methods (data set 2).

Subject	DBN [32] (2018)	CapsNet [33] (2019)	SGRM [34] (2019)	NCFS [35] (2020)	CSP-FBLBP
B01	70.38	78.75	77.3	**79.25**	76.25
B02	**70.34**	55.71	59.1	63.48	61.25
B03	71.2	55	51.5	56.65	50
B04	71.24	95.93	97	**99.28**	97.5
B05	71.21	83.12	87.4	88.67	**92.5**
B06	70.52	**83.43**	72.5	79.96	78.75
B07	70.79	75.62	86.7	**88.76**	87.5
B08	70.49	91.25	84.7	92.66	**95**
B09	70.32	**87.18**	85.6	84.95	78.75
Mean ± Std	70.72 ± 0.4	78.44 ± 14.44	78.0 ± 14.65	**81.52 ± 13.72**	79.72 ± 15.86

the frequency information of CSP. To further illustrate the problem, two subjects (A01 and B01) have been selected to show the spatial-frequency feature selection results, where features are extracted by CSP-FBLBP and selected by F-score-h. A total of six channels (B01 only has two channels) of signals are retained after CSP dimensionality reduction. The feature index 1–10 in Figure 8 corresponds to the features of the first channel signal filtered by 8–12 Hz, . . ., 26–30 Hz band-pass filter. The other feature indexes can be deduced by analogy. Figures 8(a) and 8(c) show the feature weights calculated by Fisher score; Figures 8(b) and 8(d) show the features selected by F-score-h. From the feature index, it is possible to calculate which channel and which frequency band the selected feature belongs to. It can be seen from Figure 8 that only a few features with high scores are retained. The channel (spatial information) and frequency band (frequency information) selected for different subjects are different; that is, the optimal spatial-frequency features are subject-specific. CSP-FBLBP jointly considers subject-specific spatial-frequency features, so a better classification result is achieved.

In addition, CSP-FBLBP and CSP-FB use the same method to compensate for the frequency defects of CSP, but the classification effect of CSP-FBLBP is better than CSP-FB. This result shows that the feature type extracted after CSP spatial filtering is also very critical. It can be seen from the experimental results that the logarithmic band power is better than the traditional logarithmic variance features. Based on CSP transformation, it is worth studying to further improve the feature extraction method.

Compared with the existing spatial-frequency feature extraction methods (CSP-FB, SFBCSP, and SBLCSP), CSP-FBLBP has a greater time advantage, and the feature extraction time is significantly lower than the existing methods. The feature extraction time of SFBCSP and SBLFB is the longest, mainly due to the time of band-pass filtering. SFBCSP and SBLFB decompose the original EEG signals into 17 frequency subbands. The number of subbands and channels is relatively large, and the amount of calculation is relatively large, so the feature extraction time is long. After CSP dimensionality reduction, the number of signal channels of CSP-FBLBP is greatly reduced, so its feature extraction time is significantly reduced. In addition, comparing the feature extraction time of CSP-LBP and CSP as well as CSP-FBLBP and CSP-FB can show that the calculation time and complexity of logarithmic band power features are lower than a logarithmic variance.

From the comparative analysis of the above experimental results, it can be concluded that the F-score-h feature selection method has achieved better classification results, and its feature selection time also has significant advantages. F-score-h is a hybrid of filter and wrapper feature selection methods. On the one hand, the filter feature selection method has a small amount of calculation, so the calculation time is short; on the other hand, the wrapper feature selection method uses the classification performance of the classifier as an evaluation standard, its classification performance is generally better. F-score-h takes into account the advantages of both filter and wrapper methods, so it has achieved better classification performance.

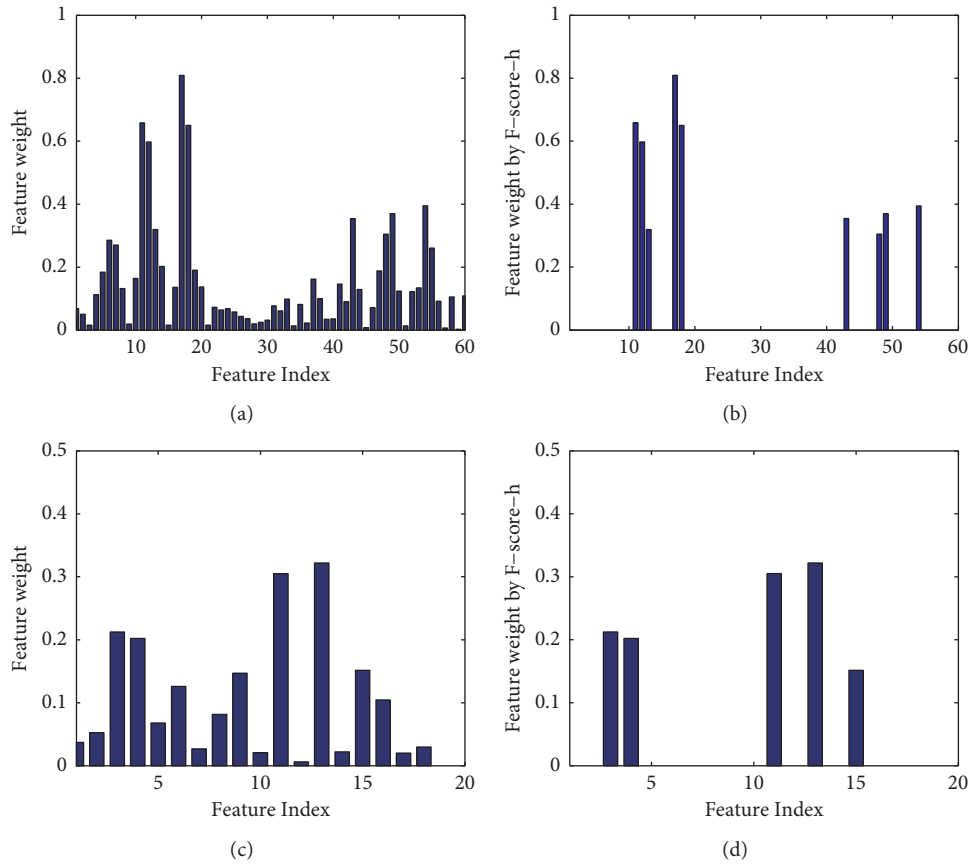


Figure 8: The feature selection by F-score-h.

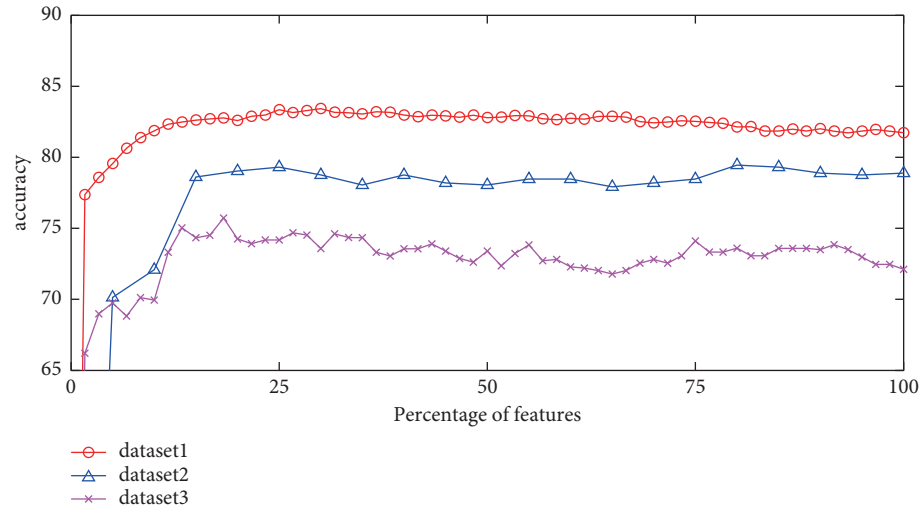


Figure 9: The variation of the average accuracy with the number of features (F-score).

In this paper, for the LASSO method, we select the features whose weight is greater than 0 as the optimal feature subset. Generally, the larger the feature weight, the more important the corresponding feature. However, the optimal feature subset is selected by the LASSO model, which is not necessarily the best on the SVM classifier [36]. Combined with LASSO and SVM for feature selection, the classification effect may be better [37]. The classification effect of BPSO, GA, and BDE is relatively poor; there are many reasons. First of all, the genetic algorithm may fall into a local optimal situation, BPSO may appear a "premature" phenomenon [25], and BDE may not be able to effectively converge. In addition, the selection of initialization parameters for BPSO, GA, and BDE also has a great influence on feature selection. How to choose more suitable model parameters is a very critical issue.

Table 8: The classification accuracy achieved by F-score and F-score-h.

Subject	F-score				F-score-h
	25%	50%	75%	100%	
Dataset 1	83.35 ± 12.23	82.78 ± 12.65	82.52 ± 12.42	81.74 ± 13.17	**83.55 ± 12.32**
Dataset 2	79.31 ± 15.30	78.06 ± 17.35	78.47 ± 16.70	78.89 ± 16.30	**79.72 ± 14.95**
Dataset 3	74.17 ± 2.73	73.40 ± 5.58	74.11 ± 7.07	72.12 ± 5.66	**75.99 ± 2.53**
Dataset all	82.02 ± 12.49	81.35 ± 13.25	81.26 ± 12.97	80.53 ± 13.45	**82.39 ± 12.41**

To further compare the advantages of F-score-h, we compared F-score-h with F-score. F-score is a filter method that uses Fisher scores to rank the features and then select the top K features for classification. The classification accuracy when a different number of features is selected is shown in Figure 9. The feature dimension extracted in data set 2 is only 20, while the feature dimension of data sets 1 and 3 is 60; in order to compare the variation of the average classification accuracy of the three data sets with the number of features on the same axis, the number of features is mapped to a percentage. For example, 25 on the abscissa of Figure 9 indicates that the selected feature number is 25% of the total number of features. It can be seen from Figure 9 that the number of features corresponding to the optimal classification accuracy of each data set is different. The filtering method selects the same number of features for all data sets, and its classification effect is not good. F-score-h selects features through feature weights, so the selected feature subset is more discriminative and contains less redundant information. In addition, F-score-h can select subject-specific features. Therefore, F-score-h is better than F-score.

The average classification accuracy of each data set when the F-score takes a different number of features is shown in Table 8. It can be intuitively observed from Table 8 that the accuracy of F-score-h is optimal in each data set.

5. Conclusion

The new feature extraction and selection methods have been proposed in this paper. In the new feature extraction method, the logarithmic band power is used to replace the logarithmic variance, and the filter bank method is used to compensate for the frequency defects of CSP. In the new feature selection method, the Fisher score is used to sort the features, and then a series of threshold parameters are set; SVM combined with cross-validation is used to select the optimal threshold parameters so as to obtain the optimal feature subset. The experimental results show that the proposed feature extraction and feature selection method has better classification performance than existing methods, and both feature extraction time and feature selection time have greater advantages, which can be applied to real-time BCI systems.

Although the proposed method has achieved good classification results, the impact of the time window on CSP is not considered in the feature extraction process, and the proposed method still has a large room for improvement. In future work, we will jointly consider efficient time-spatial-frequency feature extraction and selection methods.

Data Availability

In this study, we used three data sets for experiments. Data sets 1 and 2 are public BCI competition data sets, which have been deposited on the BCI competition website. Data set 3 is self-collected by our laboratory and is not publicly available but can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest

All authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 61967004, 11901137, and 81960324), Natural Science Foundation of Guangxi Province (no. 2018GXNSFBA281023), and Guangxi Key Laboratory of Automatic Testing Technology and Instruments (nos. YQ20113 and YQ19209).

References

[1] A. Khalaf and M. Akcakaya, "A probabilistic approach for calibration time reduction in hybrid EEG–fTCD brain–computer interfaces," *BioMedical Engineering Online*, vol. 19, no. 1, pp. 1–18, 2020.

[2] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain-computer interfaces using motor-imagery: techniques and challenges," *Sensors*, vol. 19, no. 6, p. 1423, 2019.

[3] N. Gursel Ozmen, L. Gumusel, and Y. Yang, "A biologically inspired approach to frequency domain feature extraction for EEG classification," *Computational and Mathematical Methods in Medicine*, vol. 2018, 2018.

[4] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs," *Pattern Recognition*, vol. 45, no. 6, pp. 2137–2144, 2012.

[5] B. Wang, C. M. Wong, Z. Kang et al., "Common spatial pattern reformulated for regularizations in brain-computer interfaces," *IEEE Transactions on Cybernetics*, vol. 51, no. 10, pp. 5008–5020, 2020.

[6] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-r. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[7] M. Aljalal, R. Djemal, K. AlSharabi, and S. Ibrahim, "Feature extraction of EEG based motor imagery using CSP based on logarithmic band power, entropy and energy," in *Proceedings of the 2018 1st International Conference on Computer*

Applications & Information Security (ICCAIS), pp. 1–6, IEEE, Riyadh, Saudi Arabia, April 2018.

[8] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, p. 39, 2012.

[9] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface," *Journal of Neuroscience Methods*, vol. 255, pp. 85–91, 2015.

[10] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *International Journal of Neural Systems*, vol. 27, no. 02, Article ID 1650032, 2017.

[11] S. Zhang, Z. Zhu, B. Zhang, B. Feng, T. Yu, and Z. Li, "The CSP-based new features plus non-convex log sparse feature selection for motor imagery EEG classification," *Sensors*, vol. 20, no. 17, p. 4749, 2020.

[12] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," *Applied Soft Computing*, vol. 43, pp. 117–130, 2016.

[13] I. Jain, V. K. Jain, and R. Jain, "Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification," *Applied Soft Computing*, vol. 62, pp. 203–215, 2018.

[14] H. Lu, J. Chen, K. Yan, Q. Jin, Y. Xue, and Z. Gao, "A hybrid feature selection algorithm for gene expression data classification," *Neurocomputing*, vol. 256, pp. 56–62, 2017.

[15] A. S. Ghareb, A. A. Bakar, and A. R. Hamdan, "Hybrid feature selection based on enhanced genetic algorithm for text categorization," *Expert Systems with Applications*, vol. 49, pp. 31–47, 2016.

[16] N. Almugren and H. Alshamlan, "A survey on hybrid feature selection methods in microarray gene expression data for cancer classification," *IEEE Access*, vol. 7, pp. 78533–78548, 2019.

[17] C. Zuo, J. Jin, R. Xu et al., "Cluster decomposing and multi-objective optimization based-ensemble learning framework for motor imagery-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 18, no. 2, Article ID 26018, 2021.

[18] C. Zuo, Y. Miao, X. Wang, L. Wu, and J. Jin, "Temporal frequency joint sparse optimization and fuzzy fusion for motor imagery-based brain-computer interfaces," *Journal of Neuroscience Methods*, vol. 340, p. 108725, 2020.

[19] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2010.

[20] A. Shoeibi, N. Ghassemi, R. Alizadehsani et al., "A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals," *Expert Systems with Applications*, vol. 163, Article ID 113788, 2021.

[21] M. Radman, M. Moradi, A. Chaibakhsh, M. Kordestani, and M. Saif, "Multi-feature fusion approach for epileptic seizure detection from EEG signals[J]," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3533–3543, 2020.

[22] C.-C. Chang and C.-J. Lin, "Libsvm," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[23] A. Meziani, K. Djouani, T. Medkour, and A. Chibani, "A Lasso quantile periodogram based feature extraction for EEG-based motor imagery," *Journal of Neuroscience Methods*, vol. 328, Article ID 108434, 2019.

[24] J. Too, A. Abdullah, N. Mohd Saad, and W. Tee, "EMG feature selection and classification using a Pbest-guide binary particle swarm optimization," *Computation*, vol. 7, no. 1, p. 12, 2019.

[25] J. Too, A. R. Abdullah, and N. Mohd Saad, "A new Co-evolution binary particle swarm optimization with multiple inertia weight strategy for feature selection," *Informatics*, vol. 6, no. 2, p. 21, 2019.

[26] D. Datta and S. Dutta, "A binary-real-coded differential evolution for unit commitment problem," *International Journal of Electrical Power & Energy Systems*, vol. 42, no. 1, pp. 517–524, 2012.

[27] J. Liu, S. Ji, and J. Ye, "SLEP: sparse learning with efficient projections," *Arizona State University*, vol. 6, no. 491, p. 7, 2009.

[28] E. C. Djamal and R. D. Putra, "Brain-computer interface of focus and motor imagery using wavelet and recurrent neural networks[J]," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 18, no. 4, pp. 2748–2756, 2020.

[29] Y. Xu, J. Hua, H. Zhang et al., "Improved transductive support vector machine for a small labelled set in motor imagery-based brain-computer interface[J]," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 2087132, 2019.

[30] E. Dong, K. Zhou, J. Tong, and S. Du, "A novel hybrid kernel function relevance vector machine for multi-task motor imagery EEG classification," *Biomedical Signal Processing and Control*, vol. 60, Article ID 101991, 2020.

[31] Z. Yu, T. Ma, N. Fang, H. Wang, Z. Li, and H. Fan, "Local temporal common spatial patterns modulated with phase locking value," *Biomedical Signal Processing and Control*, vol. 59, Article ID 101882, 2020.

[32] Y. Chu, X. Zhao, Y. Zou, W. Xu, J. Han, and Y. Zhao, "A decoding scheme for incomplete motor imagery EEG with deep belief network," *Frontiers in Neuroscience*, vol. 12, p. 680, 2018.

[33] K.-W. Ha and J.-W. Jeong, "Motor imagery EEG classification using capsule networks," *Sensors*, vol. 19, no. 13, p. 2854, 2019.

[34] Y. Jiao, Y. Zhang, X. Chen et al., "Sparse group representation model for motor imagery EEG classification[J]," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 631–641, 2018.

[35] M. K. I. Molla, A. A. Shiam, M. R. Islam, and T. Tanaka, "Discriminative feature selection-based motor imagery classification using EEG signal," *IEEE Access*, vol. 8, pp. 98255–98265, 2020.

[36] J. Li, K. Cheng, S. Wang et al., "Feature selection: a data perspective," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, 2017.

[37] M. Miao, H. Zeng, A. Wang, C. Zhao, and F. Liu, "Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery EEG: an sparse regression and Weighted Naïve Bayesian Classifier-based approach," *Journal of Neuroscience Methods*, vol. 278, pp. 13–24, 2017.