Facial emotion recognition via stationary wavelet entropy and Biogeography-based optimization

Xiang Li[1], Chaosheng Tang[1], Junding Sun[1,*]

[1]College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, PR China

Abstract

INTRODUCTION: As one of the important research directions in the field of computer vision, facial emotion recognition plays an important role in people's daily life. How to make the computer accurately read facial emotion is an important research content.

OBJECTIVES: In the current research on facial emotion recognition, there are some problems such as poor generalization ability of network model and low robustness of recognition system. To solve above problems, we propose a novel facial emotion recognition method.

METHODS: Our method of feature extraction using the stationary wavelet entropy, which combines single hidden layer feedforward neural network with biogeography-based optimization for facial emotion recognition.

RESULTS: The simulation results show that the overall accuracy of our method is 93.79±1.24%.

CONCLUSION: This model is superior to the current mainstream facial emotion recognition models in the performance of facial emotion detection. In future research, we will try deep learning and other training methods.

_____

*Corresponding author. Email: sunjd@hpu.edu.cn

1. Introduction

With the rapid development of computer technology and neural network technology, there is a higher and higher demand in the degree of automation for current people. We yearn for communication between people and computers to be the same as that between people. In the process of communication, to make computers accurately feel facial emotions of each other, it requires computers to have the same ability of understanding emotion as human beings. Therefore, facial emotion recognition is an effective way

to achieve it. Facial emotion recognition refers to the separation of specific facial states from a given static image or dynamic video sequence, to determine the psychological emotions of the object to be recognized. On the other hand, facial emotion is an important non-verbal expression in daily life. By observing the change of facial emotions, we can better recognize the psychological changes of the other party. Similarly, facial emotion recognition is also an important research direction in the field of computer vision. Accurate recognition of facial emotions through computer vision recognition system is conducive to the smooth progress of psychological recognition, human-computer interaction, assisted driving, station security and other work.

The purpose of facial emotion recognition is to analyze emotion and use the recognize results to identify specific emotions. According to the definition of basic facial emotions given by American psychologists Ekman and Friesen [1] in 1971, we can identify facial emotions into seven emotions: happy, sadness, fear, anger, surprise, disgust and neutral. The process of facial emotion recognition can be divided into four stages: image acquisition, image preprocessing, feature extraction and classify discrimination, of which the two most important stages are feature extraction and classify discrimination. With the development of convolutional neural network and its own characteristics, many scholars tend to use convolutional neural network to extract image features. Ali, et al. [2] proposed to use support vector machine (SVM) method. Evans [3] used Haar wavelet transform (HWT) method. Ivanovsky, et al. [4] proposed to use convolutional neural network on GPU for feature extraction in facial smile emotion recognition. Hasani and Mahoor [5] proposed a converged network architecture that uses three sets of Inception-Resnet modules, and combines conditional random fields (CRF) to greatly improve the accuracy of facial emotion recognition. Yang [6] introduced cat swarm optimization (CSO) and achieved good results in facial emotion recognition. Lucy, et al. [7] proposed a double-channel convolutional neural network, the first channel inputs extracted eye features, the second channel inputs extracted mouth features. After the fusion of the above two features, input the full connection layer,

to improve the accuracy of facial emotion recognition. Through the analysis of the above, we can find that the facial emotion features extracted by the above methods are not stable in translation operation and are easy to lose the original emotional information. Moreover, network models all have poor generalization and weak robustness. When the performance environment of the network model changes slightly, the network model may get stuck in the process of running, which directly affects the speed and accuracy of facial emotion recognition.

To solve the above problems, we proposed an improved facial emotion recognition model. The stationary wavelet entropy (SWE) is introduced for feature extraction, takes single hidden layer feedforward neural network (SHLFNN) as the classifier, and uses biogeography-based optimization (BBO) as the training method of our model. Through the following simulation experiments, it can be proved that the recognition performance of the proposed facial emotion recognition method is better than that state-of-the-art approaches.

The structure of the rest is organized in the following way. The second part introduced the subjects and data sets. The third part introduced the concept and basic principle of method. The fourth part introduced and analyzed the relevant experimental results. The fifth part summarized and anticipated the paper.

2. Dataset

In order to make the experiment easier to achieve and the experimental results more comparative, the data set adopted in this paper is from [8] data set. The data set in [8] is collected by an experienced photographer who used Canon digital camera (Canon EOS 70D) to capture the facial emotions of each subject ten times for 20 subjects of different ages, different careers and different races. Including seven kinds of facial emotional images: happy, sadness, fear, anger, surprise, disgust, and neutral. Among them, reference [9] is a facial model proposed by F.Y. Shih and C.F. Chuang. To make our data set more practical, we have made slight modifications to the collected images. And our dataset contains only facial images, as shown in

Table 1 and Figure 1Table 2. In the final, we have 700 images in total.

Table 1. Dataset of face models

| Define distance between two centers of both eyes is D, then we perform the following modification. |
|---|
| 1. Width of nose is 0.8D. |
| 2. Wide of mouse is D. |
| 3. Vertical distance between eyebrows and eyes is 0.4D. |
| 4. Vertical distance between eyes and center of nostril is 0.6D. |
| 5. Vertical distance between eyes and center of mouth is D. |
| 6. Extend upper, lower, left and right margin 0.4D. |
| 7. Output the facial emotion image with 2.2D height and 1.8D width. |
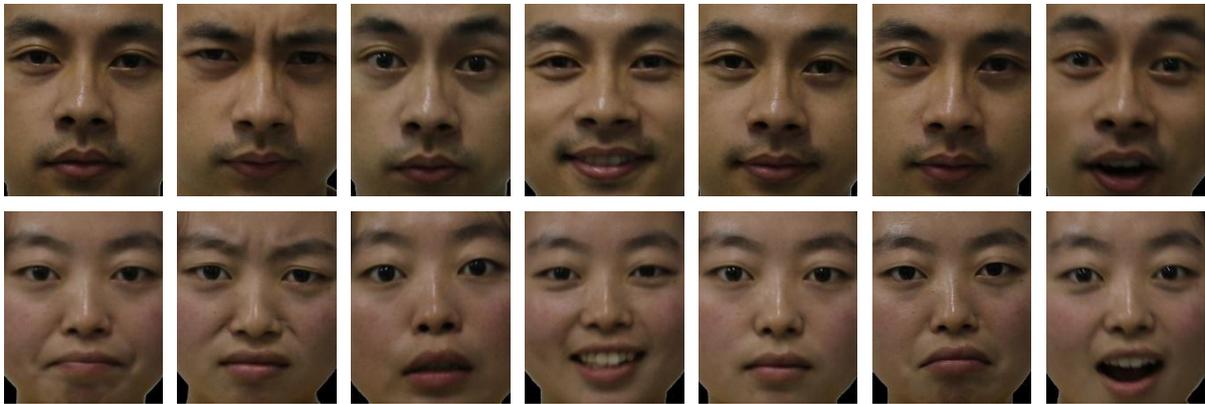
Figure 1. Samples of our dataset

3. Methodology

3.1. Stationary Wavelet Entropy

The concept of wavelet entropy (WE) is put forward by Rosso, et al. [10] when analyzing the electrical signals of the human brain. Among them, wavelet refers to the wave formed by the function in a distance. Entropy is a physical quantity used to describe the complexity or chaos of a system. The greater the entropy value, the greater the complexity or chaos. Entropy is often used as a reference in evaluating time-domain physiological signals and diagnosing diseases [11-13]. Traditional WE is obtained based on wavelet transform (WT). Since WT is sensitive to the change of signal details, it can extract the transient local detail features of non-static signal. Thus WE is often used for physiological signal analysis [14-16], such as extracting the texture features in facial emotional images. The calculation method of WE is:

$$E_h = \sum_{k=1}^{n} |m_{h(k)}|^2 \qquad (1)$$

$$T_h = \frac{\sum_{k=1}^{n} |m_{h(k)}|^2}{\sum_{h=1}^{l} \sum_{k=1}^{n} |m_{h(k)}|^2} \qquad (2)$$

where, k represents the index of state characteristics of a partially uncertain system. n represents the number of state characteristics of a partially uncertain system. h represents the index of wavelet decomposition level. l represents the number of wavelet decomposition levels. E_h represents the relative energy of the wavelet coefficient at the h scale, and T_h represents the distribution of different wavelet scales after the normalization of the original signal energy. Therefore, WE (W_E) can be calculated as follows.

$$W_E = - \sum_{h=1}^{l} T_h \cdot log(T_h) \qquad (3)$$

Spatial normalization of images is the premise for WE to have the above advantages in image feature extraction. Otherwise, when the image is slightly changed, the result of image recognition will be directly affected. In recent years, scholars have proposed SWE in response to this problem, which uses the stationary wavelet transform (SWT) instead of the WT. According to the experiments in [17-20], SWE is better than WE in hearing loss detection, gene expression and facial emotion recognition, etc. In this study, the decomposition level is set to 4. Some other advanced wavelet descriptors [21-25] will be used in our future studies.

3.2. Single-hidden-layer Feedforward Neural Network

According to the definition of neural network in Wikipedia, the full name of neural network is artificial neural network (ANN), which is a computational model that simulates the structure and function of the neural system of biological brain and is used to estimate or approximate the function. Neural network belongs to the field of machine learning and cognitive science. It has many neurons that can change their internal structure based on external information and acquiring the ability to learn. SHLFNN is one of the structures of artificial neural networks.

The structure of SHLFNN consists of three parts: the input layer, the hidden layer and the output layer. The connection mode between each layer of neurons is a full connection. Figure 2 is the diagram of SHLFNN network structure. According to the universal approximation theorem [26], SHLFNN has a strong approximation ability and can arbitrarily approximate the corresponding expected output. Since its simplicity and practicality, difficult nonlinear mapping can be accomplished with less cost [27-30]. Therefore, SHLFNN is one of the most important network models in the field of feedforward neural network and has been widely used. The output of SHLFNN can be expressed as [31]:

$$S(X) = \sum_{j=1}^{H} g_j \alpha \left(\sum_{q=1}^{p} w_{jq} x_q + b_j \right)$$

$$= \sum_{j=1}^{H} g_j \alpha \left(W_j^T \cdot X + b_j \right) \qquad (4)$$

where, $X = (x_1, x_2, x_3, \cdots, x_q)^T$ represents the input vector of q dimensions. g_j represents the output weight of the jth neuron in the hidden layer. $W_j = (w_{j1}, w_{j2}, w_{j3}, \cdots, w_{jq})^T$ represents the input weight of the jth neuron in the hidden layer. b_j represents the bias of the jth neuron in the hidden layer. α represents the activation function of each neuron in the hidden layer. $S(X)$ represents the output of SHLFNN. Deep learning methods [32-35] and transfer learning [36-39] are not used in our method, since our dataset is small.
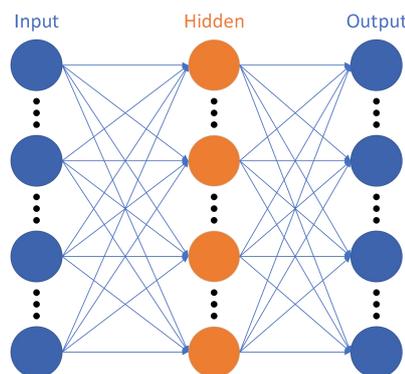

Figure 2. Diagram of a SHLFNN

3.3. Biogeography-based optimization

BBO algorithm comes from the theory of biogeography, which is a swarm intelligence optimization algorithm based on the general rules of migration and variation of different populations in different habitats [40-43].

Since the BBO algorithm was proposed, it has the characteristics of less parameters, stronger global search ability and faster convergence speed, so it is often used to solve the high-dimensional multi-objective optimization problems. The core of BBO algorithm is the population migration and mutation between different habitats, which are also the main steps to achieve the optimization effect in the process of solving the problem. In this paper, BBO is used to optimize the weight and bias parameters of

SHLFNN to further improve network performance.

BBO algorithm is a combination of both biogeography and engineering application, which mainly involves the terms as follows. Habitat, where species live and reproduce and represent a solution to a problem to be optimized. Habitat suitability index (HSI), the value of HSI corresponds to the number of existing species. The larger HSI, the larger number of species, indicating the higher fitness of the solution. And species will move from the habitat with higher value of HSI to the habitat with lower value of HSI to improve the habitability of the habitat with lower HSI value. Suitable index vector (SIV) represents factors such as rainfall, sunlight and area of the habitat, which corresponding to the variables of habitat.
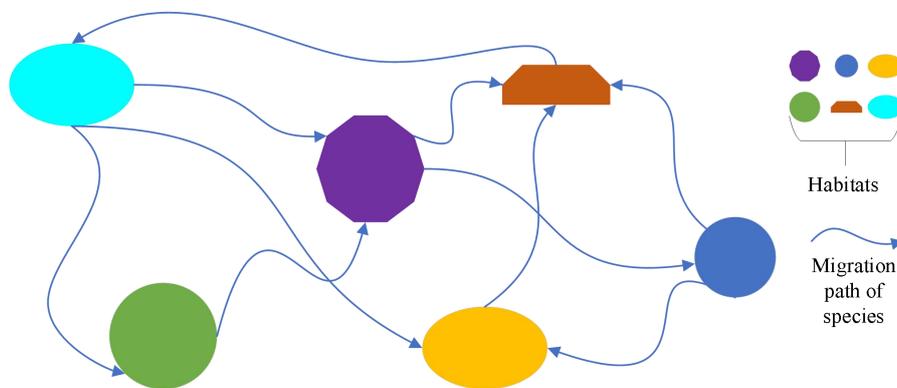
And better SIV could produce higher HSI. The basic idea of BBO is as follows: for a problem to be solved, some candidate schemes (i.e. biological habitats) are proposed. The score of the candidate schemes is quantified through the analysis of the adaptability of the candidate schemes, and the order is made according to the score size. According to the ranking of scores, the optimal solution of the problem is selected. Among them, the first candidate scheme is regarded as the optimal solution of the problem to be solved. At the same time, to better preserve the optimal solution, the algorithm usually introduces elitist preservation strategy.

Figure 3. Diagram of BBO habitat and migration

As shown in Figure 3, when we are faced with a problem to be optimized and a set of corresponding optimization candidates, every habitat in the diagram represents an optimization solution of the problem. At the same time, each habitat also corresponds to an HSI to measure the superiority of the habitat quality, which is the superiority of solution. The size of the HSI value depends on factors such as the climate, rainfall and vegetation diversity of the habitat, etc. Various factors describe the habitat through SIV. When solving the problem, BBO algorithm will simulate emergencies in some habitats, prompting a species to move from a higher HSI value habitat to a lower HSI value habitat. Subsequently, the information of population migration and mutation operation is shared among the habitats, so as to improve the HSI value of the habitats where mutation operation occurs and realize continuous optimization of the habitat

system [44-46].

Migration operation

The optimization effect of BBO algorithm mainly depends on migration operator, which is used to realize information sharing and interaction between different habitats. The HSI value corresponds to the number of species in the habitat. The higher the HSI value is, the higher the number of species in the habitat is, the higher the rate of emigration rate is and the lower the rate of immigration is. Conversely, the lower the HSI value is, the lower the number of species in the corresponding habitat is, the lower the rate of emigration is and the higher the rate of immigration is. Therefore, we can get the following formula.

$$\lambda_Z = L\left(1 - \frac{z}{z_{max}}\right) \qquad (5)$$

$$\mu_Z = \frac{U \cdot Z}{Z_{max}} \tag{6}$$

where, λ_Z represents the rate of immigration when the number of species is Z. L represents the maximum rate of immigration, $L \in (0,1]$. Z_{max} represents the maximum number of species. Z represents the index of the number of species. μ_Z represents the rate of emigration when the number of species is Z. U represents the maximum rate of emigration, $U \in (0,1]$. According to the above formula, we can obtain the linear model of population migration as shown in Figure 4.
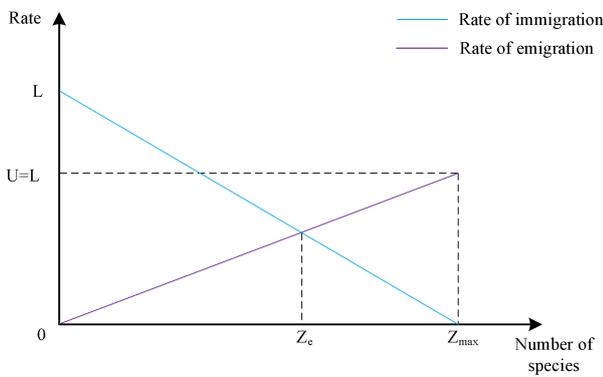

Figure 4. Linear model of population migration

Mutation operation

As mentioned above, BBO algorithm choose the habitat with low HSI value perform mutation operation by changing a certain SIV. And introduce elitist preservation strategy to preserve the best solution based on the elite parameters defined by us. Therefore, the mutation probability V of a habitat mainly depends on its species number probability M, which means the species number Z of a habitat is closely related to the mutation probability V. From this we can obtain the following formula:

$$V_Z = V_{max} \cdot \left(\frac{1-M_Z}{M_{max}}\right) \tag{7}$$

$$\ddot{M}_Z =$$
$$\begin{cases} -(\lambda_Z + \mu_Z)M_Z + \mu_{Z+1}M_{Z+1}, \ Z = 0 \\ -(\lambda_Z + \mu_Z)M_Z + \lambda_{Z-1}M_{Z-1} + \mu_{Z+1}M_{Z+1}, 1 \le Z \le Z_{max} - 1 \\ -(\lambda_Z + \mu_Z)M_Z + \lambda_{Z-1}M_{Z-1}, \ Z = Z_{max} \end{cases}$$
$$\tag{8}$$

where, V_Z represents the probability of mutation operation with the number of species Z. V_{max} represents the defined maximum rate of mutation. M_Z represents the species probability of the habitat with the number of species Z. $\ddot{M}_Z$ represents the updated value of M_Z. M_{max} represents the maximum probability of the species.

The detailed description of the BBO algorithm process is shown in Table 2. The overall flow diagram of BBO algorithm is shown in Figure 5.

Table 2. Algorithm process

Initial parameters: the number of habitats, the rate of mutation, the rate of immigration, the rate of emigration, maximum number of species, elitist parameters

for each habitat

    initialize SIV

end

do

    for each habitat

        calculate HSI values, sort them from high to low, and save them

        optimal habitat preservation

    end

    for each habitat

        calculate the rate of immigration and the rate of emigration

        perform the migration operation

        calculate the mutation rate

```
            perform mutation operation
        end
    while    the maximum number of iterations is not reached
            perform the iteration again
    else     satisfy the stop condition
            stop the iteration, output the result
```
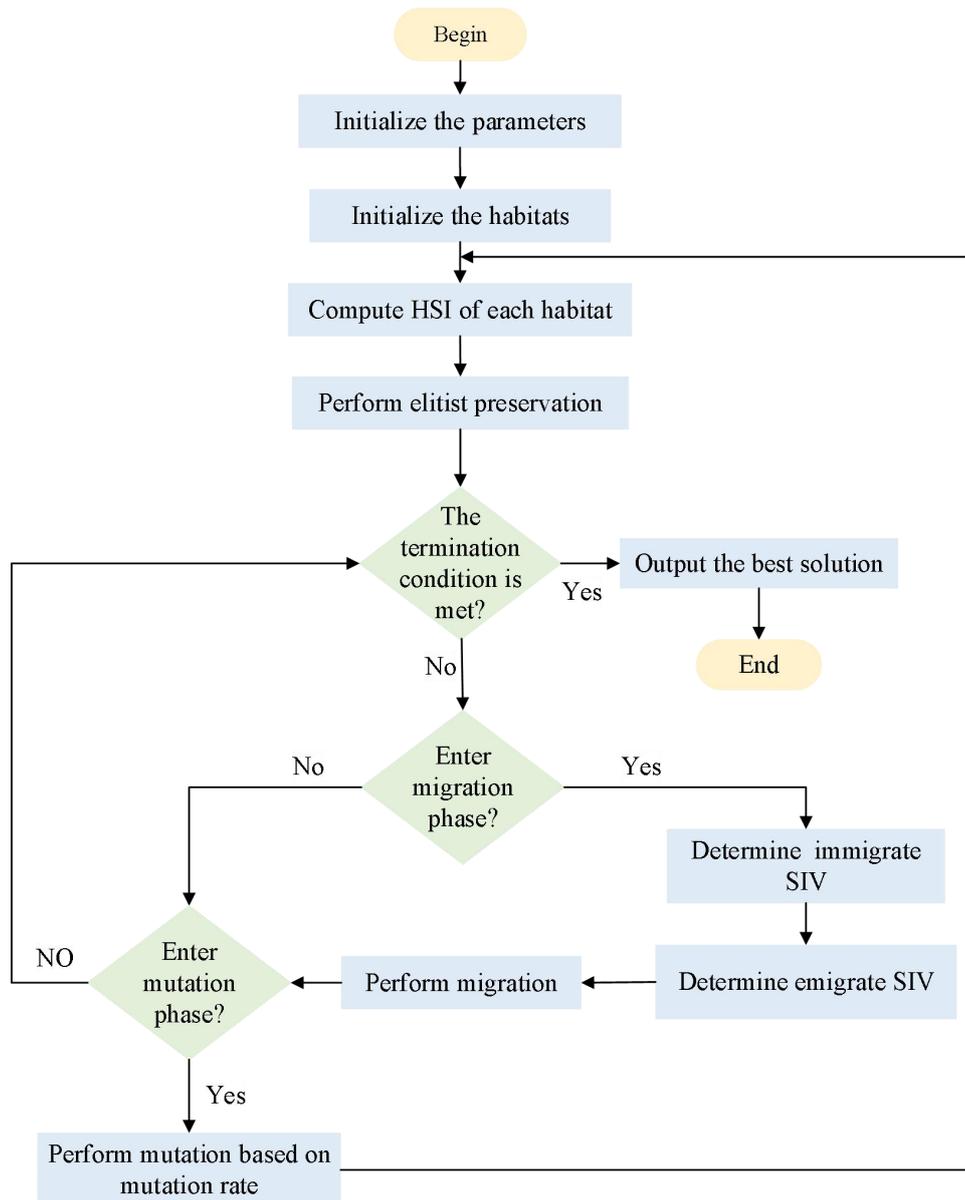

Figure 5. Diagram of BBO algorithm

3.4. Implementation

The overall flow of the proposed facial emotion recognition algorithm is shown in Figure 6. First, we built a face model, used SWE to extract the features in the target image. Then input the features into SHLFNN, and next trained and optimized SHLFNN by BBO. Finally performed the identification of output results.
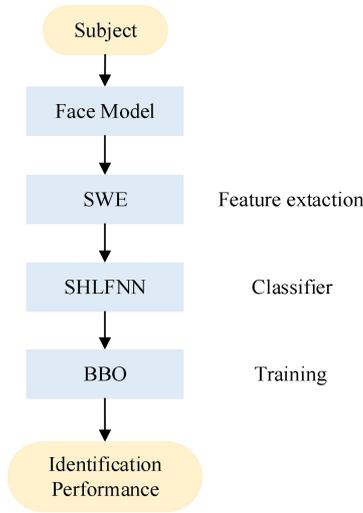
Figure 6. Block diagram of our developed system

In the execution process, BBO generates the corresponding output $y(i)$ of N habitats. We sort y from large to small, which means

$$Y(I)_{min} = Y(1) = y(i)_{max}, I, i = 1,2,3,\cdots,N \quad (9)$$

where Y represents the ranking number of the value of HSI. y represents the fitness of the habitats (candidate solutions). I represents the index of the ranking number. i represents the index of the habitats. N represents the number of habitats (the number of HSI values). As can be seen from Figure 4, the smaller $Y(I)$ is, the larger $y(i)$ it corresponds to, the smaller $\lambda(i)$ it corresponds to, and the larger $\mu(i)$ it corresponds to. When performing the immigration operation to the habitat i, compare the probability of $G(i)$ produced by the random function with that of $\lambda(i)$. If $G(i) < \lambda(i)$, perform the immigration operation. And the subsequent habitats do the same operation. Similarly, if $G(i) < \mu(i)$, then perform the emigration operation from all the groups that meet the criteria to provide data for the immigration operation. After that, perform the mutation operation. In each iteration, elitist preservation also saves the best individual. BBO can train SHLFNN in this way to optimize and improve the performance of the network.

3.5. Measure

In the experimental process, in order to avoid overfitting, we chose the 10-fold cross validation technique [47]. Each group contained 10 images of each seven emotions: happy,

sadness, fear, anger, surprise, disgust and neutral. In the performance of 10-fold cross validation, eight of these groups were used for training, one for validation, and the remaining one for testing. For a more concise representation, we introduced the confusion matrix (CM). Therefore, the ideal $C(r = 1, d = 1)$ should be as follows:

$$C(r = 1, d = 1) =$$

$$\begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 10 \end{bmatrix} \quad (10)$$

where, C is the confusion matrix, r is the number of runs, and d is the number of folds. The above matrix is the representation of 1 group of ideal confusion matrices in 1 iteration. Thus, the ideal $C(r = 1, d = 10)$ is:

$$C(r = 1, d = 10) =$$

$$\begin{bmatrix} 100 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 100 \end{bmatrix} \quad (11)$$

where, the elements on the diagonal of $C(r = 1, d = 10)$ are the structure of matrix summation for the test sets of 10 experiment groups. In general, to improve the accuracy of the experiment and reduce the random error, we performed 10 runs of 10-fold cross validation and summarized CM. Thus, the ideal $C(r = 10, d = 10)$ can be obtained as:

$$C(r = 10, d = 10) =$$

$$\begin{bmatrix} 1000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1000 \end{bmatrix}$$

$$(12)$$

For the sensitivity and overall accuracy (OA) of the network after the performance of $r = 10, d = 10$, we can obtain the following formula to define:

$$E(t) = \frac{C_{tt}(r=10, d=10)}{\sum_{u=1}^{7} C_{tu}(r=10, d=10)} \quad (13)$$

$$OA = \frac{\sum_{u=1}^{7} C_{uu}(r=10,d=10)}{\sum_{u=1}^{7} \sum_{v=1}^{7} C_{uv}(r=10,d=10)} \qquad (14)$$

where, $E(t)$ represents the sensitivity of class $t(t \in [1,7], t \in N^+)$, which means the t th element on the diagonal of confusion matrix divided by the sum of the tth row. OA represents the overall accuracy, which means taking the sum of the diagonal elements of confusion matrix divided by the sum of all elements of confusion matrix. C_{uv} represents class u recognized as class v in the confusion matrix.

4. Experiment Results and Discussions

4.1. Confusion matrix of proposed method

Table 3 is the confusion matrix presented by our model, which presents the experimental data of seven emotions: anger, disgust, fear, happy, neutral, sadness and surprise. Figure 7 is the bar chart corresponding to Table 3. It can be seen from Table 3 and Figure 7 that shows the accuracy of the whole model is 93.79%. The three emotions with the highest precision are: anger (94.9%), fear (94.8%), and sadness (94.1%). The three emotions with the highest specificity are: neutral (93.98%), anger (93.94%), and happiness (93.93%). The sensitivity analysis is shown in Table 4. The above data is closely related to the facial muscles that correspond to emotions, with lip funneler and nose wrinkles making their early facial emotions similar. And expressions such as jaw drop and the upper lid raiser also make emotions based on the same muscle movement characteristics look similar [48].

Table 3. Confusion matrix of our method

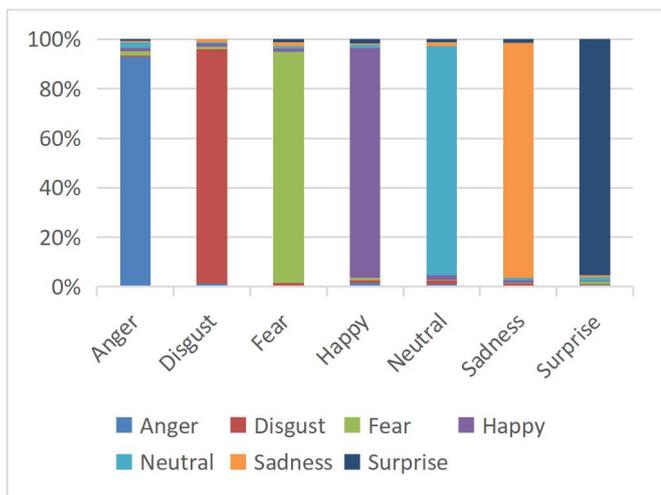| CM	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise
Anger	949	14	3	15	11	0	8
Disgust	6	937	14	12	15	14	2
Fear	17	11	948	9	3	1	11
Happy	14	12	15	929	18	10	2
Neutral	24	5	8	13	924	11	15
Sadness	4	12	17	5	14	941	7
Surprise	8	0	12	17	13	13	937

Figure 7. The corresponding bar chart in Table 3

4.2. Statistical results

Table 4 shows the sensitivity analysis of the seven emotions running 10 times. According to the data from Table 4 and Figure 8, the sensitivity of each emotion is as follows: 94.90±1.52% (anger), 93.70±2.00% (disgust), 94.80±0.92% (fear), 92.90±2.13% (happy), 92.40±2.22% (neutral), 94.10±1.52% (sadness), and 93.70±2.36% (surprise). From this we can get: the emotion of anger is the most sensitive and easy to recognize. Followed by the emotion of fear, and the third is the emotion of sadness. According to Table 5 and Figure 8, the overall average accuracy of the system after 10 runs is 93.79±1.24%.

Table 4. Statistical analysis on the sensitivities of each class

|	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise
Run 1	95.00	94.00	96.00	91.00	94.00	92.00	95.00
Run 2	97.00	96.00	94.00	92.00	94.00	95.00	94.00
Run 3	94.00	94.00	95.00	92.00	94.00	96.00	94.00
Run 4	95.00	93.00	96.00	96.00	94.00	95.00	95.00
Run 5	96.00	96.00	94.00	95.00	94.00	96.00	96.00
Run 6	94.00	89.00	95.00	92.00	93.00	93.00	93.00
Run 7	92.00	93.00	93.00	93.00	90.00	93.00	92.00
Run 8	95.00	93.00	95.00	94.00	90.00	92.00	94.00
Run 9	94.00	94.00	95.00	89.00	88.00	94.00	88.00
Run 10	97.00	95.00	95.00	95.00	93.00	95.00	96.00
Average	94.90± 1.52	93.70± 2.00	94.80± 0.92	92.90± 2.13	92.40± 2.22	94.10± 1.52	93.70± 2.36
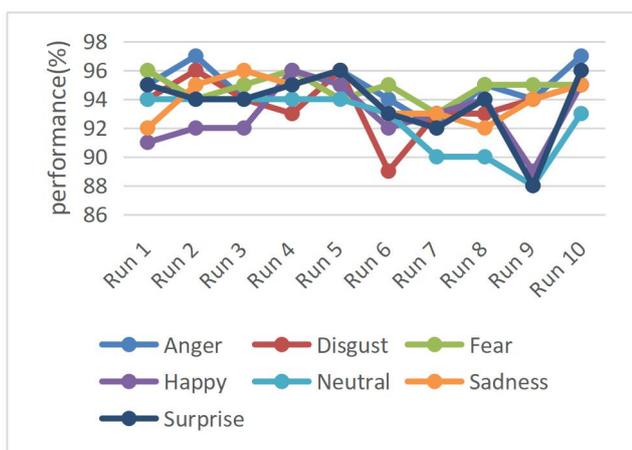
Figure 8. The trend of the sensitivities of each class
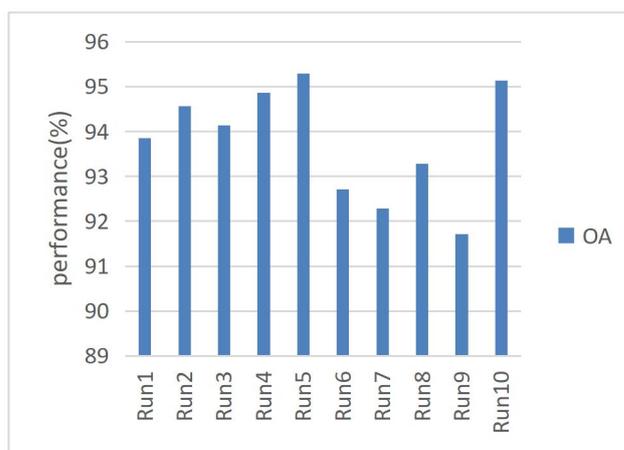

Figure 9. Overall accuracy comparison

4.3. Optimal Decomposition Level

We compared our result (level = 4) against other configurations (level= 1, 2, and 3). According to Table 6, we can see that different decomposition levels show different sensitivity to each emotion. When decomposition level = 1, the most sensitive and easily recognizable emotion is surprise, with the sensitivity of $89.20 \pm 2.86\%$. Followed by the emotion of neutral, with the sensitivity of $89.10 \pm 1.93\%$. And the third is the emotion of sadness, with the sensitivity of $89.00 \pm 2.75\%$. When decomposition level = 2, the most sensitive and easily recognizable emotion is happy, with the sensitivity of $90.40 \pm 2.27\%$.

Followed by the emotion of fear, with the sensitivity of $90.30 \pm 2.26\%$. And the third is the emotion of surprise, with the sensitivity of $90.20 \pm 2.10\%$. When decomposition level = 3, the most sensitive and easily recognizable emotion is sadness, with the sensitivity of $92.60 \pm 0.84\%$. Followed by the emotion of anger, with the sensitivity of $92.50 \pm 2.84\%$. And the third is the emotion of neutral, with the sensitivity of $92.40 \pm 1.90\%$. At the same time, we can also see that with the increase of decomposition levels, the average sensitivity of all emotions shows an upward trend, which indicates that in terms of recognition performance: L = 3> L = 2> L = 1.

Table 5. Statistical analysis on the overall accuracies

Run	1	2	3	4	5	6	7	8	9	10	Average
OA	93.86	94.57	94.14	94.86	95.29	92.71	92.29	93.29	91.71	95.14	93.79± 1.24

Table 6. Comparison with other decomposition levels (L = 1,2,3)

L = 1							
Run	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise	OA
Run 1	90.00	88.00	90.00	88.00	85.00	90.00	93.00	89.14
Run 2	89.00	89.00	90.00	88.00	91.00	92.00	84.00	89.00
Run 3	87.00	86.00	92.00	93.00	89.00	84.00	91.00	88.86
Run 4	89.00	89.00	90.00	89.00	90.00	92.00	91.00	90.00
Run 5	88.00	90.00	91.00	89.00	91.00	88.00	91.00	89.71
Run 6	87.00	89.00	91.00	90.00	88.00	89.00	85.00	88.43
Run 7	88.00	88.00	90.00	87.00	91.00	92.00	89.00	89.29
Run 8	91.00	93.00	84.00	86.00	87.00	87.00	91.00	88.43
Run 9	88.00	88.00	85.00	90.00	89.00	90.00	89.00	88.43
Run 10	92.00	87.00	86.00	88.00	90.00	86.00	88.00	88.14
Average	88.90± 1.66	88.70± 1.89	88.90± 2.81	88.80± 1.93	89.10± 1.97	89.00± 2.75	89.20± 2.86	88.94± 0.61

L = 2							
Run	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise	OA
Run 1	91.00	89.00	93.00	89.00	90.00	89.00	89.00	90.00
Run 2	92.00	92.00	89.00	94.00	89.00	91.00	90.00	91.00
Run 3	89.00	93.00	89.00	92.00	91.00	88.00	94.00	90.86
Run 4	93.00	92.00	91.00	92.00	90.00	94.00	90.00	91.71
Run 5	89.00	87.00	93.00	87.00	93.00	85.00	88.00	88.86

Run							
Run 6	90.00	88.00	89.00	92.00	89.00	89.00	92.00	89.86
Run 7	80.00	86.00	92.00	89.00	92.00	93.00	89.00	88.71
Run 8	92.00	89.00	89.00	88.00	90.00	93.00	93.00	90.57
Run 9	89.00	92.00	92.00	89.00	89.00	91.00	89.00	90.14
Run 10	93.00	91.00	86.00	92.00	87.00	88.00	88.00	89.29
Average	89.80± 3.79	89.90± 2.42	90.30± 2.26	90.40± 2.27	90.00± 1.70	90.10± 2.81	90.20± 2.10	90.10± 0.97

$L = 3$

Run	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise	OA
Run 1	91.00	95.00	94.00	91.00	93.00	92.00	94.00	92.86
Run 2	95.00	92.00	96.00	92.00	93.00	92.00	90.00	92.86
Run 3	87.00	89.00	90.00	91.00	92.00	94.00	90.00	90.43
Run 4	94.00	90.00	92.00	94.00	93.00	93.00	95.00	93.00
Run 5	89.00	93.00	92.00	91.00	93.00	91.00	89.00	91.14
Run 6	95.00	91.00	94.00	90.00	91.00	93.00	92.00	92.29
Run 7	96.00	93.00	94.00	95.00	94.00	93.00	94.00	94.14
Run 8	92.00	94.00	89.00	94.00	95.00	93.00	94.00	93.00
Run 9	93.00	93.00	89.00	95.00	92.00	92.00	93.00	92.43
Run 10	93.00	93.00	93.00	90.00	88.00	93.00	91.00	91.57
Average	92.50± 2.84	92.30± 1.83	92.30± 2.36	92.30± 2.00	92.40± 1.90	92.60± 0.84	92.20± 2.10	92.37± 1.07

A summarized comparison is shown in Table 7 and Figure 10. As can be seen from the Table 7, the overall accuracy of the model increases with the increase of decomposition level in the experimental range of decomposition level, which indicates that the higher the decomposition level is, the better the effect of emotion recognition is. At the same time, we can also see from Table 7 that when level = 4, the most sensitive and easily recognizable emotion is anger, with the sensitivity of $94.90 \pm 1.52\%$. Followed by the emotion of fear, with the sensitivity of $94.80 \pm 0.92\%$. And the third is the emotion of sadness, with the sensitivity of $94.10 \pm 1.52\%$. As expected, this conclusion is consistent with Table 3 and Table 4. In addition, Figure 10 shows this result more vividly.

Table 7. Select optimal decomposition level

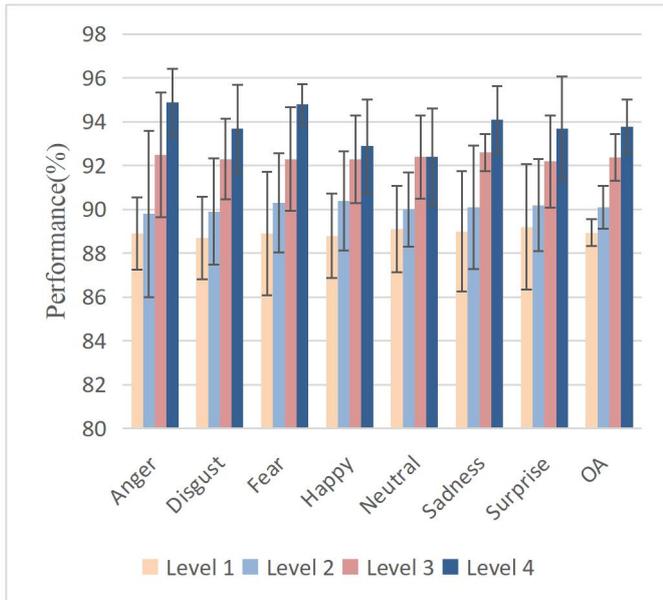| Level	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise	OA
1	88.90± 1.66	88.70± 1.89	88.90± 2.81	88.80± 1.93	89.10± 1.97	89.00± 2.75	89.20± 2.86	88.94± 0.61
2	89.80± 3.79	89.90± 2.42	90.30± 2.26	90.40± 2.27	90.00± 1.70	90.10± 2.81	90.20± 2.10	90.10± 0.97
3	92.50± 2.84	92.30± 1.83	92.30± 2.36	92.30± 2.00	92.40± 1.90	92.60± 0.84	92.20± 2.10	92.37± 1.07
4	94.90± 1.52	93.70± 2.00	94.80± 0.92	92.90± 2.13	92.40± 2.22	94.10± 1.52	93.70± 2.36	93.79± 1.24

Figure 10. Error bar of different decomposition level

4.4. Comparison to State-of-the-art Approaches

The OA of the "SWE+BBO" method used in this experiment is compared with that of the other three methods, which are SVM [2], HWT [3] and CSO [6]. The results are shown in Table 8. OA of SVM is $83.43 \pm 2.15\%$. OA of HWT is $78.37 \pm 1.50\%$. OA of CSO is $89.49 \pm 0.76\%$. We can clearly see that the method of "SWE+BBO" has the highest accuracy ($93.79 \pm 1.24\%$). Followed by CSO. And the third is SVM, while the lowest is HWT.

As can be seen from Table 3 and Table 4 that the highest OA obtained by "SWE+BBO" method mainly depends on: (i) the ability of SWE to extract image features. (ii) the classifier of SHLFNN. (iii) the excellent training ability of BBO. And the next best method is CSO, which mainly combines the seeking mode and the tracing mode in the algorithm through mixture ratio to achieve global optimization.

We should note that there are currently several variants of BBO algorithms, such as parallel hybrid BBO, alternated chaotic BBO and adaptive option-based BBO. In the future research, we will test their performances. And

we will also attempt to use deep learning approaches [49-53] to realize facial emotion recognition.

Table 8. Comparison with other methods

Approach	OA
SVM	83.43±2.15
HWT	78.37±1.50
CSO	89.49±0.76
SWE+BBO (Ours)	93.79± 1.24

5. Conclusions

In this paper, we proposed an improved facial emotion recognition system. We use static wavelet entropy for feature extraction and BBO algorithm to train single hidden layer feedforward neural network. The facial emotion recognition model has achieved a good recognition effect. However, there are still some aspects that can be explored in our method, such as the influence of different order of wavelet entropy on feature extraction.

In the future research, we will continue to focus on the research of facial emotion recognition and try to collect more emotional images than in this paper. At the same time, we plan to put forward a better algorithm to optimize the hyperparameters of single hidden layer feedforward neural network, such as weights and biases. And we will also try such optimization algorithms based on evolution and swarm-intelligence as gray wolf optimization and ant colony optimization to improve the performance of single hidden layer feedforward neural network.

References

[1] P. Ekman and W. V. Friesen, "Facial Action Coding System (FACS): A technique for the measurement of facial action," *Rivista Di Psichiatria*, vol. 47, no. 2, pp. 126-38, 1978.

[2] H. Ali, M. Hariharan, S. Yaacob, and A. H. Adom, "Facial Emotion Recognition Based on Higher-Order Spectra Using Support Vector Machines," (in English), *Journal Of Medical Imaging And Health Informatics*, Article; Proceedings Paper vol. 5, no. 6, pp. 1272-1277, Oct 2015.

[3] F. Evans, "Haar Wavelet Transform Based Facial Emotion Recognition," *Advances in Computer Science Research,* vol. 61, pp. 342-346, 2017/03 2017.

[4] L. Ivanovsky, V. Khryashchev, A. Lebedev, and I. Kosterin, "Facial expression recognition algorithm based on deep convolution neural network," 2018.

[5] B. Hasani and M. H. Mahoor, "Spatio-Temporal Facial Expression Recognition Using Convolutional Neural Networks and Conditional Random Fields."

[6] W. Yang, "Facial Emotion Recognition via Discrete Wavelet Transform , Principal Component Analysis, and Cat Swarm Optimization," *Lecture Notes in Computer Science,* vol. 10559, pp. 203-214, 2017.

[7] N. Lucy, W. Hui, L. Jiang, U. Ishaq, Y. Xiaokun, and Z. Ting, "Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts," in *Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts*, 2017.

[8] H. M. Lu, "Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Vector Machine, and Stratified Cross Validation," *IEEE Access,* vol. 4, pp. 8375-8385, 2016.

[9] F. Y. Shih and C. Chao-Fa, "Automatic extraction of head and face boundaries and facial features," vol. 158, no. none, pp. 117-130.

[10] O. A. Rosso *et al.*, "Wavelet entropy: a new tool for analysis of short duration brain electrical signals," *J Neurosci Methods,* vol. 105, no. 1, pp. 65-75.

[11] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology Heart & Circulatory Physiology,* vol. 278, no. 6, pp. H2039-H2049.

[12] X.-X. Zhou, "Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine," *Simulation,* vol. 92, no. 9, pp. 861-871, 2016.

[13] D. R. Nayak, "Detection of unilateral hearing loss by Stationary Wavelet Entropy," *CNS & Neurological Disorders - Drug Targets,* vol. 16, no. 2, pp. 15-24, 2017.

[14] R. Alcaraz and J. J. Rieta, "Application of Wavelet Entropy to Predict Atrial Fibrillation Progression from the Surface ECG," *Computational & Mathematical Methods in Medicine,* vol. 2012, pp. 1-9.

[15] P. Phillips, "Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm," *Neurocomputing,* vol. 272, pp. 668-676, 2018.

[16] J. M. Gorriz, "Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization," *Journal of Alzheimer's Disease,* vol. 65, no. 3, pp. 855-869, 2018.

[17] N. Nguyen, A. Vo, I. Choi, and K.-J. Won, "- A Stationary Wavelet Entropy-Based Clustering Approach Accurately Predicts Gene Expression," vol. - 22, no. - 3.

[18] Y. Huo, "Feature Extraction of Brain MRI by Stationary Wavelet Transform and its Applications," *Journal of Biological Systems,* vol. 18, no. S, pp. 115-132, 2010.

[19] A. Liu, "Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine," *Journal of Medical Imaging and Health Informatics,* Research Article vol. 5, no. 7, pp. 1395-1403, 2015.

[20] A. Atangana, "Application of stationary wavelet entropy in pathological brain detection," *Multimedia Tools and Applications,* journal article vol. 77, no. 3, pp. 3701-3714, 2018.

[21] O. Hellel, M. Beladgham, and A. M. Lakhdar, "Study of performance of a 'second-generation wavelet video encoder with a scalable rate'," (in English), *IET Image Processing,* Article vol. 13, no. 4, pp. 698-706, Mar 2019.

[22] J. Yang, "Identification of green, Oolong and black teas in China via wavelet packet entropy and fuzzy support vector machine," *Entropy,* Article vol. 17, no. 10, pp. 6663-6682, 2015.

[23] J. Q. Yang, J. F. Yang, A. J. Liu, and P. Sun, "A Novel Compressed Sensing Method for Magnetic Resonance Imaging: Exponential Wavelet Iterative Shrinkage-Thresholding Algorithm with Random Shift," (in English), *International Journal Of Biomedical Imaging,* Article 2016, Art. no. 9416435.

[24] M. Yang, "Dual-Tree Complex Wavelet Transform and Twin Support Vector Machine for Pathological Brain Detection," *Applied Sciences,* vol. 6, no. 6, 2016, Art. no. 169.

[25] J. M. Gorriz and J. Ramírez, "Wavelet entropy and directed acyclic graph support vector machine for detection of patients with unilateral hearing loss in MRI scanning," *Frontiers in Computational Neuroscience,* vol. 10, 2016, Art. no. 160.

[26] S. Odense and R. Edwards, "Universal Approximation Results for the Temporal Restricted Boltzmann Machine and the Recurrent Temporal Restricted Boltzmann Machine," 2016.

[27] L. N. Wu, "Improved image filter based on SPCNN," (in English), *Science In China Series F-Information Sciences,* Article vol. 51, no. 12, pp. 2115-2125, Dec 2008.

[28] L. Wu, "Weights optimization of neural network via improved BCO approach," *Progress in Electromagnetics Research,* vol. 83, pp. 185-198, 2008.

[29] G. Wei, "A New Classifier for Polarimetric SAR Images," *Progress in Electromagnetics Research,* vol. 94, pp. 83-104, 2009.

[30] N. Naggaz, "Remote-sensing Image Classification Based on an Improved Probabilistic Neural Network," *Sensors,* vol. 9, no. 9, pp. 7516-7539, 2009.

[31] S. Chakraverty and S. Mall, "Single layer Chebyshev neural network model with regression-based weights for solving nonlinear ordinary differential equations," (in English), *Evolutionary Intelligence,* Article; Early Access p. 8.

[32] C. Pan, "Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling," *Journal of Computational Science,* vol. 27, pp. 57-68, 2018.

[33] C. Tang, "Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform," *Multimedia Tools and Applications,* vol. 77, no. 17, pp. 22821-22839, 2018.

[34] X. Jiang and L. Chang, "Classification of Alzheimer's Disease via Eight-Layer Convolutional Neural Network with Batch Normalization and Dropout Techniques," *Journal of Medical Imaging and Health Informatics,* vol. 10, no. 5, pp. 1040-1048, 2020.

[35] X. Jiang, "Chinese Sign Language Fingerspelling Recognition via Six-Layer Convolutional Neural Network with Leaky Rectified Linear Units for Therapy and Rehabilitation," *Journal of Medical Imaging and Health Informatics,* vol. 9, no. 9, pp. 2031-2038, 2019.

[36] S. Lu, "Pathological Brain Detection based on AlexNet and Transfer Learning," *Journal of Computational Science,* vol. 30, pp. 41-47, 2019.

[37] J. Hong, "Detecting cerebral microbleeds with transfer learning," *Machine Vision and Applications,* vol. 30, no. 7-8, pp. 1123–1133, 2019.

[38] X. Yu, "Utilization of DenseNet201 for diagnosis of breast abnormality," *Machine Vision and Applications,* vol. 30, no. 7-8, pp. 1135-1144, 2019/10/01 2019.

[39] V. V. Govindaraj, "High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model," *Journal of Medical Imaging and Health Informatics,* vol. 9, no. 9, pp. 2012-2021, 2019.

[40] A. K. Sangaiah, G. B. Bian, S. M. Bozorgi, M. Y. Suraki, A. A. R. Hosseinabadi, and M. B. Shareh, "A novel quality-of-service-aware web services composition using biogeography-based optimization algorithm," (in English), *Soft Computing,* Article vol. 24, no. 11, pp. 8125-8137, Jun 2020.

[41] S. Wang *et al.*, "A new logistic map based chaotic biogeography-based optimization approach for cluster analysis," in *International Conference on Progress in Informatics and Computing (PIC)*, Shanghai, China, 2016, pp. 88-92: IEEE.

[42] G. Yang, "Automated classification of brain images using wavelet-energy and biogeography-based optimization," (in English), *Multimedia Tools and Applications,* vol. 75, no. 23, pp. 15601-15617, May 2016.

[43] L. Wei, "Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization," *Entropy,* vol. 17, no. 8, pp. 5711-5728, 2015.

[44] Y.-J. Li, "Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization," *Multimedia Tools and Applications,* vol. 77, no. 9, pp. 10393-10417, 2018.

[45] P. Li and G. Liu, "Pathological Brain Detection via Wavelet Packet Tsallis Entropy and Real-Coded Biogeography-based Optimization," *Fundamenta Informaticae,* vol. 151, no. 1-4, pp. 275-291, 2017.

[46] X. Wu, "Smart detection on abnormal breasts in digital mammography based on contrast-limited adaptive histogram equalization and chaotic adaptive real-coded biogeography-based optimization," *Simulation,* vol. 92, no. 9, pp. 873-885, September 12, 2016 2016.

[47] P. Tomczak and H. Puszkarski, "Interpretation of ferromagnetic resonance experimental results by cross-validation of solutions of the Smit-Beljers equation," (in English), *Journal of Magnetism and Magnetic Materials,* Article vol. 507, p. 7, Aug 2020, Art. no. 166824.

[48] R. E. Jack, O. G. B. Garrod, and P. G. J. C. B. Schyns, "Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time," vol. 24, no. 2, pp. 187-192.

[49] Y. Chen, "Cerebral micro‑bleeding identification based on a nine‑layer convolutional neural network with stochastic pooling,"

Concurrency and Computation: Practice and Experience, vol. 31, no. 1, p. e5130, 2020.

[50] A. K. Sangaiah, "Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization," *Neural Computing and Applications,* vol. 32, pp. 665-680, 2020.

[51] X. Jiang, "Fingerspelling identification for Chinese sign language via AlexNet-based transfer learning and Adam optimizer," *Scientific Programming,* vol. 2020, 2020, Art. no. 3291426.

[52] Z. Rezaei, H. Ebrahimpour-Komleh, B. Eslami, R. Chavoshinejad, and M. Totonchi, "Adverse Drug Reaction Detection in Social Media by Deep Learning Methods," (in English), *Cell Journal,* Article vol. 22, no. 3, pp. 319-324, Fal 2020.

[53] X. Jiang, S. C. Satapathy, L. Yang, S.-H. Wang, and Y.-D. Zhang, "A Survey on Artificial Intelligence in Chinese Sign Language Recognition," *Arabian Journal for Science and Engineering,* 2020/07/22 2020.