

A Hybrid Classification Method of Medical Image Based on Deep Learning

Yulong Wang¹ · Xiaofeng Liao² · Dewen Qiao¹ · Jiahui Wu¹

Abstract With the rapid development of modern medical science and technology, medical image classification has become a more and more challenging problem. However, in most traditional classification methods, image feature extraction is difficult, and the accuracy of classifier needs to be improved. Therefore, this paper proposes a high-accuracy medical image classification method based on deep learning, which is called hybrid CQ-SVM. Specifically, we combine the advantages of convolutional neural network (CNN) and support vector machine (SVM), and integrate the novel hybrid model. In our scheme, quantum-behaved particle swarm optimization algorithm (QPSO) is adopted to set its parameters automatically for solving the SVM parameter setting problem, CNN works as a trainable feature extractor and SVM optimized by QPSO performs as a trainable classifier. This method can automatically extract features from original medical images and generate predictions. The experimental results show that this method can extract better medical image features, and achieve higher classification accuracy.

Keywords Medical image classification · Convolutional neural network (CNN) · Quantum-behaved particle swarm optimization (QPSO) · Support vector machine (SVM) · Feature extraction

Yulong Wang
E-mail: ahwangedu@163.com

Xiaofeng Liao (Corresponding author)
E-mail: xfliao@cqu.edu.cn

Present address:

¹College of Electronic and Information Engineering, Southwest University, Chongqing, 400715, P.R.China

²College of Computer Science, Chongqing University, Chongqing 400044, P.R.China

1 INTRODUCTION

With the rapid development of medical technology and the continuous emergence of various medical devices, more and more medical images appear in clinical diagnosis. Therefore, medical image analysis has been widely used, and how to effectively and accurately classify these image information has become an urgent problem.

In recent decades, in order to solve the problem of classification, many effective methods which involved in statistical learning [1], pattern recognition [2], computer vision [3] and signal processing [4] have been proposed. Medical image classification and recognition are important research direction of pattern recognition, which has been widely concerned by researchers [5–9]. In many classification methods, support vector machine (SVM) [10] stands out and becomes one of the most popular methods, which is because SVM model has strong generalization ability. Different from the previous classification methods based on empirical risk minimization, SVM is a new machine learning method based on statistical learning theory, which has strong nonlinear processing ability. In recent years, SVM has become a standard tool in the field of machine learning and is increasingly being applied to pattern recognition, data mining, intelligent control, prediction, etc [11–14]. However, SVM still has some disadvantages because of the dubious determination of the parameters, such as penalty parameter C and kernel parameter σ . Sometimes, from parameters optimization to model selection, SVM may fall into an over-fitting problem [15]. Therefore, when applying SVM to solve the classification problem, the parameters including the penalty parameter C and the kernel parameter σ directly affect the final classification accuracy.

For a long time, researchers have done a lot of research on the parameters optimization of SVM. For

example, a grid search (GS) method based on cross validation (CV) idea to find optimal parameters, Gao et al. predicted the status of the Tennessee Eastman (TE) Process by optimized SVM model [16]. Sopyla et al. proposed to apply Barzilai Borwein update steps in the classical SGD algorithm to solve the optimization of the SVM model [17]. The partial derivatives proposed by Rubio et al. [18] are used to improve the procedure for parameter optimization of LS-SVM. However, these methods are easy to fall into local optimum when optimizing the parameters of SVM model, and result from certain difficulties to the improvement of SVM performance. In recent years, with the development of machine learning, evolutionary algorithms (EA) such as genetic algorithm [19], particle swarm optimization algorithm (PSO) [20], and quantum-behaved particle swarm optimization algorithm (QPSO) [21] have emerged. These algorithms are improved on the basis of non-derivative learning methods, which makes the optimization method much more flexible. Among them, QPSO is a swarm-based stochastic optimization technology. Compared with standard PSO, QPSO has more global search ability and less control parameters. Therefore, we employ QPSO to solve SVM parameters optimization problem.

Deep learning method is a new technology in the field of pattern recognition, which imitates the automatic learning process of human brain. It is developed by researchers in the early neural network and has been successfully applied to various types of image processing tasks [22–26]. In recent years, due to the breakthrough of deep learning method in image processing [27], the application of deep learning method in the field of medical image classification has also begun [28–30]. Hua et al. [31] preliminarily discussed the application of deep belief network model and deep convolution neural network to the classification of solitary pulmonary nodule. David et al. [32] used image processing technology and image classifier based on artificial neural network to analyze the retinal image automatically, and classified the image according to the disease situation. Ramirez et al. [33] applied artificial neural network to brain images to analyze the condition of Alzheimer’s patients. Derya Avci et al. [34] proposed a new approach based on adaptive wavelet entropy energy and neural network classifier for urine cell recognition. Among them, convolutional neural network (CNN), as a classical learning model in deep learning, has been widely used. CNN has the ability of representation learning, which can do shift-invariant classification for input information according to its hierarchical structure. In addition, the rapid development of modern computers provides strong computational support for CNN training. Through thou-

sands or even more training, the parameters in the network are constantly updated, and finally the convergence state is arrived. Traditional medical image classification methods mostly need to extract features such as brightness, shape, gradient, gray level and texture, and select relevant features. The process is complicated, while the deep learning method can automatically train the image and extract multi-level features, and does not need complex image processing steps, so the process is simple. However, as far as we know, in many literatures, few people use CNN as a feature extraction method. In our work, CNN acts as a feature extractor to extract the features of the original image, which can avoid the complex image processing steps in the process of feature extraction and retain the important features of the image. And compared with the current popular feature extraction method of principal component analysis (PCA) [35].

On the basis of the previous studies, this paper proposes a new medical image classification method, which combines CNN, QPSO and SVM, named hybrid CQ-SVM model. Specifically, as a trainable feature extractor, CNN extracts the important features of the original image. SVM optimized by QPSO is used as a trainable output classifier for image classification and prediction. In order to evaluate the performance of the hybrid CQ-SVM model, we carry out experiments based on the image of breast cancer cells, and confirm the validity of the model.

In summary, the main contributions of this paper are as follows.

(1) We use CNN to design a trainable feature extractor, which can automatically extract the advanced features of the original image, making the image distinguishable.

(2) We use QPSO to optimize SVM, design a trainable high-precision output classifier, and do a lot of comparative experiments to verify its effectiveness.

(3) Combining the advantages of CNN and SVM, we propose a new hybrid CQ-SVM model to improve the accuracy of medical image classification.

The rest of this paper is organized as follows. Section 2 introduces some preliminary knowledge. Section 3 constructs the hybrid CQ-SVM model. Analysis for the experiments and results are provided in Section 4. Section 5 gives the conclusions and future research direction.

2 PRELIMINARIES

In this section, we introduce CNN model, QPSO algorithm and SVM model in turn.

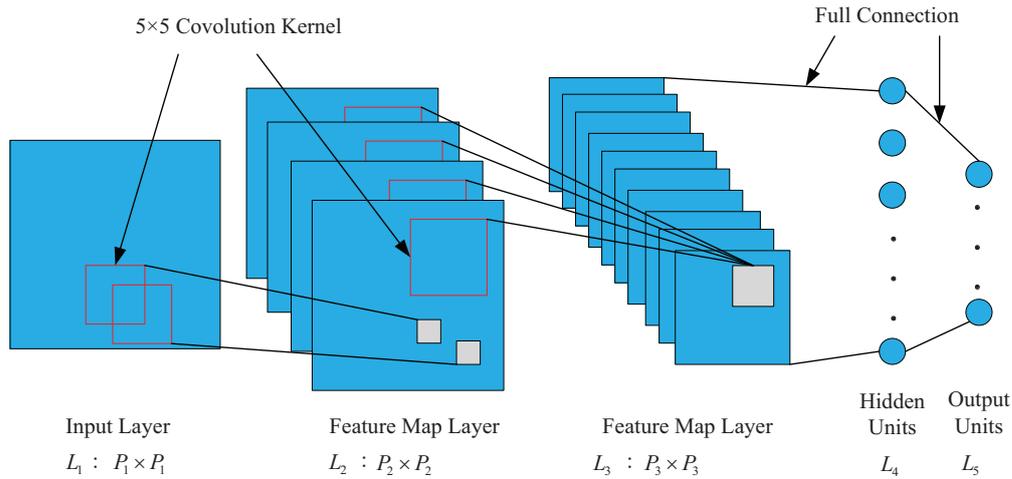


Fig. 1: Structure of the adopted CNN.

2.1 CNN Model

CNN is a kind of multi-layer neural network with deep supervised learning structure [36]. It can be regarded as two parts: an automatic feature extractor and a trainable classifier. The feature extractor consists of several feature mapping layers, which can extract data features with high recognition rate by convolutional filtering and down sampling. The trainable classifier is usually a fully connected multi-layer perceptron. The back-propagation algorithm is used to train the weight and bias in the feature extractor.

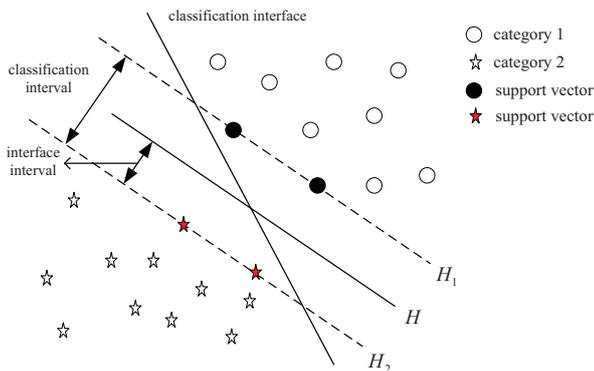


Fig. 2: Diagram of the optimal classification surface.

Fig. 1 presents an example of CNN structure, which consists of five layers. The input layer L_1 is a matrix with the size of $P_1 \times P_2$. The original image will be transferred to the input layer after preprocessing. L_2 and L_3 are feature mapping layers, which are used to calculate data features. Among them, the size of the convolutional filter kernel is 5×5 , and the down sam-

pling operation ratio after filtering is 2. Each neuron on the layer is connected to the previous layer, and each neuron has 25 inputs in total. All neurons in a feature mapping layer share the same kernel and connection weight, which is called “weight sharing”. Each feature map layer reduces the feature size from the previous feature size P to $\lceil (P-4)/2 \rceil$. As shown in Fig. 1, $P_2 = \lceil (P_1 - 4)/2 \rceil$, and $P_3 = \lceil (P_2 - 4)/2 \rceil$, where $\lceil x \rceil$ denotes the largest integer not exceeding x . L_4 and L_5 are hidden layers, which together form a fully connected multi-layer perceptron. In our experiments, we use a more complex CNN network named LeNet-5 [37], and use the optimized SVM classifier as the trainable classifier to replace the multi-layer perceptron. This process will be described in detail in section 3.

2.2 QPSO Algorithm

QPSO algorithm is an improvement of PSO algorithm. Its main goal is to search the position close to the optimal solution in the search space. In the QPSO algorithm, each particle is a point in the search space, which represents a solution of the problem in the d -dimensional space. The number of particles N is determined by the user. The positions of all particles are randomly initialized. The position of the i th particle in the k th iteration is $\mathbf{X}^{(k)}(i) = [\mathbf{X}^{(k)}(i, 1), \dots, \mathbf{X}^{(k)}(i, d)]$. $\mathbf{P}^{(k)}(i) = [\mathbf{P}^{(k)}(i, 1), \dots, \mathbf{P}^{(k)}(i, d)]$ represents the current optimal position of the particle. $\mathbf{P}^{(k)}(g)$ is called the global optimal position, which represents the position of particles with the best fitness value. The update

equation of particle position is as follows:

$$\begin{cases} \mathbf{X}^{(k+1)}(i) = \mathbf{p}(i) + \alpha \cdot \left| \mathbf{M}_{best} - \mathbf{X}^{(k)}(i) \right| \cdot \ln\left(\frac{1}{\theta}\right) & \text{if } t \geq 0.5, \\ \mathbf{X}^{(k+1)}(i) = \mathbf{p}(i) - \alpha \cdot \left| \mathbf{M}_{best} - \mathbf{X}^{(k)}(i) \right| \cdot \ln\left(\frac{1}{\theta}\right) & \text{if } t < 0.5, \end{cases} \quad (1)$$

where \mathbf{M}_{best} is the average of $\mathbf{P}(i)$, and represents the best position in the average particle history. It is calculated as the following equation.

$$\begin{aligned} \mathbf{M}_{best} &= \frac{1}{N} \sum_{i=1}^N \mathbf{P}^{(k)}(i) \\ &= \left(\frac{1}{N} \sum_{i=1}^N \mathbf{P}^{(k)}(i, 1), \dots, \frac{1}{N} \sum_{i=1}^N \mathbf{P}^{(k)}(i, d) \right), \end{aligned} \quad (2)$$

$\mathbf{p}(i)$ is a design parameter, called a local attractor. It is an important parameter for the convergence of QPSO algorithm. It is calculated as follows:

$$\mathbf{p}(i) = \lambda \mathbf{P}^{(k)}(i) + (1 - \lambda) \mathbf{P}^{(k)}(g). \quad (3)$$

In Eq. (1), α is called updated parameter, which is used to control the convergence rate of the algorithm. Its calculation formula is as Eq. (4). α_0 is the initial value of α , and α_1 is the final value of α . k_{\max} is the maximum number of iterations. In our experiments, we set these two values to 0.5 and 0.9, respectively [38].

$$\alpha = \alpha_0 + (k_{\max} - k) \cdot (\alpha_1 - \alpha_0) / k_{\max}. \quad (4)$$

In the above formulas, θ , t and λ are the uniformly generated random numbers in the range of [0,1].

2.3 SVM Model

As shown in Fig. 2, the main idea of SVM is to optimize the linear separable classification plane. The ultimate goal is not only to reduce the sample classification error, but also to maximize the sample classification interval. H is the optimal hyperplane. H_1 and H_2 are class interfaces.

(1) **Mathematical Modeling of SVM:** Suppose there is a training sample set $\{(\mathbf{x}_i, y_i), i = 1, 2, \dots, N\}$ of size N . It consists of two categories, $y \in \{-1, 1\}$. The goal of the SVM classifier is to find an optimal hyperplane $f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b$ in the sample space, so that the hyperplane can divide the samples into two categories correctly. The optimization problems to be solved by SVM are as follows:

$$\begin{cases} \min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i, \\ \text{s.t. } y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \xi_i > 0, \end{cases} \quad (5)$$

where \mathbf{w} is an n-dimensional vector known as weight or support vector, and b is scalar called bias. ξ_i is a relaxation variable, which is used to solve the approximate separable problem in SVM. C is a regularization parameter (also called penalty factor), which controls the penalty degree of wrong samples, and achieves the trade-off between the ratio of wrong samples and algorithm complexity.

(2) **Kernel Function:** The function of kernel is to transform the completely indivisible problem into the separable problem or to reach the approximately separable state. Choosing proper kernel function is very important for constructing SVM model with good classification performance. The common kernel functions are linear kernel function, polynomial kernel function, Gaussian radial basis function and Sigmoid kernel function. In this paper, we choose radial basis function (RBF) as the kernel function of SVM. The specific expression is as follows:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right), \quad (6)$$

where σ is a kernel parameter, which is related to the number of support vectors and the determination of training time.

3 HYBRID CQ-SVM MODEL

In this section, we introduce the proposed hybrid CQ-SVM model. The model is designed with reference to CNN classification model and SVM classification model. First, the hybrid model based on CNN is described, and then QPSO is used to optimize SVM in the hybrid model based on CNN to build its corresponding hybrid model.

3.1 Overview of Hybrid CQ-SVM Model

The design idea of the hybrid CQ-SVM classifier model proposed by us comes from the advantages of CNN and SVM [39]. CNN has a strong learning ability, and can automatically learn the important features of images. It does not need complex image processing steps. SVM has powerful generalization ability. The two combinations will undoubtedly form a better model. The architecture of CQ-SVM model is designed by replacing the last output layer of CNN model with SVM classifier. Moreover, in order to further improve the classification performance of SVM, QPSO is used to optimize its parameters. For the output units in the last layer of CNN network, they are the estimated probabilities of input

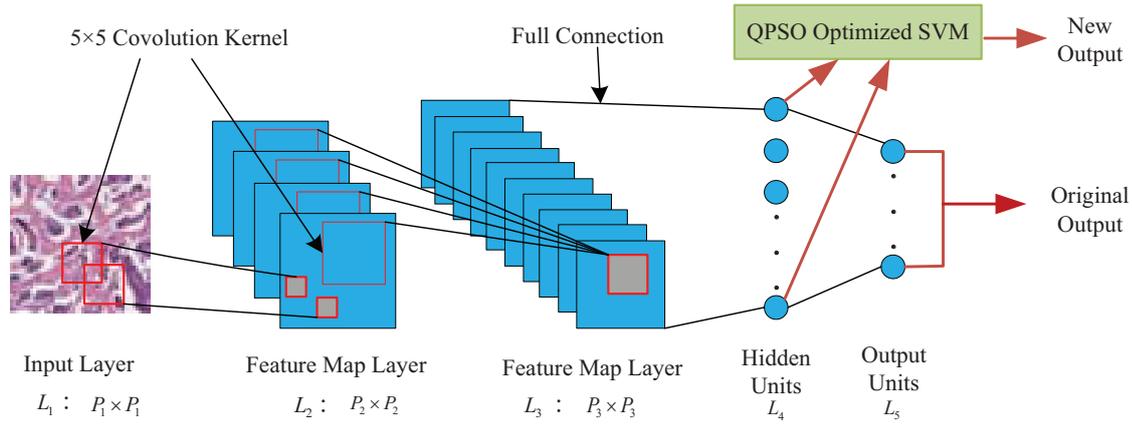


Fig. 3: Structure of the hybrid CQ-SVM model.

samples. Each output probability is calculated by the activation function. The input of the activation function is a linear combination of the output of the previous hidden layer and the weight, plus a bias term. It is meaningless to see the output value of the hidden layer. It is only meaningful to the network itself. However, these output values can be used as the feature data of the original image, so they can be used as the input of SVM classifier.

The structure of the hybrid CQ-SVM model is shown in Fig. 3. Firstly, the preprocessed image data should be inputted to the original CNN for network structure training, and the network weight and bias parameters should be updated until the training process converges. What should be emphasized here is that the CNN used in our experiment is LeNet-5 network structure. Then, the trained CNN network is used as the feature extractor of CQ-SVM model. After CNN feature extraction, SVM optimized by QPSO takes the output of hidden layer (L_4) as the new feature vector of training. Finally, the SVM classifier gets a good optimization and training. It will perform the recognition task, and make a new prediction for the image with this automatic feature extraction.

3.2 Optimizing SVM with QPSO

Fig. 4 shows the process of optimizing SVM with QPSO. First, the parameters of QPSO algorithm are initialized, including the number of particles, the maximum number of iterations and the initial position of particles. C and σ are also initialized and provided to SVM. Then, SVM uses the training set for model training, and uses the test set to test the trained model. The red part in the figure represents the training and testing process of the model. Finally, the test results of the

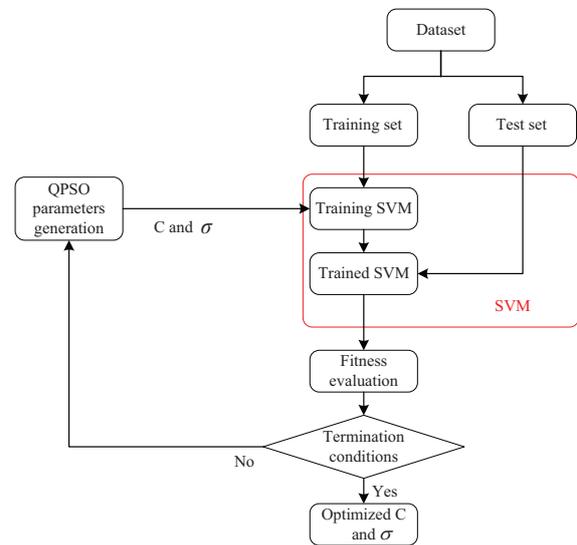


Fig. 4: Flowchart of the QPSO-SVM model.

model are used as model fitness evaluation. At the end of each iteration, condition judgment will be carried out. The parameter optimization is executing until the termination of the model training, then C and σ are obtained as the results.

In this paper, QPSO is used to search for the best combination of these two parameters. The specific steps are as follows:

Step 1: Code parameters to be determined, $\mathbf{x} = \{C, \sigma\}$.

Step 2: Establish training sample set and test sample set.

Step 3: Determine the structure of SVM and the choice of kernel function. The RBF function is chosen in our experiments.

Step 4: Calculate the optimal location of particles and particle swarm.

Step 5: Update the particle position based on the update formula.

Step 6: After the end of this iteration, if the number of iterations exceeds the preset value or the calculation error reaches the preset value, the accuracy requirements will be met and the calculated value will be output. Otherwise, go to **step 4**.

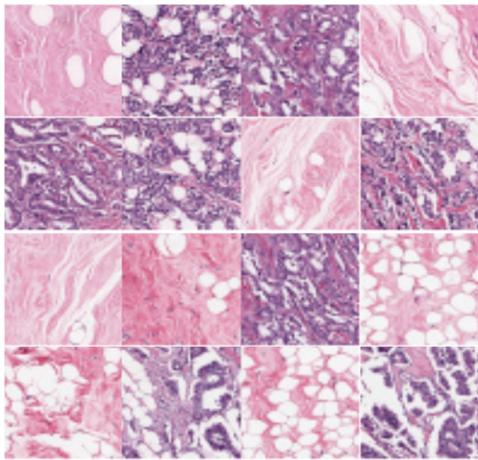


Fig. 5: The mixed breast slice images.

4 EXPERIMENTAL RESULTS

In this section, we first introduce the experimental process and related parameter settings, then introduce the medical image data set used in the experiment, and finally analyze the experimental results.

4.1 Data Set

The medical image data set used in the experiment is from the breast histopathology image on the kaggle website [40]. The original data set consists of a complete slide image of 162 breast cancer specimens, each scanned 40 times, and 277524 color blocks (IDC negative 198738, IDC positive 78786) with an extraction size of 50×50 . The mixed breast slice image is shown in Fig. 5. We select 12000 images as the experimental data set, including 6000 negative images and 6000 positive images.

4.2 Experiment Process and Parameter Setting

(1) The flowchart of the experiment is shown in Fig. 6. The left is the training process, and the right is the

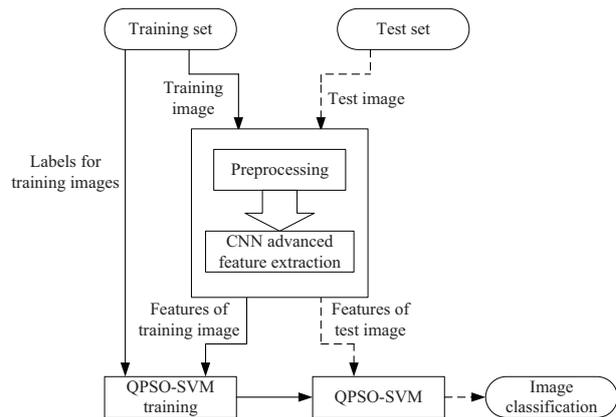


Fig. 6: Experiment flow chart.

testing process. In all experiments, the symbols ‘CNN’ and ‘PCA’ refer to the corresponding feature extraction methods based on CNN and PCA. Firstly, two kinds of images are extracted from breast histopathology images, and training set and test set are obtained. After preprocessing (binarization, grayscale, etc.), the data are extracted by CNN. Finally, SVM classifier optimized by QPSO is used to classify and recognize the data.

(2) For CNN, we use a typical LeNet-5 framework, whose network structure is described in [38]. Here, we select the convolution kernel size as 5×5 , and the pooling size as 2×2 to extract the characteristics of medical images. At the same time, we also use another typical feature extraction method: PCA, for which we use 90% components on purpose to ensure a fair comparison. For QPSO-SVM, the searching ranges of C and σ are all set as $[0.1, 100]$. The swarm size N is set to 100. The maximum number of iterations k_{\max} is set to 50. For comparison, we select another two highly related optimization methods PSO and CV. For PSO-SVM, the maximum number of iterations k_{\max} , the swarm size, and the searching ranges of C and σ are set as the same with QPSO-SVM. For CV-SVM, three-fold CV is used and the searching ranges of C and σ are all set as $[2^{-10}, 2^{-9}, \dots, 2^{10}]$. We should emphasize that the QPSO and PSO based optimization methods only need the extreme values of the ranges of C and σ without defining a step value for these two parameters. The QPSO-SVM, PSO-SVM and CV-SVM classification methods are all implemented under the package of Libsvm. Here, we should also emphasize that we empirically select the values for most of the involved parameters. Although this strategy might be suboptimal, it has been observed to produce good results in practice. Another important reason is that all methods share the same parameter settings, which is fair for comparison.

Table 1: Test accuracy of NO-SVM(%).

Samples	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test																																																																																													
	1000	200	2000	400	3000	600	4000	800	5000	1000	6000	1200	6000	1000	7000	1400	8000	1600	9000	1800	10000	2000																																																																																					
1	89.00	89.75	86.50	86.88	90.10	89.10	88.90	90.38	90.62	91.38	91.25	2	91.50	88.00	85.83	86.38	88.20	89.00	88.60	90.85	90.00	90.22	91.25	3	88.00	88.50	88.33	89.13	87.80	88.42	88.10	89.85	89.85	90.67	91.25	4	91.00	88.50	86.67	88.38	92.00	88.33	88.90	89.57	90.18	90.44	91.25	5	88.50	85.50	90.50	88.00	89.00	88.83	88.30	90.14	90.12	91.00	91.25	6	89.50	87.75	88.00	87.00	88.90	87.50	89.60	90.28	90.50	90.78	91.25	7	86.00	88.50	89.50	89.75	87.20	89.58	89.50	89.64	90.68	90.17	91.25	8	86.50	86.25	88.50	87.38	88.30	89.75	89.10	90.78	90.06	90.00	91.25	Mean	88.75	87.84	87.98	87.86	88.94	88.81	88.88	90.19	90.25	90.58	91.25
Mean	88.75	87.84	87.98	87.86	88.94	88.81	88.88	90.19	90.25	90.58	91.25																																																																																																

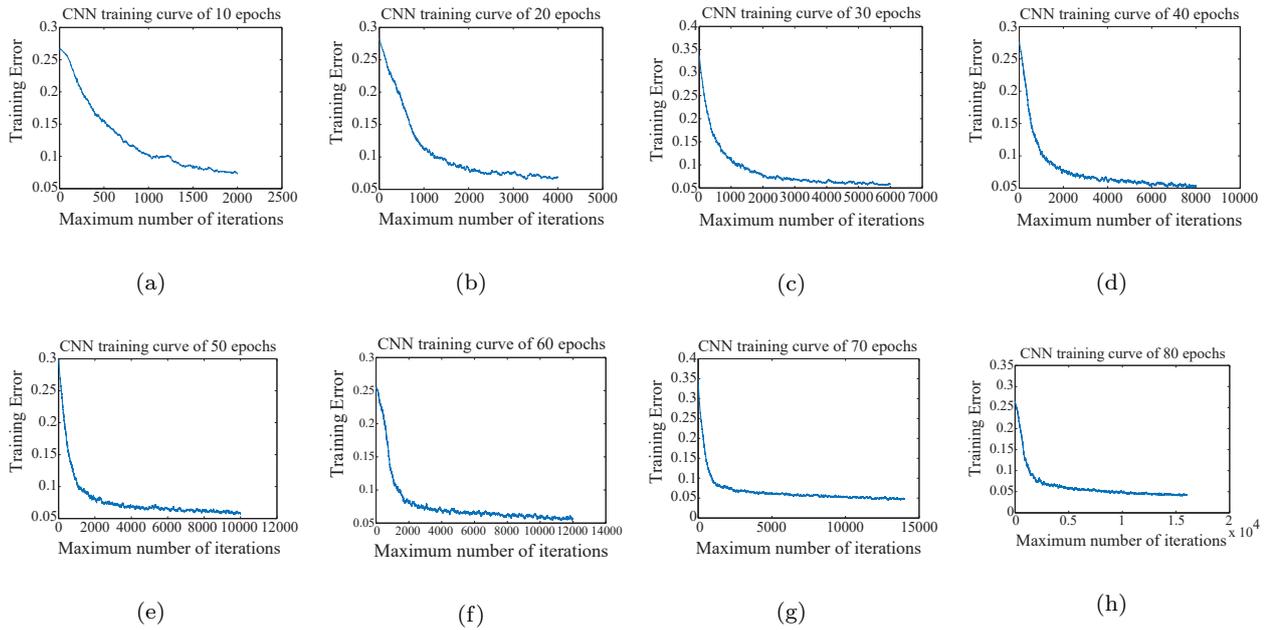


Fig. 7: CNN model evaluation chart.

(3) Finally, we would like to emphasize that the experiments were conducted under Matlab R2014a in a processor Intel(R) Core(TM) i7-8700CPU@3.20GHz with an operating system of 64-bit.

4.3 Experiments with Medical Images

(1) First, we randomly select 10000 images as training sets and 2000 images as test sets to train CNN. Fig. 7 shows the training errors of the CNN model with 10~80 training epochs. It can be seen that the curve of the training error drops, becomes more and more smooth, and tends to be stable. The test errors of Fig. 7a ~ Fig. 7h are 4.14%, 4.40%, 3.35%, 3.00%, 3.10%, 5.45%, 2.95%, 3.10%, respectively. The curves in Fig. 7a and Fig. 7b are not smooth, and there are many spikes. The curves in Fig. 7e and Fig. 7f tend to be stable, but there are many spikes. The curve in Fig. 7g is relative-

ly smoothest among the 8 curves, and there are fewer spikes. The above results show that the training error does not decrease with the increase of the number of iterations. Therefore, in the following experiments, we choose the corresponding model of Fig. 7g as the CNN feature extractor.

(2) In this part, we use multiple experiments to prove the effectiveness of CNN feature extraction. In the experiment, we use the SVM classifier which comes with Matlab toolkit, hereinafter referred to as NO-SVM. First of all, we use 12000 pictures to do multiple experiments, and gradually increase the number of pictures in training set and test set. We used NO-SVM and CNN-SVM to do comparative experiments. The experimental results are shown in Table 1 and Table 2. Each group is carried out eight experiments. Fig. 8 shows the average value of the training accuracies in each group. As you can see from the figure, as the size of the data

Table 2: Test accuracy of CNN-SVM(%).

Samples	Train	Test																				
	1000	200	2000	400	3000	600	4000	800	5000	1000	6000	1200	6000	1000	7000	1400	8000	1600	9000	1800	10000	2000
1	91.50		91.75		91.17		92.63		93.00		92.83		93.10		92.50		91.94		92.94		93.15	
2	95.00		92.00		92.17		90.63		91.40		93.00		92.30		90.79		92.88		92.57		93.15	
3	92.50		91.75		91.83		92.13		91.40		92.92		91.80		92.00		92.06		93.00		93.15	
4	92.00		89.25		92.83		89.50		92.90		93.42		91.40		92.00		91.81		92.61		93.15	
5	89.50		90.00		91.00		92.13		92.60		91.33		92.40		92.14		93.31		92.44		93.15	
6	88.00		91.50		91.83		91.25		92.50		93.42		93.70		91.36		92.38		92.72		93.15	
7	91.00		87.25		90.17		92.13		93.10		91.83		92.40		93.64		90.81		92.17		93.15	
8	90.00		91.00		91.67		91.13		91.80		93.42		91.50		93.36		92.63		92.61		93.15	
Mean	91.19		90.56		91.58		91.44		92.34		92.77		92.33		92.22		92.23		92.63		93.15	

Table 3: Test accuracy of SVM(%).

Experiments	1	2	3	4	5	6	7	8	9	10	Mean
CNN-SVM	93.30	92.30	92.10	94.20	92.60	93.90	92.10	93.20	92.00	93.20	92.89
PCA-SVM	92.90	92.80	91.50	92.80	92.10	92.30	91.40	91.40	92.80	92.80	92.28
NO-SVM	88.90	88.60	88.10	88.90	88.30	89.60	89.50	89.10	88.50	88.10	88.76

set increases, the blue line is always higher than the red line. It shows that the average accuracy of feature extraction by CNN is greatly improved. It should be noted that both experiments use SVM classifier with default parameters in Matlab toolkit. So the comparative experiment is fair. Under the condition that all the pictures of the data set participate in the experiment, the accuracy of NO-SVM test set is stable at 91.25%, and that of CNN-SVM test set is stable at 93.15%.

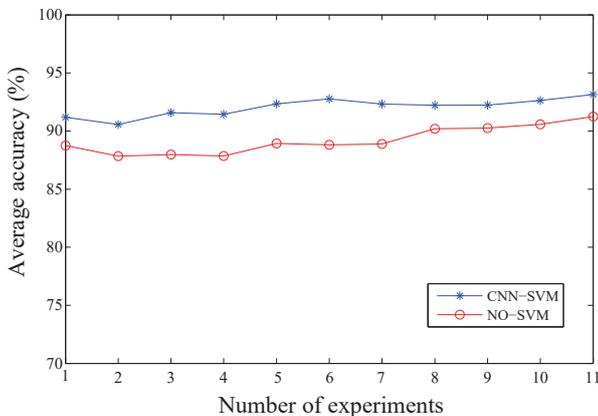


Fig. 8: Experimental comparison chart.

(3) In this part, we carry out a comparative test of feature extraction methods. 6000 and 1000 pictures are randomly selected as training set and test set, and CNN-SVM, PCA-SVM and NO-SVM are used for multiple groups of experiments. The experimental results are recorded in Table 3. It should be noted that the training set and data set used in each group of experiments are the same. In order to make the results more intuitive, we made a line chart of test accuracy,

as shown in Fig. 9. Blue line and dots represent CNN-SVM, red line and dots represent PCA-SVM, and black line and dots represent NO-SVM respectively. Three lines of different colors represent the corresponding average value. It can be seen from the figure that after the feature extraction method is added, the test accuracy is significantly improved. Compared with PCA, the test accuracy of CNN is higher than that of PCA.

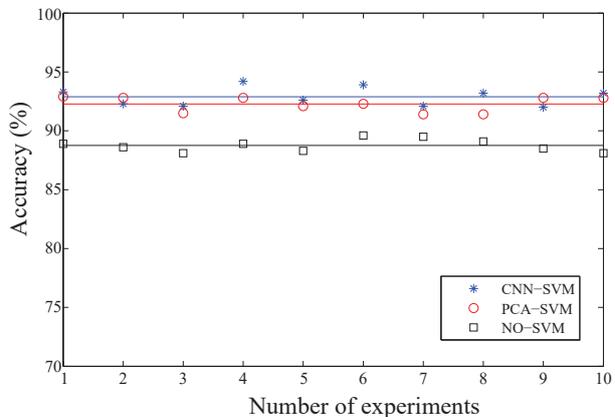


Fig. 9: Experimental comparison chart.

(4) In this part, we still use 6000 pictures as the training set and 1000 pictures as the test set. We prove the effectiveness of the proposed CQ-SVM classification method for medical images. First, we obtain the fitness curves of QPSO-SVM, PSO-SVM, and the fitness surface of CV-SVM based on CNN. As shown in Fig. 10, we can see the fitness values reach to 94.65%, 94.42%, and 94.31% for QPSO-SVM, PSO-SVM, and CV-SVM on train samples, respectively. The average fitness curves tell the average fitness between the current position and

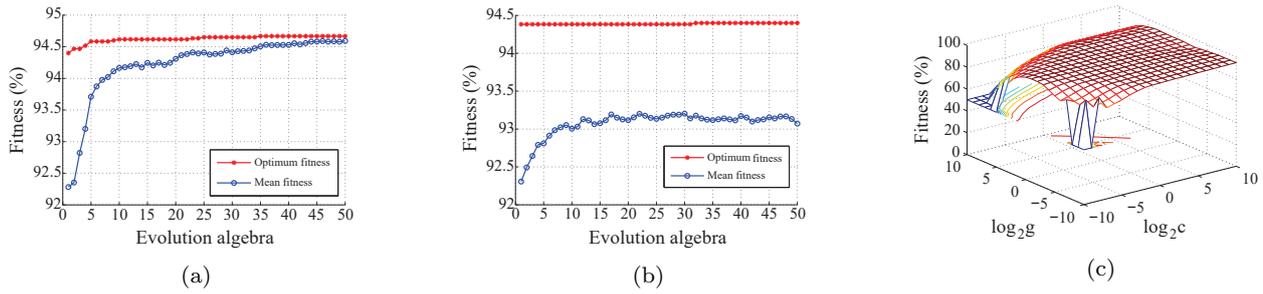


Fig. 10: (a) Fitness curves of QPSO-SVM. (b) Fitness curves of PSO-SVM. (c) Fitness curves of CV-SVM.

its optimal position for each particle during the evolution. While, the best fitness curves (the red lines) tell the overall fitness of the swarm, which are stable during the evolution and reflect the final outcome accuracy. Concerning CV-SVM, the 3-D graphics in those figures directly illustrate the accuracy since CV conducts a simple grid search of different C and σ parameters within a predefined range. Therefore, from Fig. 10 we can see our parameter settings for QPSO-SVM, PSO-SVM, and CV-SVM are appropriate for deriving a desired and stable classification performance. Similarly, the method of averaging the results of multiple groups of experiments. We give the experimental results of using different feature extraction methods and optimization methods on the same training set and test set. The results are shown in Table 4. We highlight the best ones (the highest classification accuracy) with bold typeface for each cases. It can be seen that the proposed CQ-SVM model has the highest test accuracy.

Table 4: Test accuracy of SVM(%).

Classifier	Feature extraction			Optimization method		Accuracy(%)
	SVM	NOPCA	CNN	CV	PSO	
1	✓					88.40
2		✓				91.84
3			✓			94.22
4		✓		✓		95.35
5		✓			✓	96.00
6		✓				✓ 96.20
7			✓	✓		96.75
8			✓		✓	96.90
9			✓			✓ 97.85

¹ “✓” means the corresponding method is used in the experiment.

5 CONCLUSION

This paper has proposed a new hybrid CQ-SVM model based on deep learning to solve the problem of med-

ical image classification. This model has used a CNN based feature extraction method and SVM optimized by QPSO for classification. CNN is a trainable automatic feature extractor, and SVM optimized by QPSO is a trainable output classifier. In order to prove the effectiveness of the model, we have carried out multiple groups of classification accuracy comparison experiments, and evaluated the performance of CQ-SVM classification model under different feature extraction methods and parameter optimization methods as well as their combination. Further experiments with additional scenes and comparison methods should be conducted in the future. Furthermore, we also envisage the following future perspectives for the development of the presented work.

(1) In this paper, we only give the experiment with low dimensional optimization application by adopting QPSO. In our future work, we will dig high dimensional practical problem to validate the advantages of QPSO.

(2) We will study and improve the optimization method of SVM, so that it can optimize the parameters of SVM better.

(3) We will use more complex CNNs to perform feature extraction on higher-dimensional pictures, such as AlexNet, VGGNet.

(4) We only compare the traditional feature extraction method such as PCA. In our future work, we will include some state-of-the-art methods.

Acknowledgements This work is supported in part by National Key Research and Development Program of China (No. 2018AAA0100101), in part by National Natural Science Foundation of China (Grant no. 61932006, 61772434, 61806169), in part by China Postdoctoral Science Foundation under Grant 2018M643085, and in part by Fundamental Research Funds for the Central Universities (Grant no. XDJK2018D005, XDJK2019C020).

Compliance with Ethical Standards

Conflict of interest Author Yulong Wang declares that he has no conflict of interest. Author Xiaofeng Liao declares that he has no conflict of interest. Author Dewen Qiao declares that he has no conflict of interest. Author Jiahui Wu declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, Advances in hyperspectral image classification: Earth monitoring with statistical learning methods, *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 45–54, 2013.
- Q. Fu, D. S. Mansjur, and B. H. Juang, Empirical system learning for statistical pattern recognition with non-uniform error criteria, *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4621–4633, 2010.
- H. Venkateswara, S. Chakraborty, and S. Panchanathan, Deep-learning systems for domain adaptation in computer vision: Learning transferable feature representations, *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 117–129, 2017.
- P. Ray, G. K. Budumuru, and B. K. Mohanty, A comprehensive review on soft computing and signal processing techniques in feature extraction and classification of power quality problems, *Journal of Renewable and Sustainable Energy*, vol. 10, no. 2, 2018.
- C. B. R. Ferreira and D. L. Borges, Automated mammogram classification using a multi-resolution pattern recognition approach, in *14th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2001)*, 15–18 October 2001, Florianopolis, Brazil, 2001.
- Y. Zhou, X. He, L. Huang, L. Liu, F. Zhu, S. Cui, and L. Shao, Collaborative learning of semi-supervised segmentation and classification for medical images, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- A. S. M. Sohail, P. Bhattacharya, S. P. Mudur, and S. Krishnamurthy, Classification of ultrasound medical images using distance based feature selection and fuzzy-svm, *Lecture Notes in Computer Science*, vol. 6669, pp. 176–183, 2011.
- Z. Huang, X. Zhu, M. Ding, and X. Zhang, Medical image classification using a light-weighted hybrid neural network based on pcanet and densenet, *IEEE Access*, to be published, DOI: 10.1109/ACCESS.2020.2971225, 2020.
- B. H. Yilmaz, C. M. Yilmaz, and C. Kose, Diversity in a signal-to-image transformation approach for eeg-based motor imagery task classification, *Medical and Biological Engineering and Computing*, vol. 58, no. 2, pp. 443–459, 2020.
- C. Saunders, M. O. Stitson, J. Weston, R. Holloway, L. Bottou, B. Scholkopf, and A. Smola, Support vector machine, *Computer Science*, vol. 1, no. 4, pp. 1–28, 2002.
- J. Cui, Z. Li, G. Jian, L. Rui, and X. Xu, The application of support vector machine in pattern recognition, in *IEEE International Conference on Control and Automation*, 2007.
- K. Manimala, K. Selvi, and R. Ahila, Optimization techniques for improving power quality data mining using wavelet packet based support vector machine, *Neurocomputing*, vol. 77, no. 1, pp. 36–47, 2012.
- N. J. Tian, N. M. Gao, and N. Y. Xiang, Intelligent optimized control of flocculation process of sewage treatment based on support vector machine, in *2006 IEEE International Conference on Information Acquisition*, 2006.
- S. Gupta, R. Kambli, S. Wagh, and F. Kazi, Support-vector-machine-based proactive cascade prediction in smart grid using probabilistic framework, *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2478–2486, 2015.
- G. C. Cawley and N. L. C. Talbot, On over-fitting in model selection and subsequent selection bias in performance evaluation, *Journal of Machine Learning Research*, vol. 11, no. 1, pp. 2079–2107, 2010.
- X. Gao and J. Hou, An improved svm integrated gs-pca fault diagnosis approach of tennessee eastman process, *Neurocomputing*, vol. 174, no. JAN.22PT.B, pp. 906–911, 2016.
- K. Sopyla and P. Drozda, Stochastic gradient descent with barzilaicborwein update step for svm, *Information Sciences*, vol. 316, pp. 218–233, 2015.
- G. Rubio, H. Pomares, I. Rojas, L. J. Herrera, and A. Guilln, Efficient optimization of the parameters of ls-svm for regression versus cross-validation error, in *Artificial Neural Networks - ICANN 2009, 19th International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings, Part II*, 2009.
- H. S. Madraswala, Modified genetic algorithm solution to unit commitment problem, in *2017 International Conference on Nascent Technologies in Engineering (ICNTE)*, 2017.
- Eberhart and Y. Shi, Particle swarm optimization: developments, applications and resources, in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, 2002.
- A. Tharwat and A. E. Hassanien, Quantum-behaved particle swarm optimization for parameter optimization of support vector machine, *Journal of Classification*, vol. 36, no. 3, pp. 576–598, 2019.
- L. Liu, W. Sun, and B. Ding, Offline handwritten chinese character recognition based on dbn fusion model, in *2016 IEEE International Conference on Information and Automation (ICIA)*, 2016, pp. 1807–1811.
- G. Amaranageswarao, S. Deivalakshmi, and S.-B. Ko, Wavelet based medical image super resolution using cross connected residual-in-dense grouped convolutional neural network, *Journal of Visual Communication and Image Representation*, p. 102819, 2020.
- G. WANG and Y. HUANG, Medical-image-fusion algorithm based on a detail-enhanced and pulse-coupled neural network model stimulated by parallel features, 2020.
- H. Zarrabi, A. Emami, P. Khadivi, N. Karimi, and S. Samavi, Blessmark: a blind diagnostically-lossless watermarking framework for medical applications based on deep neural networks, 2020.
- E. Ahn, A. Kumar, D. Feng, M. Fulham, and J. Kim, Unsupervised feature learning with k-means and an ensemble of deep convolutional neural networks for medical image classification, 2019.
- Y. Lecun, Y. Bengio, and G. Hinton, Deep learning, *Nature*, vol. 521, no. 7553, p. 436, 2015.
- H. Greenspan, B. van Ginneken, and R. M. Summers, Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique, *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

29. A. Khatami, A. Khosravi, T. Nguyen, C. P. Lim, and S. Nahavandi, Medical image analysis using wavelet transform and deep belief networks, *Expert Systems with Applications*, vol. 86, pp. 190–198, 2017.
30. Geert, Litjens, Thijs, Kooi, Babak, Ehteshami, Bejnordi, Arnaud, Arindra, and Adiyoso, A survey on deep learning in medical image analysis, *Medical Image Analysis*, vol. 42, no. 9, pp. 60–88, 2017.
31. K. Hua, C. Hsu, S. C. Hidayati, W. Cheng, and Y. Chen, Computer-aided classification of lung nodules on computed tomography images via deep learning technique, *OncoTargets and Therapy*, vol. 8, pp. 2015–2022, 2015.
32. J. David, R. Krishnan, and K. A. Suresh, Neural network based retinal image analysis, in *Image and Signal Processing, 2008. CISP 08. Congress on, 2008*.
33. J. Ramirez, R. Chaves, J. M. Grriz, I. lvarez, and F. Segovia, Functional brain image classification techniques for early alzheimer disease diagnosis, in *Bioinspired Applications in Artificial and Natural Computation, Third International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2009, Santiago de Compostela, Spain, June 22-26, 2009, Proceedings, Part II, 2009*.
34. D. Avci, M. K. Leblebicioglu, M. Poyraz, and E. Dogantekin, A new method based on adaptive discrete wavelet entropy energy and neural network classifier (adweenn) for recognition of urine cells from microscopic images independent of rotation and scaling, *Journal of Medical Systems*, vol. 38, no. 2, pp. 1–9, 2014.
35. I. T. Jolliffe, *Principal component analysis*. 2nd ed, *Weather*, vol. 98, no. 10, 2005.
36. Lecun, Y., Bottou, L., Bengio, Haffner, and P., Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
37. F. Lauer, C. Y. Suen, and G. Bloch, A trainable feature extractor for handwritten digit recognition, *Pattern Recognition*, vol. 40, no. 6, pp. 1816–1824, 2007.
38. Sun, Jun, Wei, Fang, Xiaojun, Wu, Palade, and Vasile, Quantum-behaved particle swarm optimization: Analysis of individual particle behavior and parameter selection. *Evolutionary Computation*, vol. 20, no. 3, pp. 349–393, 2012.
39. X. Niu and C. Y. Suen, A novel hybrid cnn-svm classifier for recognizing handwritten digits, *Pattern Recognition*, vol. 45, no. 4, pp. 1318–1325, 2012.
40. P. Mooney, Breast histopathology images, <https://www.kaggle.com/paultimothymooney/breast-histopathology-images>.