THE ECG SIGNAL CLASSIFICATION BASED ON ENSEMBLE LEARNING OF PSO-ELM ALGORITHM

W. Li[*], *B. Li*[*,†], *H.L. Guo*[*], *Y.X. Fang*[*], *F.J. Qiao*[*], *S.W. Zhou*[†]

Abstract: ECG anomaly detection plays an important role in clinical medicine. So far, a number of ECG recognition technologies have emerged in this field, but most often suffer from slow training and instability. Considering that the Extreme Learning Machine (ELM) and Particle Swarm Optimization (PSO) algorithm have the advantages of fast learning speed and strong generalization ability, this paper integrates multiple independent PSO-ELM model and proposes a novel ensemble learning framework termed as E-PSO-ELM to realize ECG signals recognition. More specifically, the individual PSO-ELM adopts the input weight and hidden layer deviation of ELM as the particles in the PSO algorithm, and takes the root mean square error of ELM training sample as the adaptive value of the particles, so as to enhance the stability of the network and realize high ECG recognition rate. The simulation results on MIT-BIH Arrhythmia Database show that E-PSO-ELM has a high classification accuracy rate of 98.23 %. In addition, compared with other algorithms, the stability of E-PSO-ELM is more prominent, which can reduce the probability of operating errors. Therefore, E-PSO-ELM has a high practical value.

Key words: *extreme learning machine, particle swarm optimization, ensemble learning*

Received: January 13, 2020
Revised and accepted: August 30, 2020

DOI: 10.14311/NNW.2020.30.018

1. Introduction

According to incomplete statistics from the National Cardiovascular Center, about 290 million people suffer from cardiovascular diseases, with a mortality rate as high as 40 percent. To make matters worse, the prevalence of cardiovascular diseases is still on the rise.

Electrocardiogram plays an important role in clinical detection of cardiovascular diseases. The classification of ECG signals has become a hot topic in modern

[*]Wei Li; Bin Li – Corresponding author; Hongli Guo; Yixian Fang; Fengjuan Qiao; School of Mathematics and Statistics, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China, E-mail: ribbenlee@126.com

[†]Bin Li; Shuwang Zhou; Shandong Computer Science Center (National Supercomputer Center in Jinan), Jinan 250041 China.

research. This article applies integration ideas to traditional machine learning algorithms to build a new type of integration model which is used for ECG signal classification.

In most ECG signal classification methods, the QRS wave detection is the basis for analyzing the electrocardiogram. Robust QRS complex detection can locate the R peak and other peaks more accurately, thereby extracting more comprehensive feature information and improving the classification accuracy of the model. Fig. 1 shows the normal ECG signals and temporal characteristics, where the QRS wave contains a large amount of cardiac status information [1]. For this reason, over the past decades, researchers have proposed various improved QRS detection methods, which shows a tendency of cross merging based on traditional techniques. For example, the Pan-Tompkins algorithm [2] is a classic QRS wave detection algorithm with a recognition rate of 99.3 %. Marchon et al. improved the Pan-Tompkins algorithm by using two-stage elliptical IIR bandpass filters and achieved 0.012 % and 0.044 % errors on database adfecgdb and PhyC 2013, respectively [3]. Yazdani [4] combined mathematical morphology and adaptive structure to detect QRS waves. [5] and [6], based on the wavelet transform, respectively applied synchronous compression wavelet and inverse biorthogonal wavelet decomposition combined with nonlinear filtering to detect R waves.

The learning process of ECG signals includes two aspects: feature extraction and classification. In terms of feature extraction, threshold-based methods [7], digital filter-based methods [8, 9], Hermite functions [10], Hermite polynomials, frequency-based features [11], ECG morphology [12], higher order cumulant features [13] and statistical features [14, 15] and so on, have been widely used, respectively. Among these methods mentioned above, wavelet transform is a local transform of space (time) and frequency, which can effectively extract information from signals. In this paper, wavelet transform is also used to extract QRS features as input of classifiers.

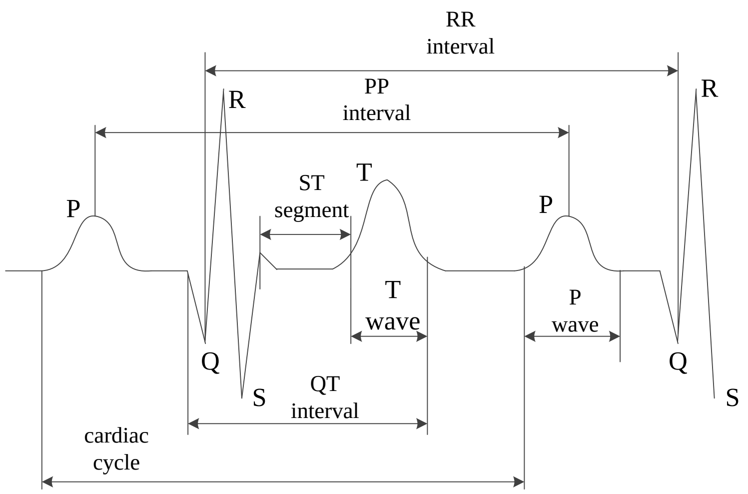

Fig. 1 *A normal ECG signal and its temporal features of one period.*

Meanwhile, a stable and accurate recognition model is needed for the classification of ECG signals. As an efficient machine learning algorithm, ELM has been widely used in identifying classification problems. ELM, a single hidden layer feedforward neural network proposed by Huang et al. [16], which can randomly generate input weights and hidden layer deviations, and calculate the output weights by generalized inverses of hidden neurons. Therefore, ELM shows the advantages of simple network structure, fast running speed, strong generalization performance and so on, making up for the shortcomings of the traditional feedforward neural networks.

In view of the instability of ELM, in recent years, researchers proposed a number of improvement methods to optimize parameters and improve the overall performance of the networks. The reference [17–19] constructed the entire network in the form of incremental nodes, which could effectively reduce errors. In [20], Xu et al. proposed the PSO-ELM algorithm to solve some prediction problems. In [21], Han et al. optimized the parameters using an improved PSO-ELM(IPSO-ELM) with inertia weight adaptive adjustment. In [22], Zhu et al. introduced a differential evolution algorithm to adjust the parameters and improve the performance of the ELM algorithm.

In addition, the ensemble learning can further improve the generalization performance of the neural networks. For example, reference [23] proposed a simple integration model (GE-ELM), which used genetic algorithm to generate the initial network, and then integrated the new network structure with special sorting strategy. Reference [24] proposed voting-ELM algorithm to solve EXtensible Markup Language (XML) document classification. Lan et al. effectively increased the stability of a single ELM by integrating the output averaging of multiple Online Sequential Extreme Learning Machines (OS-ELM) [25]. Reference [26] proposed a selective evolution learning algorithm based on differential evolution (DESE-ELM) to solve the classification problem.

In order to overcome the shortcomings of ELM, combine PSO optimization algorithm and integration idea, this paper proposes an integration strategy based on PSO-ELM for the classification of ECG signals. According to the size of adaptive value of PSO, we use ordinal to optimize the individual, select M optimal individual by sorting, generate M independent neural network structures, and obtain the optimal input weight and hidden layer deviation of the group M ELMs. Finally, by comparing the actual output with the expected output, the classification accuracy of different ELMs is calculated with the maximum voting principle. In order to improve the robustness of the network and reduce the classification error rate, the proposed approach integrates multiple classifiers, which effectively improves the classification accuracy of the whole neural network is better than the DESE-ELM algorithm in stability. In summary, the main advantages of E-PSO-ELM are as follows:

1) E-PSO-ELM adopts PSO to optimize the randomly generated parameters of input weights and hidden biases in ELM, which improves the generalization performance of the network. Moreover, it selects the best ELM for integration, which further improves the network generalization ability and enhances the classification accuracy of the model.

2) Using PSO to optimize ELM parameters, the proposed model is more stable. In addition, several weak classifiers are integrated into a strong classifier, which further improves the stability of the model. The comparison experiment with several common algorithms also verifies the above conclusion.

The frame diagram of the proposed method termed as E-PSO-ELM is shown in Fig. 2.

This paper is organized as follows. In Section 2, ELM and PSO are briefly introduced. The detailed description of E-PSO-ELM is given in Section 3. Experimental results and analysis are reported in Section 4. A conclusion is drawn in Section 5.

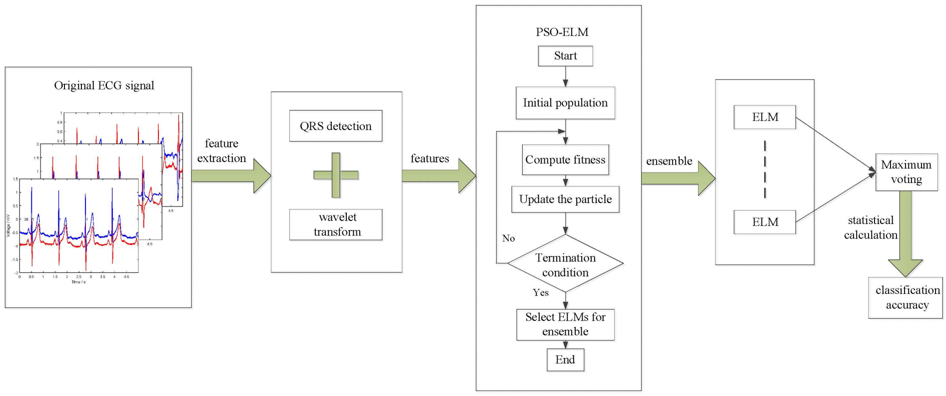

Fig. 2 *The framework of E-PSO-ELM learning algorithm*

2. Related work

In this section, we briefly introduce ELM and PSO algorithms to provide the necessary background knowledge for the E-PSO-ELM in Section 3.

2.1 Extreme Learning Machine

Extreme Learning Machine (ELM) [16] is a single hidden layer feedforward neural network (SLFN) whose structure is shown in Fig. 3. ELM does not need to update iteration, but directly generates input weight and hidden layer deviation randomly, so it has a faster learning speed.

Given N input samples $\{(x_i, t_i)\}_{i=1}$, $\tilde{N}$ hidden layer neurons, the ELM learning algorithm can approximate these N samples with zero error, then, the ELM model can be formulated as

$$f_{\tilde{N}}(x_i) = \begin{bmatrix} \sum_{j=1}^{\tilde{N}} \beta_{j1} G(w_j \cdot x_i + b_j) \\ \sum_{j=1}^{\tilde{N}} \beta_{j2} G(w_j \cdot x_i + b_j) \\ \vdots \\ \sum_{j=1}^{\tilde{N}} \beta_{jm} G(w_j \cdot x_i + b_j) \end{bmatrix} = t_i, i = 1, 2, \ldots, N, \tag{1}$$

where $G(\cdot)$ is the corresponding activation function, w_j is the weight vector connecting the jth hidden neuron and the input neurons, b_j is bias (or impact factor) of the jth the hidden neuron. The Eq. (1) can also be simplified as

$$\mathbf{H}\beta = T. \tag{2}$$

$\mathbf{H}$, β and T are respectively

$$\mathbf{H} = \begin{bmatrix} G(w_1 \cdot x_1 + b_1) & \cdots & G(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ G(w_1 \cdot x_N + b_1) & \cdots & G(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}},$$

$$\beta = \begin{bmatrix} b_1^T \\ \vdots \\ b_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m},$$

where $\mathbf{H}$ is the output matrix of the hidden layer. β is the matrix of the output weights and T is the target matrix of ELM.

By randomly setting the parameters of the hidden layer node, the output matrix $\mathbf{H}$ [27] can uniquely be determined, and the output weight β can be obtained by Eq. (3).

$$\beta = \mathbf{H}^{+} T, \tag{3}$$

where $\mathbf{H}^{+}$ denotes the Moore-Penrose inverse of matrix $\mathbf{H}$.

2.2 Particle swarm optimization

Particle Swarm Optimization (PSO) [28] is a population and heuristic optimization algorithm that has been widely used in scientific research and engineering applications. Next, we briefly introduce the basic flow of PSO.

In an N-dimensional search space, (1) initialize a population of m individuals $P = (x_1, x_2, \ldots, x_i, \ldots, x_m)^T$, where the position and velocity information of particle i in the population are expressed as $x_i = (x_{i1}, x_{i2}, \ldots, x_{iN})^T$, $v_i = (v_{i1}, v_{i2}, \ldots, v_{iN})^T, i = 1, 2, \ldots, m$, respectively; (2) then calculate the fitness values of all particles, find the optimal solution, and update the position information and velocity information of the current particle according to Eq. (4) and Eq. (5).

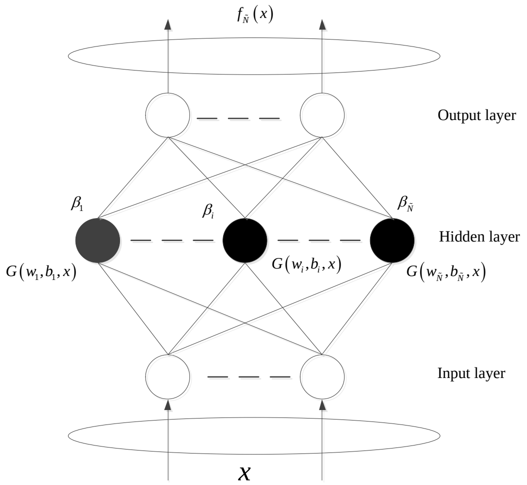

Fig. 3 *The network structure of Extreme Learning Machine.*

$$\begin{aligned} v_i^{t+1} &= wv_i^t + c_1 * rand() * \left(P_i^t - x_i^t\right) + \\ &\quad +c_2 * rand() * g\left(P_g^t - x_g^t\right), \end{aligned} \tag{4}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1}, \tag{5}$$

where $rand()$ is a random number in the interval 0 and 1. w denotes the inertia weight. c_1 and c_2 are the acceleration constants, which are used to adjust individual extremum P_i and global optimal solution P_g. In this paper, the time-varying acceleration coefficients c_1 and c_2 are represented the same as the references [29–31], and their dynamic adjustment formulas are as follows:

$$c_1 = \frac{(-2) \times (Iteration)^3}{(Max\_iteration)^3 + 2}, \tag{6}$$

$$c_2 = \frac{2 \times (Iteration)^3}{(Max\_iteration)^3}, \tag{7}$$

where $Iteration$ denotes the number of current iterations. $Max\_iteration$ stands for the maximum iteration number of the algorithm.

In order to ensure that the algorithm has higher exploration ability in the initial stage of iteration, the convergence speed of the algorithm is accelerated in the later stage of the iteration, and the dynamic inertia weight w is defined as follows:

$$w(t) = w_{\max} - \text{Iteration} \times \left[(w_{\max} - w_{\min}) / (\max - \text{Iteration})\right], \tag{8}$$

where $w_{\max}$ and $w_{\min}$ are the initial and terminal values of inertia weight in the iteration process, respectively.

3. Ensemble learning based on ELM and PSO algorithms

In this part, we propose an ensemble learning strategy based on ELM and PSO algorithms (PSO-ELM), which is called E-PSO-ELM learning algorithm. Specifically, we add a selection strategy on the basis of the PSO-ELM algorithm, generate candidate neural network set, and integrate the target neural network into the candidate set.

More specifically, in the PSO-ELM learning algorithm, the input weight and hidden layer deviation of ELM are taken as the particles in the particle swarm optimization algorithm, and the root mean square error (RMSE) of the ELM training sample shown as Eq. (9) is taken as the adaptive value of the particle. According to the adaptive value of each particle, the position and velocity of the particles are updated according to Eq. (4) and Eq. (5). The maximum number of iterations is selected as the termination condition of the iteration.

$$\text{fitness} = RMSE = \sqrt{\frac{\sum\limits_{j=1}^{N} \left\| \sum\limits_{i=1}^{\tilde{N}} \beta_i G\left(w_i \cdot x_j + b_i\right) - t_j \right\|_2^2}{N}}. \tag{9}$$

The number of initialization samples is N, the number of hidden layer neurons is $\tilde{N}$, and the number of candidate networks is M. The individuals in the last iteration are regarded as potential candidate neural networks, the fitness values of all individuals are sorted, and the first M individuals are selected as candidate neural network sets. After the candidate set is substituted into the network for learning, the classification results of M groups can be obtained through the calculation of M independent networks. The actual output is compared with the expected output, the correct result of classification is 1, and the wrong result of classification is 0. Then the maximum voting principle is adopted to obtain the final classification result.

According to the statistical analysis, the formula for calculating the accuracy rate is as follows:

$$\text{accuracyrate} = \frac{\sum_{i=1}^{N} n_{1i\left(\max\left(\sum_{r=1}^{M} n_{0r}, \sum_{r=1}^{M} n_{1r}\right)\right)}}{N}, \tag{10}$$

where N: the total number of samples; M: the number of integrated neural networks. $\sum_{r=1}^{M} n_{0r}$ and $\sum_{r=1}^{M} n_{1r}$ are the number of wrong classification and correct classification of the same sample among the M integrated neural networks, respectively. Function max(): if $\sum_{r=1}^{M} n_{0r} \geq \sum_{r=1}^{M} n_{1r}$, returns a value of 0; if $\sum_{r=1}^{M} n_{0r} < \sum_{r=1}^{M} n_{1r}$, returns a value of 1. $\sum_{r=1}^{M} n_{1i}$ is the number of correctly classified samples.

In order to describe the E-PSO-ELM algorithm more intuitively, the basic process is summarized in Algorithm 1.

Algorithm 1 E-PSO-ELM algorithm.

Randomly initialize the population. Initialize the random velocity and position of the particle.
Training phase
repeat
  The fitness of each particle is calculated according to the fitness function. Then update the velocity and position of the particle.
until Output the fitness values of all particles for all last iterations.
Ensemble phase
repeat
  Sort all fitness values. Select particles for integration.
  Build a weak classifier.
  {The parameters of the selected particles are substituted into the PSO-ELM network for training, and several weak classifiers are constructed.}
  Calculate actual output.
  {Calculate the actual output of the weak classifier.}
  Build a strong classifier.
  {The output of the weak classifier is voted, and the one with more votes is the final actual output.}
until Output the actual output.
Calculate the classification accuracy rate.

4. Experiment

4.1 Pre-processing

To test the effectiveness of the proposed algorithm, we perform validation on MIT-BIH Arrhythmia Database which is the most commonly used and authoritative database in the ECG field. The MIT-BIH database consists of 48 dual-lead ECG records, each containing a large amount of data [32]. Fig. 4 shows an ECG signal diagram of number 100. It can be seen from Fig. 4 that the MIT database is dual-lead. The number 1 in the figure represents the normal heart beat, and the number 28 represents the rhythm change.

Compared with other waveforms, QRS is the largest, most obvious and easiest to detect. Aiming at the singularity of QRS, the feature extraction is carried out by integrating wavelet transforms and wavelet coefficients are used as features. The wavelet coefficients fully reflect the time-frequency domain changes of the signal.

In order to better extract the QRS features, the original ECG signals are preprocessed to make it a signal composed of a single wave peak in a single mode. Secondly, to better detect the QRS position and prevent the missing and wrong detection of QRS, the adaptive threshold method is adopted. Finally, target cardiac beats are intercepted to generate the collection of target types of cardiac beats needed, respectively "normal (N)", "left bundle branch block (LBBB)", "right bundle branch block (RBBB)" and "ventricular premature beat (VPB)". For the fairness of the data, this paper selects 5000 samples from each type of heartbeat as the model input, as shown in Tab. I.

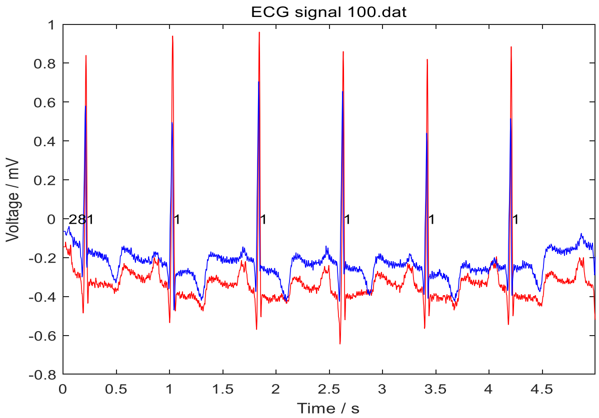

Fig. 4 *The ECG signal diagram of number 100*

Heart beat type	Sample number $\times$ Characteristic number
Normal(N)	5000×250
Left Bundle Branch Block(LBBB)	5000×250
Right Bundle Branch Block(RBBB)	5000×250
Ventricular Premature beat(VPB)	5000×250

Tab. I *Sample data of the target cardiac beat.*

4.2 Experimental analysis

This paper compares the classification accuracy of ECG signal classification in terms of ELM, DE-ELM, PSO-ELM, DESE-ELM, E-PSO-ELM, SVM, CNN-ELM [33], CNN [34], E-SVM [35], respectively. Initial parameters of the PSO-ELM and the E-PSO-ELM are shown in Tab. II, and initial parameters of the DE-ELM and the DESE-ELM are shown in Tab. III. In this paper, the parameters of the SVM learning algorithm for the ECG signal classification are shown in Tab. IV.

The parameter descriptions in Tab. II and Tab. III are as follows, NP: number of population; *iteration*: number of iterations; c_1 and c_2: acceleration constants; w: dynamic inertia weight; $w_{\max}$: initial value of inertia weight; $w_{\min}$: terminal value of inertia weight; CR: crossover probability; *refresh*: mutation strategy; *strategy*: crossover strategy; F: scaling factor; T: tolerance.

NP	*iteration*	c_1	c_2	w	$w_{\max}$	$w_{\min}$
500	100	2	2	2	0.9	0.5

Tab. II *PSO algorithm parameters for the ECG signal classification.*

Fig. 5 shows the variation of the testing classification accuracy of the classification problem on MIT-BIH data set with the number of hidden layer neurons. As

NP	$iteration$	CR	$refresh$	$strategy$	F	T
500	100	0.8	1	3	1	0.02

Tab. III *DE algorithm parameters for the ECG signal classification.*

Penalty coefficient	Gamma
2	1

Tab. IV *The SVM parameters for the ECG signal classification.*

can be seen, the complexity of the neural networks has a great influence on the classification accuracy. Therefore, when dealing with different classification problems, in addition to adjusting the parameters of the classifier for different data sets, the complexity of the neural networks and the number of hidden neurons should appropriately be adjusted to improve the generalization performance of the neural networks model.

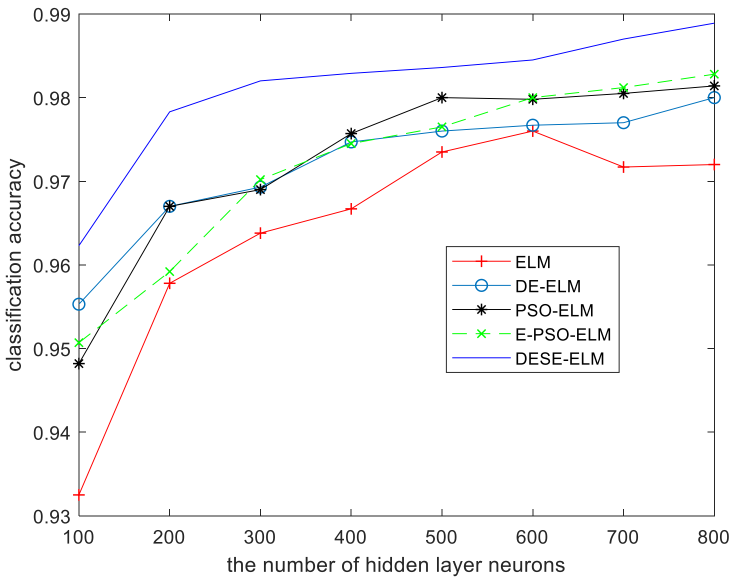

Fig. 5 *The relationship between classification accuracy and the number of hidden layer neurons.*

In this experiment, we set the number of hidden layer neurons in the model as 800, the activation function as sigmoid function, the number of iterations as 100 and the number of neural network integration M as 10. When the number of hidden neurons in the network increased to 900 and 1000, the generalization ability of the network did not improve, while the training time increased significantly. Therefore, the number of hidden neurons selected in this paper is 800. Fig. 6 shows the curve of the number of hidden neurons versus the training time. From Fig. 6, it can be seen that when the number of hidden neurons is 800, the training time of E-PSO-ELM is less than that of DESE-ELM.

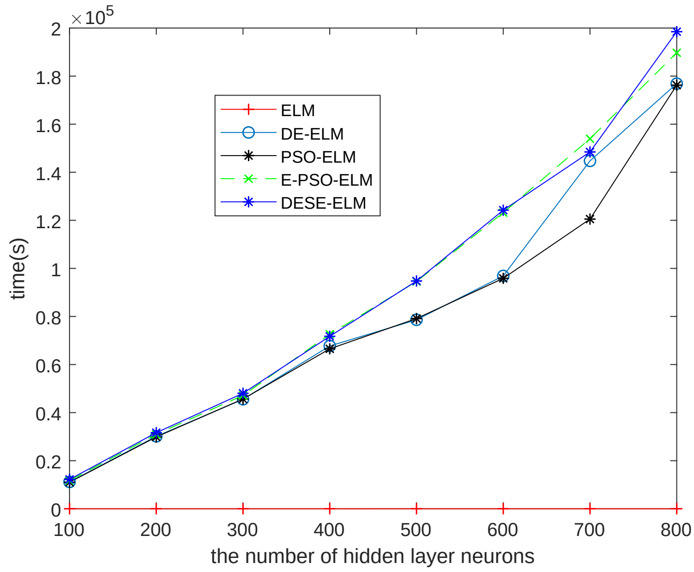

Fig. 6 *The relationship between time and the number of hidden layer neurons.*

In order to eliminate the influence of random initialization, we take the average of 10 experimental results as the final experimental results, and calculate the standard deviation as the stability analysis of the model. The experimental results are shown in Tab. V.

It can be seen from Tab. V that the improved PSO-ELM has improved classification accuracy compared with the original PSO-ELM, but the stability of PSO-ELM is relatively higher. PSO-ELM is the optimal particle adopted as the final input of the network, while E-PSO-ELM selects 10 individuals which can generate better 10 independent network structures according to the adaptive value of particles. Not all of these neural network structures are applicable to the classification problem, so the stability of the final ensemble neural network is relatively poor. For the classification of ECG signals, the classification accuracy rate of the E-PSO-ELM algorithm reaches 0.9823 %, which has good generalization performance. Although the classification accuracy of the E-PSO-ELM algorithm is lower than that of DESE-ELM, it can be seen from the comparison of STD in the Tab. V that PSO is better than DE, and the stability of E-PSO-ELM is better than DESE-ELM. The stability of the model is high, and the probability of error is small when dealing with actual problems. In the running time of the algorithms, the E-PSO-ELM algorithm takes less time than the DESE-ELM algorithm in Fig. 6. Compared with SVM, CNN, E-SVM and CNN-ELM algorithm, E-PSO-ELM has higher classification accuracy and better network generalization performance.

Algorithms	MIT-BIH		
	CPU time(s)	Training error STD	Testing error STD
ELM	3.3750	0.9901	0.9731
		8.1165e-04	0.0025
DE-ELM	1.7686e+05	0.9868	0.9761
		0.0025	0.0022
PSO-ELM	1.7623e+05	0.9909	0.9814
		8.2489e-04	0.0011
E-PSO-ELM	1.89709e+05	0.9910	0.9823
		7.5196e-04	4.8408e-04
DESE-ELM	1.9849e+05	0.9901	0.9889
		8.1165e-04	5.2546e-04
SVM	2.077476	–	0.9688
		–	0.0021
CNN-ELM	–	–	0.8833
CNN [34]	–	–	0.9770
E-SVM [35]	–	–	0.9440

Tab. V *Classification accuracy of different algorithms under MIT-BIH data set.*

In a word, the performance comparison shows that E-PSO-ELM can not only achieve a relatively high classification accuracy rate, but also maintain a strong stability. Therefore, when dealing with practical problems, E-PSO-ELM has high practical value.

5. Conclusions

Based on ensemble learning and PSO-ELM learning algorithm, this paper proposes a new classification recognition algorithm named E-PSO-ELM which has good generalization ability. The network architecture combines the advantages of PSO-ELM algorithm and ensemble learning. The neural network structure of the PSO-ELM is relatively more stable, and the ensemble idea can further improve the generalization performance of the network. The simulation results of the proposed algorithm on the MIT-BIH database are as high as 98.23 %, making it a powerful tool for ECG signal recognition.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (61973185), the development plan of Youth Innovation Team in Colleges and universities of Shandong Province (2019KJN011), the Shandong Province Key Research and Development Program (2018GGX103054), the Natural Science Foundation of Shandong Province under Grant ZR2017MEE033 and Young Doctor Cooperation Foundation of Qilu University of Technology (Shandong Academy of Sciences) (2018BSHZ2008).

References

[1] KONLER B.U., HENNIG C., ORGLMEISTER R. The Principles of Software QRS Detection. *IEEE Engineering in Medicine and Biology Magazine*. 2002, 21(1), pp. 42–57, doi: 10.1109/51.993193.

[2] PAN J., TOMPKINS W.J. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*. 1985, 32(3), pp. 230–236, doi: 10.1109/TBME.1985.325532.

[3] MARCHON N., NAIK G. QRS Detector For Maternal Abdominal ECG. *International Conference on Signal and Information Processing*. 2016, pp. 1–5, doi: 10.1109/ICONSIP.2016.7857463.

[4] YAZDANI S., VESIN J.M. Extraction of QRS Fiducial Points From The ECG Using Adaptive Mathematical Morphology. *Digital Signal Processing*. 2016, 56, pp. 100–109, doi: 10.1016/j.dsp.2016.06.010.

[5] SHARMA T., SHARMA K.K. QRS Complex Detection in ECG Signals Using The Synchrosqueezed Wavelet Transform *IETE Journal of Research*. 2016, 62(6), pp. 885–892, doi: 10.1080/03772063.2016.1221744.

[6] KHOLKHAL M., REGUIG F.B. Efficient Automatic Detection of QRS Complexes in ECG Signal Based on Reverse Biorthogonal Wavelet Decomposition and Nonlinear Filtering. *Measurement*. 2016, 94, pp. 663–670, doi: 10.1016/j.measurement.2016.09.014.

[7] CHOUHAN V.S., MEHTA S.S. Threshold-based Detection of P and T-wave in ECG Using New Feature Signal. *IEEE Transactions on Biomedical Engineering*. 2008, 8(2), pp. 144–153, doi: 10.1080/00045603909357325.

[8] MAHMOUD S.A., BAMAKHRAMAH A., Al-TUNAIJI S.A. Six Order Cascaded Power Line Notch Filter for ECG Detection Systems with Noise Shaping. *Circuits, Systems, and Signal Processing*. 2014, 33(8), pp. 2385–2400, doi: 10.1007/s00034-014-9761-1.

[9] PASOLLI E., MELGANI F. Active Learning Methods for Electrocardiographic Signal Classification. *IEEE Transactions on Information Technology in Biomedicine*. 2010, 14(6), pp. 1405–1416, doi: 10.1109/TITB.2010.2048922.

[10] DOQUIRE G., DE LANNOY G., FRANCOIS D. Feature Selection for Interpatient Supervised Heart Beat Classification. *Computational Intelligence and Neuroscience*. 2011, pp. 643816, doi: 10.1155/2011/643816.

[11] NASEER N., NASEER H. Classification of Normal and Abnormal ECG Signals Based on Their PQRST Intervals. *Conference: 2017 International Conference on Mechanical, System and Control Engineering*, 2017, pp. 388–391, doi: 10.1109/ICMSC.2017.7959507.

[12] BABAEIZADEH S., RICHARD E. Real Time Atrial Fibrillation Monitoring . *Google Patents*. 2013, doi: US8560058B2.

[13] AFKHAMI R.G., AZARNIA G., TINATI M.A. Cardiac Arrhythmia Classification Using Statistical and Mixture Modeling Features of ECG Signals. *Pattern Recognition Letters*. 2016, 70(Jan.15), pp. 45–51, doi: 10.1016/j.patrec.2015.11.018.

[14] LADAVICH S., GHORAANI B. Rate-Independent Detection of Atrial Fibrillation By Statistical Modeling of Atrial Activity. *Biomedical Signal Processing and Control*. 2015, 18, pp. 274–281, doi: 10.1016/j.bspc.2015.01.007.

[15] ZHOU X.L., DING H.X., WU W.Q. A Real-Time Atrial Fibrillation Detection Algorithm Based on The Instantaneous State of Heart Rate. *PLOS ONE*. 2015, 10(9), pp. e0136544., doi: 10.1371/journal.pone.0136544.

[16] HUANG G.B., WANG D.H., LAN Y. Extreme Learning Machines: A Survey. *International Journal of Machine Learning and Cybernetics*. 2011, 2(2), pp. 107–122, doi: 10.1007/s13042-011-0019-y.

[17] HUANG G.B., CHEN L., SIEW C.K. Universal Approximation Using Incremental Constructive Feedforward Networks With Random Hidden Nodes. *IEEE Transactions on Neural Networks*. 2006, 17(4), pp. 879–892, doi: 10.1109/TNN.2006.875977.

[18] HUANG G.B., CHEN L. Convex Incremental Extreme Learning Machine. *Neurocomputing*. 2007, 70(16-18), pp. 3056–3062, doi: 10.1016/j.neucom.2007.02.009.

[19] HUANG G.B., CHEN L. Enhanced Random Search Based Incremental Extreme Learning Machine. *Neurocomputing*. 2008, 71(16-18), pp. 3460–3468, doi: 10.1016/j.neucom.2007.10.008.

[20] XU Y., YANG S. Evolutionary Extreme Learning Machine-Based on Particle Swarm Optimization. *Springer Berlin Heidelberg*. 2006,36(3), pp. 644–652, doi: 10.1007/11759966_95.

[21] HAN F., YAO H.F., LING Q.H. An Improved Extreme Learning Machine Based on Particle Swarm Optimization. *Neurocomputing*. 2011,116, pp. 699–704, doi: 10.1016/j.neucom.2011.12.062.

[22] ZHU Q.Y., QIN A.K., SUGANTHAN P.N. Evolutionary Extreme Learning Machine. *Pattern Recognition*. 2005, 38(10), pp. 1759–1763, doi: 10.1016/j.patcog.2005.03.028.

[23] XUE X., YAO M., WU Z.H, YANG J.H. Genetic Ensemble of Extreme Learning Machine. *Neurocomputing*. 2014, 129, pp. 175–184, doi: 10.1016/j.neucom.2013.09.042.

[24] ZHAO X.G., WANG G.R., BI X.,GONG P.Z. XML Document Classification Based on ELM. *Neurocomputing*. 2011, 74(16), pp. 2444–2451, doi: 10.1016/j.neucom.2010.12.038.

[25] LAN Y., SOH Y.C., HUANG G.B. Ensemble of Online Sequential Extreme Learning Machine. *Neurocomputing*. 2009, 72(13-15), pp. 3391–3395, doi: 10.1016/j.neucom.2009.02.013.

[26] ZHANG Y., LIU B., YANG F. Differential Evolution Based Selective Ensemble of Extreme Learning Machine. *2016 IEEE Trustcom/BigDataSE/ISPA*. 2016, pp. 1327–1333, doi: 10.1109/TrustCom.2016.0211.

[27] HUANG G.B. Learning Capability and Storage Capacity of Two-Hidden Layer Feedforward Networks. *IEEE Transactions on Neural Networks*. 2003,14(2), pp. 274–281, doi: 10.1109/TNN.2003.809401.

[28] GAO Y. M., XING Y. F. The Application of BP Neural Network Based on A Particle Swarm Optimization To The Catalytic Cracking Reaction-Regeneration Process. *Computers and Applied Chemistry*. 2017,34(11), pp. 899–903, doi: 10.16866/j.com.app.chem201711009.

[29] RATNAWEERA A., HALGAMUGE S.K., WATSON H.C. Self-Organizing Hierarchical Particle Swarm Optimizer With Time-Varying Acceleration Coefficients. *IEEE Transactions on Evolutionary Computation*. 2004, 8(3), pp. 240–255, doi: 10.1109/TEVC.2004.826071.

[30] GUO H.L., LI B., LI W., QIAO F.J. Local Coupled Extreme Learning Machine Based on Particle Swarm Optimization. *Algorithms*. 2018, 11(11), pp. 174–189, doi: 10.3390/a11110174.

[31] LI B., LI Y.B., RONG X.W. A Hybrid Optimization Algorithm for Extreme Learning Machine. *Proceedings of the 2015 Chinese Intelligent Automation Conference*. 2015, pp. 297–306, doi: 10.1007/978-3-662-46469-4_31.

[32] MOODY G.B., MARK R.G. The Impact of The MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*. 2001, 20(3), pp. 45-50, doi: 10.1109/51.932724.

[33] DIKER A., AVCI E. Feature Extraction of ECG Signal by Using Deep Feature. *Proceedings of the 2019 7th International Symposium on Digital Forensics and Security*, 2019, pp. 1–6, doi: 10.1109/ISDFS.2019.8757522.

[34] HASAN N.L., BHATTACHARJEE A. Deep Learning Approach to Cardiovascular Disease Classification Employing Modified ECG Signal from Empirical Mode Decomposition. *Biomedical Signal Processing and Control*. 2019, 52, pp. 128-140, doi: 10.1016/j.bspc.2019.04.005.

[35] PANDEY S.K., JANGHEL R.R., VANI V. Patient Specific Machine Learning Models for ECG Signal Classification. *Procedia Computer Science*. 2020, 167, pp. 2181–2190, doi: 10.1016/j.procs.2020.03.269.